UNIVERSIDAD DE GRANADA

PROGRAMA OFICIAL DE POSGRADO EN ESTUDIOS AVANZADOS DE TRADUCCIÓN E INTERPRETACIÓN

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

DEPARTAMENTO DE TRADUCCIÓN E INTERPRETACIÓN

DOCTORAL THESIS

PHRASEOLOGY IN SPECIALIZED LANGUAGE AND ITS REPRESENTATION IN ENVIRONMENTAL KNOWLEDGE RESOURCES

MIRIAM BUENDÍA CASTRO

GRANADA, 2013

Editor: Editorial de la Universidad de Granada
Autor: Miriam Buendía Castro
D.L.: GR 1906-2013
ISBN: 978-84-9028-597-8

UNIVERSIDAD DE GRANADA

DOCTORAL THESIS

PHRASEOLOGY IN SPECIALIZED LANGUAGE AND ITS REPRESENTATION IN ENVIRONMENTAL KNOWLEDGE RESOURCES

MIRIAM BUENDÍA CASTRO

SUPERVISED BY

PROFESSOR PAMELA FABER and

PROFESSOR CLARA INÉS LÓPEZ RODRÍGUEZ

GRANADA, 2013

El doctorando **MIRIAM BUENDÍA CASTRO** y los directores de la tesis **DOÑA PAMELA FABER BENÍTEZ** y **DOÑA CLARA INÉS LÓPEZ RODRÍGUEZ** garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 1 de febrero de 2013

Fdo.: Dña. Pamela Faber Benítez

Director/es de la Tesis

Fdo.: Dña. Clara Inés López Rodríguez

Director/es de la Tesis

Fdo.: Miriam Buendía Castro

Doctorando

To my mother, and to the memory of my father

Acknowledgements

There are many people to whom I would like to express my sincere gratitude. Each of them has directly or indirectly lent a helping hand over the last four years.

First of all, I am deeply grateful to Professor Pamela Faber for giving me the opportunity to work with her. I also sincerely appreciate her wise guidance, and all the time and support she has given me, especially in this last stage of the thesis. Behind each line in this thesis is her voice. I will always be in debt to her.

I would also like to thank my co-director Professor Clara Inés López Rodríguez for her encouragement throughout the years. Thanks to her excellent advice (always accompanied by a smile) and her sense of perfection, many pitfalls were averted. I will always treasure her friendship and affection.

My sincere thanks also go to Professor Rita Temmerman. During my two research stays at the Erasmushogeschool, in Brussels, she made me feel at home and helped me in my research by giving me valuable insights into terminology at crucial stages of the thesis. In fact, this thesis would not have been the same without her suggestions.

Professor Silvia Montero is also a person whose affection, help, and advice has aided me greatly in both personal and professional issues. Indeed, to a certain extent, she is responsible for the topic since it carries on the research in her thesis on phraseology in Oncoterm.

I am also grateful to the members of the research group Lexicon: (i) to Beatriz Sánchez, for giving me many insights into the semantics of verb; (ii) to Ana Belén Pelegrina, who helped me to computationally implement the phraseological module in EcoLexicon; (iii) to José Manuel Ureña, for always being available to help with conference papers; (iv) to Pilar Léon, Pedro Magaña, and Antonio San Martín for their time and help, especially regarding computer issues.

I would also like to thank my colleagues from the Erasmushogeschool in Brussels, especially to Lara Sanz, for given me so many ideas for this research over lunch. Thanks also go to Andrés Domínguez, Koen Kerremans, and Toon Cox for always being there to answer any doubts, and to Professor Philippe Humblé for giving me the opportunity of giving lectures abroad.

I would also like to especially thank my friends, Loli, Marta, Leyre, Elsa, Silvia, and Carmen for their always support, particularly in difficult moments.

My gratitude especially goes to my parents and my brother. Their endless love, affection and support have been extremely important during this time. Without their love, I would never have been able to finish.

Finally, I would like to thank Ernesto for having become the most important person in my life, for his love and his endless support, and for the twins we are expecting. For a lifetime together.

Table of contents

List of Tables

List of Figures

List of Abbreviations

ADs	Accounting Dictionaries
AVP	adverbial phrase
BBI	BBI Combinatory Dictionary of English
BootCat	Bootstrapping Corpora and terms from the web
BNC	British National Corpus
CIDE	Cambridge International Dictionary of English
CT	candidate term
CG	Cognitive Grammar
COBUILD	Collins Cobuild Dictionary of English
CTT	Communicative Theory of Terminology
CREA	Corpus de Referencia del Español Actual
PRÁCTICO	Diccionario combinatorio práctico del español contemporáneo
DICE	Diccionario de colocaciones del Español
DRAE	Diccionario de la Real Academia Española
DTEFC	Diccionario de términos económicos, financieros y comerciales inglés-español, Spanish-English
DUE	Diccionario de Uso del Español
DICOFE	Dictionnaire contextuel du français économique
DAFA	Dictionnaire d'apprentissage du français des affaires
DiCoEnviro	Dictionnaire fondamental de l'environnement
DiCoInfo	Dictionnaire fondamental de l'informatique et de l'Internet
ETL	electronic text library
EIONET	Environment Information and Observation Network
EE	Environmental Event
esp.	especially
EEA	European Environment Agency
EAGLES	Expert Advisory Group on Language Engineering Standards
ECD	Explanatory Combinatorial Dictionary
FE	frame element
FBT	Frame-based Terminology
FLM	Functional Lexematic Model
GEMET	GEneral Multilingual Environmental Thesaurus
GTT	General Terminology Theory
HTML	Hypertext Markup Language
ICM	idealized cognitive model
KWIC	key word in context
KP	knowledge pattern
LOB corpus	Lancaster-Oslo/Bergen Corpus of British English
LGP	language for general purposes
LSP	language for specific purposes
DEC	Le dictionnaire explicatif et combinatoire du français contemporaine
LCM	Lexical Constructional Model
LD	lexical domain
LF	lexical function
LGM	Lexical Grammar Model
LU	lexical unit
LS	logical structure
LDOCE	Longman Dictionary of Contemporary English

MCD	Macmillan Collocations Dictionary
MTT	Meaning-Text Theory
NLP	Natural Language Processing
NSM	Natural Semantic Metalanguage
noun phrase	NP
OCR	Optical character Recognition
OALDCE	Oxford Advanced Learner's Dictionary of Current English
Oxford	Oxford bilingual English Spanish dictionary
OCD	Oxford Collocations Dictionary for Students of English
PT	phrase type
PP	prepositional phrase
REDES	Redes: Diccionario combinatorio del español contemporáneo
RRG	Role and Reference Grammar
SERP	search engine report page
SkE	Sketch Engine
STT	Sociocognitive Terminology Theory
sb	somebody
sth	something
SL	source language
STTR	standardized type/token ration
TL	target language
OPUS	The Open Parallel Corpus
TLD	top level domain
TTR	type/token ratio
UNISDR	United Nations Office for Disaster Risk Reduction
UoU	unit of understanding
URL	Uniform Resource Locator
WaC	web as corpus
Wacky project	Web as corpus kool ynitiative project
WfC	web for corpus

Abstract

Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources

This research presents the design of a template for encoding and describing phraseological information in the environmental knowledge base, *EcoLexicon* (ecolexicon.ugr.es). One of the objectives of this study was to formulate a methodology that could also be used to codify phraseological information in other specialized knowledge domains. The focus of our analysis was on verb collocations (verb+noun, noun+verb) because of the need to enrich specialized knowledge resources with information regarding predicates and their arguments. As is well known, verbs are an extremely important part of language since, in a manner of speaking, they set the scene or establish the structure for the rest of the sentence. However, very few specialized knowledge resources include them.

The phraseological template proposed in this research takes its theoretical premises from cognitive approaches to Terminology, namely, Sociocognitive Terminology and Frame-based Terminology. In addition, assumptions from linguistic models that deal with predicate argument structure were also included, such as the Lexical Grammar Model, Lexical Constructional Model, Role and Reference Grammar, and FrameNet.

The practical guidelines for the design of our template were based on a detailed analysis of the most representative lexicographic and terminographic resources that contain phraseological information. A set of the most important meaning-based resources for verb description was also analyzed.

The underlying idea of our research is that verbs and their potential arguments can be classified and structured in a set of conceptual-semantic categories typical of a given specialized domain. In this context, when semantic roles and macroroles are specified as well as the resulting phrase structure, it is then possible to establish templates that represent this meaning for entire frames. In this regard, within the context of a specialized knowledge domain, the range of verbs generally associated with potential arguments can be predicted within the frame of a specialized event. This occurs, of course, because the nature of arguments is constrained by verb meaning. However, it is

our assertion that this influence is reciprocal since in specialized language, verbs are also to some extent constrained by their arguments. In this regard, there is an interaction between the meaning components of the entities and processes activated in specialized knowledge event representations.

Finally, the results of our analysis have been implemented in EcoLexicon. As shown, the methodology proposed in this thesis for encoding and describing verbal collocations in terminographic resources is useful both for text comprehension and text production.

Keywords: *collocations, specialized resources, frames*

Summary of this doctoral thesis in Spanish

Aunque la mayoría de los terminólogos, lexicógrafos, terminógrafos y los usuarios de sus productos estén de acuerdo en que las colocaciones son absolutamente imprescindibles en los diccionarios y el resto de productos terminográficos, muy pocos recursos las integran (L'Homme y Leroyer 2009: 260) y, cuando lo hacen, no presentan ningún tipo de sistematización (Montero y Buendía 2012). En concreto, para el proceso traductor y el aprendizaje de lenguas, el aspecto combinatorio entre unidades léxicas resulta fundamental. Los pocos recursos que las incluyen suelen limitarse a su enumeración, bien como entradas independientes o bien dentro de la unidad nominal de la que forman parte, sin ningún tipo de información semántica o acerca de su uso. Además, en la mayoría de estos productos terminográficos se ha favorecido el estudio de las unidades nominales, en detrimento de las unidades verbales, a pesar de que los verbos se consideran los principales responsables de la transmisión de conocimiento (L'Homme 1998).

La presente tesis se desarrolla en el marco del grupo Lexicon (http://lexicon.ugr.es), cuya directora, la Profesora Pamela Faber, desarrolló un nuevo enfoque de naturaleza cognitiva para el estudio de la Terminología, conocido como *Terminología basada en Marcos.* La aplicación práctica de la Terminología basada en Marcos se ha materializado en la implementación de una base de conocimiento especializada en el medio ambiente llamada *EcoLexicon* (http://ecolexicon.ugr.es). EcoLexicon representa la estructura conceptual del dominio del medio ambiente a través de un tesauro en el que los conceptos se organizan en redes semánticas de conocimiento. EcoLexicon cuenta a día de hoy con más de 3.500 conceptos y alrededor de 18.500 términos en inglés, español, alemán, griego moderno, ruso y francés. Lo que confiere a EcoLexicon su carácter especial e innovador es que, a diferencia de los tesauros convencionales, no se limita a agrupar las distintas unidades léxicas por campos semánticos, sino que proporciona además información gráfica, conceptual y lingüística. De esta forma, para cada concepto se ofrece su definición, imágenes que evocan el concepto y ayudan a su comprensión, así como todos los términos asociados al concepto en las distintas lenguas. Asimismo, para cada término, se proporciona información acerca de su categoría gramatical (sustantivo, verbo, adjetivo o adverbio); el tipo de término (término principal, sinónimo, variante geográfico, o acrónimo); el

género, para las lenguas que lo requieran (masculino, femenino y neutro), así como contextos de uso que muestran el término en contexto.

Sin embargo, a pesar de la gran cantidad de información lingüística que proporciona para cada unidad terminológica, hasta el momento EcoLexicon no ofrecía ningún tipo de información fraseológica, salvo la inclusión de unidades fraseológicas nominales, a pesear de que la información fraseológica constituye una parte indispensable del conocimiento lingüístico y es de extrema utilidad para un gran número de usuarios, entre los que se encuentra el traductor. Para constatarlo, elaboramos un cuestionario para estudiar la satisfacción de los usuarios de EcoLexicon y lo distribuimos entre dos grupos de alumnos (44 en total) de tercer curso de la Licenciatura en Traducción e Interpretación de la Universidad de Granada (López, Buendía y García 2012; §1.5). Entre otras cosas, los alumnos destacaron la importancia de disponer de información fraseológica en los diccionarios y resto de productos terminográficos, y echaron de menos la inclusión de este tipo de información terminológica en EcoLexicon.

Así pues, esta tesis propone la incorporación de un módulo fraseológico en EcoLexicon, que permita incluir información colocacional para las unidades de significación especializada contenidas en EcoLexicon. En concreto, en esta tesis abordamos la inclusión de las colocaciones tipo *verbo + nombre* o *nombre + verbo* porque recientes estudios han destacado el papel fundamental del verbo en el discurso especializado (López 2007, Lorente 2007, Buendía 2012) y, desafortunadamente, en los estudios terminológicos siempre se ha favorecido el estudio de las unidades nominales en detrimento de las unidades verbales (L'Homme 1998), por lo que hay muy pocos recursos terminográficos que incorporen el estudio del verbo. Por tanto, los objetivos específicos de esta tesis son:

- Definir el perfil de usuario y las situaciones de comunicación de nuestra propuesta.
- Sentar la base teórica de nuestro modelo de codificación de información fraseológica verbal.
- Establecer las directrices para decidir el tipo de información fraseológica que se debería incluir en una entrada lexicográfica o terminográfica, teniendo en cuenta el perfil de usuario de un traductor.
- Analizar descriptivamente un corpus de textos de medio ambiente, y más concretamente de desastres naturales, destacando la importancia de combinar

metodologías más tradicionales de la Web para recopilar un corpus *(Web for Corpus, WfC*), con metodologías más novedosas de la Web como corpus (*Web as Corpus, WaC*).

- Diseñar un protocolo para la evaluación de la fiabilidad de los recursos web que asegure la calidad y fiabilidad de los textos que posteriormente pasen a formar parte de un corpus.
- Determinar un inventario de categorías semánticas y de roles semánticos relevantes para el subdominio del EVENTO EXTREMO.
- Establecer la configuración de los marcos activados en el EVENTO EXTREMO.

Para llevar a cabo este estudio, nos centramos en un subdominio del Medio Ambiente, el de los RIESGOS NATURALES, también conocido como EVENTO EXTREMO, en consonancia con la relevancia otorgada a los *eventos* en la Terminología basada en Marcos (§1.4.3). El motivo es el creciente interés que está suscitando este campo debido a que la Tierra está sufriendo cada vez más desastres naturales de mayor intensidad y con mayor frecuencia. Todo ello ha hecho que el medio ambiente haya pasado de ser un campo relegado al ámbito especializado a convertirse en una disciplina de interés para distintos perfiles de usuarios en todo el mundo.

Para poder sentar las bases teóricas de nuestro análisis fraseológico verbal ha sido necesaria, por un lado, una revisión de las teorías terminológicas propiamente dichas (§1), así como de las principales teorías lingüísticas que abordan el estudio de la *estructura argumento-predicado* (§3). Asimismo, para poder diseñar una plantilla que permita la codificación y descripción de las colocaciones verbales en EcoLexicon, hemos revisado nociones fundamentales en Fraseología y hemos analizado los principales recursos lexicográficos y terminográficos que incluyen información fraseológica (§2.2-§2.4), así como los recursos más representativos dedicados a la descripción del verbo (§3.2).

En este sentido, en el primer capítulo de esta tesis, se describen las teorías terminológicas cognitivas que sientan las bases teóricas de este trabajo. Para ello, se incluye primeramente una breve introducción acerca de la dicotomía Lexicología/Lexicografía y Terminología/Terminografía (§1.2), seguida de un resumen de los principios fundamentales de la Teoría de las funciones lexicográficas de Bergenholtz y Tarp (§1.3). Esta teoría nos servirá para delimitar el usuario potencial de nuestro estudio, así como sus necesidades básicas (§1.5). En el apartado 1.4 se entra de

lleno en la discusión de las teorías terminológicas con aproximación cognitiva en las que se basa nuestra investigación: la Teoría Sociocognitiva de Temmerman (§1.4.2), y en especial, la Terminología basada en Marcos (§1.4.3). Es por esto por lo que gran parte del capítulo se dedica a la descripción minuciosa de la información contenida en Ecolexicon y a destacar la necesidad de dotar a EcoLexicon de información fraseológica.

El capítulo dos se centra en el estudio de la Fraseología desde la Lexicografía y la Terminografía. Comenzamos haciendo un breve recorrido por la historia reciente de la fraseología para poder delimitar el concepto de fraseología y de unidad fraseológica (§2.1.2). Asimismo se detallan los enfoques actuales a la hora de abordar el estudio de la fraseología (§2.13): a) el enfoque semántico, que sostiene que existe una relación de significado entre los distintos componentes de una colocación (§2.1.3.1); y b) el enfoque basado en frecuencias estadísticas, en el que para que una combinación de palabras sea considerada colocación tiene que tratarse de una combinación con una alta frecuencia de aparición conjunta (§2.1.3.2). Conviene destacar que, dentro del enfoque de corte semántico, se dedica una parte importante a Mel'čuk y a su Teoría Sentido Texto *(Meaning-Text Theory),* dado que esta teoría ha sido y continúa siendo la más influyente en la mayoría de los diccionarios combinatorios en cualquier lengua. Seguidamente se describe la aproximación al concepto de colocación que se aplicará para cumplir los objetivos de esta tesis. En esta investigación entendemos por colocación la combinación frecuente de dos o más palabras, formada por los patrones *nombre+verbo* o bien *verbo+nombre,* en la que el nombre es la base y el verbo es el colocativo o colocador. En nuestro enfoque y en el subdominio de los RIESGOS NATURALES, el significado del colocativo (el verbo) está condicionado —y en muchos casos queda impuesto— por el significado de la base (el nombre), al mismo tiempo que el verbo restringe los posibles argumentos que pueden combinarse con él. Por ejemplo, en la combinación 'el fuego arde', el predicado 'arder' sólo admite nombres que se refieran a entidades que puedan estar en combustión (por ejemplo, el fuego), y por otro lado, 'fuego' precisa de un verbo que designe un proceso de combustión ('arder', por ejemplo). Así pues, podemos concluir diciendo que nuestra aproximación al concepto de colocación aúna tanto premisas de los enfoques semánticos como de los estadísticos y establece, por ende, que el significado de la colocación depende tanto del colocativo como de la base. Por ello, en la fase de análisis se estudiará tanto el comportamiento de los argumentos como el de los verbos, y en la implementación de la información

fraseológica en EcoLexicon, se podrá acceder a información colocacional tanto a partir de la base, como a partir del colocativo.

Posteriormente, se describe la forma en la que los diccionarios de la lengua general, tanto en inglés como en español, abordan el fenómeno de la colocación en sus entradas. Se establece la siguiente clasificación: a) diccionarios monolingües generales; b) diccionario bilingües; c) diccionarios combinatorios o exclusivos de colocaciones. Dentro de los diccionarios combinatorios, se revisan los diccionarios para la lengua inglesa: (i) *BBI Combinatory Dictionary of English* (1986, 1997, 2009), (ii) *Oxford Collocations Dictionary for students of English* (2002, 2009), y (iii) *Macmillan Collocations Dictionary* (2010); y a continuación, se revisan los diccionarios combinatorios para la lengua española: (i) *Redes: Diccionario combinatorio del español contemporáneo* (Bosque 2004), (ii) *Diccionario combinatorio práctico del español contemporáneo* (Bosque 2006), y (iii) *Diccionario de colocaciones del Español* (Vincze, Mosqueira, and Alonso 2011). Seguidamente se hace lo mismo con algunos de los recursos especializados que incorporan colocaciones. En concreto se analizan los siguientes recursos: (i) *Lexique de cooccurrents—Bourse et conjuncture économique* (Cohen 1986), especializado en el dominio de la Bolsa; (ii) *Vocabulaire et cooccurrents de la comptabilité* (Caignon 2001), diccionario monolingüe en francés de Contabilidad con equivalencias en inglés; (iii) *Dictionnaire d'apprentissage du français des affaires* (DAFA) (Binon et al. 2001), diccionario monolingüe en francés especializado en el dominio de los Negocios disponible tanto en papel como en versión electrónica con equivalencias en inglés, español, alemán, neerlandés, e italiano; (iv) *Dictionnaire contextuel du français économique* (DICOFE) (Verlinde et al. 1993-2003), cuya versión en línea permite realizar búsquedas de información colocacional relativa al dominio de los Negocios desde el neerlandés hacia el francés; (v) *Accounting Dictionaries* (ADs), grupo de cinco diccionarios de contabilidad; (vi) *Diccionario de términos económicos, financieros y comerciales (inglés-español, Spanish-English),* diccionario en papel bilingüe inglés-español de Economía, Finanzas y Comercio; (vii) *Internet. Répertoire bilingue de combinaisons lexicales spécialisées français anglais* (Meynard 2000), diccionario ingles-francés de términos relativos a Internet; (viii) *Dictionnaire fondamental de l'informatique et de l'Internet* (DiCoInfo), diccionario en línea en francés, inglés y español, especializado en el dominio de Internet y la computación; (ix) *Dictionnaire fondamental de l'environnement* (DiCoEnviro), diccionario incipiente en línea en francés, inglés y español, centrado en el subdominio del cambio climático; (x)

Termium Plus®, base de datos terminológica y lingüística de enormes dimensiones, que cubre casi todas las disciplinas del saber en inglés, francés, español y portugués.

Por último, con los aspectos positivos y negativos derivados del análisis de los distintos recursos lexicográficos y terminográficos, se establecen las directrices para decidir el tipo de información fraseológica que se debería incluir en una entrada, teniendo en cuenta el perfil de usuario de un traductor (§2.4).

En el capítulo tres, se describen las teorías lingüísticas que pueden aplicarse al estudio de la fraseología verbal en Terminología, y más concretamente, aquellas cuyas premisas han resultado particularmente útiles para nuestro estudio. En este sentido, nuestra investigación incorpora la noción de *dominio léxico* y su categorización en términos de relaciones paradigmáticas y sintagmáticas del Modelo de la Gramática Léxica (*Lexical Grammar Model*; §3.1.2.1); la caracterización de roles temáticos y macrorroles de la Gramática del Papel y la Referencia (*Role and Reference Grammar;* §3.1.2.2), la idea subyacente de plantilla léxica del Modelo Léxico Construccional (*Lexical Constructional Model;* §3.1.3.4), así como algunas nociones de FrameNet (§3.1.3.2).

En el capítulo tres se presentan asimismo los principales recursos léxicos que contienen información verbal. Nos referimos a WordNet (§3.2.1), VerbNet (§3.2.2) y PropBank (§3.2.3), para la lengua inglesa, y a ADESSE (§3.2.4) y SenSem (§3.2.5), para el español. Las conclusiones derivadas del tratamiento de los predicados en estos recursos, unidos a las conclusiones del análisis de recursos lexicográficos y terminográficos llevado a cabo en el capítulo dos, nos sirvieron para la posterior implementación de nuestra plantilla fraseológica.

El capítulo cuatro describe los materiales y métodos usados para llevar a cabo nuestro análisis. En consonancia con la Terminología basada en Marcos, en esta tesis se ha seguido tanto un enfoque *top-down* como *bottom-up.* En primer lugar, se delimitan las fronteras de nuestro estudio dentro del subdominio de los RIESGOS NATURALES tras la revisión de obras de referencia, tesauros y diccionarios especializados *(top-down)* y el análisis de corpus para extraer los términos, y por ende, los conceptos más frecuentes en este subdominio *(bottom up).* En este sentido se determina que los conceptos fundamentales que subyacen en este ámbito son los siguientes: VOLCÁN, DESLIZAMIENTO DE TIERRA, INUNDACIÓN, AVALANCHA, CICLÓN TROPICAL (incluyendo TIFÓN y HURACÁN), TORNADO, SEQUÍA, TERREMOTO, FUEGO y TSUNAMI. Seguidamente se describe el diseño, compilación y composición de nuestro corpus. Para ello,

primeramente, se hace una revisión del concepto de corpus (§4.2.1.1), así como de la tipología de corpus (§4.2.1.2) y de los métodos o enfoques utilizados en la actualidad para la compilación de corpus: (i) *la Web para recopilar un corpus* (*Web for Corpus, WfC*) (§4.2.2.1), en la que la Red se utiliza como fuente de textos en formato electrónico para la posterior compilación de corpus *off-line*; (ii) *la Web como corpus* (*Web as Corpus, WaC*) (§4.2.2.2), que utiliza la Red directamente como si fuera un corpus propiamente dicho. En nuestro análisis se han aunado los dos enfoques al estudio de corpus. Por un lado, compilamos un corpus especializado en el dominio de los RIESGOS NATURALES en inglés y otro en español, cuyas características están descritas en 4.2.3. A este respecto, para asegurar la calidad de los textos que iban a formar parte de nuestro corpus, diseñamos un protocolo para la evaluación de la fiabilidad de los recursos en línea de los que se extraen los textos que posteriormente pasan a formar parte de un corpus (§4.2.2.1.2). Por otro lado, también hicimos uso de la WaC para encontrar más ejemplos de uso acerca de un término cuando los contextos de nuestro corpus resultaban insuficientes, o simplemente queríamos comprobar o refutar una hipótesis a través del estudio de más contextos de uso. Dentro de la WaC, nos resultaron de especial interés, el buscador Google, así como los programas de pre-post procesamiento *Webcorp* (http://www.webcorp.org.uk/live/) y el de la editorial Springer, *Exemplar* (http://www.springerexemplar.com/).

Por último, describimos las herramientas informáticas utilizadas en nuestro análisis; a saber, el extractor de términos *TermoStat* (§4.3.1), y el programa de análisis de corpus, *WordSmith Tools* (§4.3.2). TermoStat nos sirvió para seleccionar los verbos más frecuentes de nuestro dominio y para, posteriormente, extraer los verbos objeto de análisis en esta tesis. En relación con Wordsmith Tools, la opción de concordancias fue de gran utilidad para analizar el comportamiento de los verbos seleccionados.

En el capítulo cinco, se describe con minuciosidad el proceso de análisis seguido, los resultados obtenidos, así como el diseño de una plantilla de análisis de verbos y la implementación de un nuevo módulo de información fraseológica en EcoLexicon.

En primer lugar, se extrajeron los verbos más recurrentes de nuestro corpus a través del extractor de términos *TermoStat*, mediante dos búsquedas, una para inglés y otra para español. Del total de verbos ofrecidos por TermoStat, se descartaron aquellos que, a pesar de haber sido etiquetados como verbos, correspondían a otras categorías sintácticas. Los demás verbos ofrecidos por el extractor de términos, aunque muchos de ellos no eran específicos del campo de los RIESGOS NATURALES o EXTREMO EVENTO, se

clasificaron en dominios y subdominios del Modelo de la Gramática Léxica (§3.1.2.1), de acuerdo con su definición, tal y como se detalla en 5.4.1. Esta clasificación preliminar nos permitió, asimismo, comprobar cuáles eran los dominios más prototípicos activados en el EXTREMO EVENTO y los que posteriormente serían objeto de nuestro estudio. Nos referimos a los dominios de EXISTENCIA, ACCIÓN, MOVIMIENTO, POSICIÓN y CAMBIO.

Una vez que los verbos estuvieron clasificados en dominios y subdominios de acuerdo a su definición, el significado de los verbos se perfiló a través del estudio de su comportamiento en los textos, es decir, a través del análisis de las líneas de concordancias con WordSmith Tools (*bottom-up*). Se analizaron todas las líneas de concordancia de cada verbo y, tal como se ha mencionado con anterioridad, se acudió a la web cuando se necesitaban más instancias. De hecho, este análisis sirvió para descartar por completo aquellos verbos que no estaban directamente asociados a los RIESGOS NATURALES. El criterio seguido fue el de desechar todos los verbos que no llevaran como argumentos alguna realización lingüística que apuntara a un desastre natural o fenómeno relacionado.

En concreto, en este análisis *bottom up*, se analizaron tanto los argumentos, por un lado, como los predicados, por otro. Para ello, las líneas de concordancia ofrecidas por WordSmith Tools para cada verbo se registraron en un archivo Excel. El proceso seguido para el análisis de los argumentos fue el siguiente:

1. Se identificaron las diversas realizaciones lingüísticas para el mismo argumento.
2. Para el conjunto de realizaciones lingüísticas que designaban el mismo tipo de argumento se asignó una misma etiqueta semántica. La idea subyacente es que las distintas realizaciones lingüísticas de un mismo tipo de argumento, evocarán el mismo tipo de información conceptual y presentarán un comportamiento semántico y sintáctico similar. Esta premisa sirvió para posteriormente establecer una serie de categorías semánticas pertinentes a nuestro dominio de RIESGOS NATURALES. Las categorías establecidas fueron las siguientes: NATURAL DISASTER, ATMOSPHERIC AGENT, WATER AGENT, ATMOSPHERIC CONDITION, MATERIAL ENTITY, AREA, CONSTRUCTION, ENERGY, HUMAN BEING, LANDFORM, WATER COURSE, DEATH, DAMAGE, LOSS OF LIFE/PROPERTY, PLANT, y EXPLOSIVE. El análisis completo para la asignación de etiquetas puede consultarse en 5.4.2.1.1.

3. Se asignó un rol semántico y macrorrol a cada argumento. Los roles semánticos de nuestro estudio coinciden con los roles temáticos más generales de la Gramática de la Referencia y el Papel (§3.1.2.2) y con los roles argumento propuestos por Goldberg (§3.1.3.3). Otros se extrajeron del inventario de roles contenidos en los recursos VerbNet (§3.2.2), ADESSE (§3.2.4) y Sensem (§3.2.5). Así pues, el inventario de roles semánticos utilizados en esta investigación son: AGENT, NATURAL FORCE, DESTINATION, EXPERIENCER, FREQUENCY, GEOGRAPHICAL LOCATION, MANNER, PATH, PATIENT, SITUATION/EXPERIENCE, ORIGIN, THEME, TIME y RESULT. Asimismo, junto con las roles temáticos, se proporcionó una categorización más general de cada argumento de acuerdo con su macrorrol, es decir en términos de ACTOR y UNDERGOER, en consonancia con la noción de macrorroles de la Gramática de Referencia y el Papel (§3.1.2.2).
4. Se clasificó cada argumento según su estructura morfosintáctica, distinguiendo entre las siguientes categorías: (i) sintagma nominal; (ii) sintagma adverbial; y (iii) sintagma preposicional. En el caso de los sintagmas preposicionales, las preposiciones más frecuentes del sintagma se ofrecieron por separado.

Posteriormente, todos los verbos que activaban el mismo número y tipo de categorías semánticas se agruparon. A ese conjunto de verbos se les asignó un nombre con el que nosotros definimos cada uno de los *marcos* [frames] dentro del dominio de los RIESGOS NATURALES. A pesar de que estos marcos se parecen más bien a los subdominios establecidos por el Modelo de la Gramática Léxica, decidimos denominarlos marcos porque la Terminología basada en Marcos, como bien indica su nombre, es un enfoque basado en marcos. Una vez hecho esto, cada uno de los marcos identificados se definió teniendo en cuenta, primeramente, el dominio léxico subyacente y de acuerdo con las categorías semánticas asignadas a sus argumentos. Conviene señalar, no obstante, que una posterior evaluación fue necesaria puesto que no todos los verbos con el mismo número de argumentos y con el mismo tipo de categorías semánticas, presentaban el mismo comportamiento. Algunas veces las diferencias entre los distintos verbos eran de registro; otras veces, aunque los distintos verbos pertenecientes a un mismo marco activaban un mismo significado subyacente, algunos de los verbos imponían ciertas restricciones de uso en sus argumentos. Incluso, se comprobó, a través de la definición de los propios verbos, que algunos de los que fueron

identificados con las mismas categorías, apuntaban a marcos diferentes. Todas estas diferencias entre los verbos de un mismo marco, se incluyen en el apartado reservado para *notas* dentro de la plantilla de análisis de cada verbo.

Cabe destacar que el análisis se hizo primeramente para el inglés y después se hizo para el español siguiendo la misma metodología. En este sentido, para el análisis en español, una vez identificadas las categorías semánticas asociadas a cada argumento, se agruparon los verbos con las mismas categorías y se asociaron directamente con los marcos previamente establecidos para la lengua inglesa, en consonancia con la metodología que propone Pimentel (2012) para la asignación de verbos equivalentes en el dominio del Derecho.

Una vez hecho esto, se rellenó una plantilla con el análisis de cada verbo por separado, que sirvió para la posterior implementación en EcoLexicon del módulo de información fraseológica. Tal y como se detalla en 5.4.3, cada plantilla proporciona una descripción del marco activado por el verbo y se compone de los siguientes elementos: (i) el dominio léxico activado por el marco; (ii) el nombre del marco; (iii) la definición del marco basada en las categorías semánticas o roles semánticos activados por los argumentos con los que co-occure cada verbo específico; (iv) una nota de uso, para cuando sea necesario. Una vez descrito el marco, los argumentos activados por cada verbo se especifican en la plantilla, se clasifican y se describen de acuerdo con sus realizaciones lingüísticas, categorías semánticas, roles, macrorroles y categoría morfosintáctica (tipo de sintagma). Finalmente, además de la descripción del marco y de los argumentos de cada verbo, en la plantilla se recogen ejemplos de uso de cada verbo. Las distintas oraciones ofrecidas, a modo de ejemplo, para cada verbo se anotan con diferentes colores que permiten la identificación rápida de sus argumentos de forma parecida a como lo hace FrameNet para sus *frame elements*.

Así pues, para cada uno de los marcos identificados como pertinentes dentro del dominio de los RIESGOS NATURALES o EXTREME EVENTO, se ofrecen dos tablas: una que contiene todos los verbos activados en inglés, siguiendo el modelo de plantilla previamente descrito en el párrafo anterior; y otra que contiene los verbos en español activados en el mismo marco. Finalmente se incluye una plantilla resumen del marco, que codifica los requisitos y restricciones de cada marco de acuerdo con sus categorías semánticas, roles, macrorroles y estructura morfosintáctica, con lo que sirve para establecer generalizaciones acerca de cada marco. El resultado completo del análisis de cada marco se detalla en 5.4.3.

Al final del capítulo cinco se muestra cómo se implementó el modulo fraseológico en EcoLexicon. En este sentido, primeramente se fue guardando toda la información contenida en las tablas de análisis, en la base de almacenamiento de EcoLexicon de acceso restringido para el grupo de investigación. En concreto, en el módulo reservado para el término dentro de la base de datos, se creó una sección aparte para la fraseología. Conviene señalar que para facilitar el acceso al traductor, que no olvidemos es el usuario potencial contemplado en esta tesis, no se utilizó ningún tipo de metalenguaje en la codificación ni posterior despliegue de la información en EcoLexicon. Por esta razón, las nociones de categoría semántica, rol y macrorrol quedaron excluidas. No obstante, estas nociones nos ayudaron a configurar los marcos y en un futuro nos ayudarán a sistematizar los patrones de los verbos, y a contribuir en la implementación de la incipiente ontología que se está elaborando para Ecolexicon.

Tal y como se detalla en 5.5, las colocaciones verbales en EcoLexicon se clasifican y describen de acuerdo a su significado. Es por este motivo por el que las colocaciones de cada término se clasifican, en primer lugar, de acuerdo con su dominio léxico, seguido del marco que activan. Una vez especificado el dominio léxico y el marco, se presenta una lista de todos los verbos pertenecientes al marco y activados por el término. Al hacer clic en cada uno de los verbos, se despliega una nueva ventana que ofrece los ejemplos de uso asociados al verbo, así como una sección de notas con información relativa a restricciones de uso o significado, si el verbo lo precisa.

Finalmente, las correspondencias entre las colocaciones en inglés y en español también se pueden encontrar en EcoLexicon. Si en la sección de términos de EcoLexicon, todos los términos en las distintas lenguas se refieren al mismo concepto, en principio, creemos que en su realización lingüística activarán información colocacional con un significado subyacente similar. Esto se traduce en que, en líneas generales, los marcos activados por un término en una lengua, van a ser muy parecidos a los marcos del término en la otra lengua, aunque cada lengua los lexicalice con verbos diferentes, lo que justifica que el número de realizaciones verbales para un mismo marco en distintas lenguas varíe. Así pues, en la combinación inglés-español para encontrar en EcoLexicon la correspondencia de un marco o de un verbo en la otra lengua, bastaría con hacer clic sobre el término en la lengua que estemos interesados, y desplegar el módulo de información fraseológica de ese término en esa lengua. En el futuro se prevé vincular los marcos de una lengua con los marcos de otra para que la correspondencia sea directa. No se ha hecho ahora por la complejidad que supone desde

el punto de vista computacional, ya que supondría reestructurar por completo todas las relaciones de base de EcoLexicon.

De esta forma, tal y como se detalla en 5.5.2, el modulo fraseológico de EcoLexicon propuesto en esta tesis permite a los usuarios acceder a información colocacional tanto a través de la base, es decir, la unidad de significación especializada, como a través del colocativo, es decir, el verbo. Por ejemplo, a la colocación 'huracán asola', se puede acceder a través de la entrada 'huracán' o buscando a través del verbo 'asolar'. Este método de codificar la información es de utilidad tanto para la traducción de una L1 a una L2 (por ejemplo, para traducir hacia el inglés la expresión 'el huracán asoló'), como para la producción en L2 de una colocación en la que el colocativo precisa de un significado específico (por ejemplo, para saber qué verbo coloca con huracán, cuando se quiere transmitir la idea de *cause to change for the worse*), en cuyo caso, el sistema ofrecerá, entre otros verbos, el verbo 'asolar'.

Para concluir podemos decir que la plantilla de información fraseológica propuesta en esta tesis para codificar y describir información fraseológica verbal cumple con los diez criterios que, según el capítulo 2 (§2.4), debería tener el diseño de una entrada lexicográfica o terminográfica ideal que quiera dar cuenta de información fraseológica:

1. EcoLexicon es un recurso que está disponible en línea, por lo que las entradas con información fraseológica para cada término se proporcionan en formato electrónico.
2. Se ofrece información fraseológica tanto para inglés como para español, y el usuario puede encontrar fácilmente las correspondencias entre las unidades fraseológicas en las dos lenguas.
3. El modelo de plantilla con información fraseológica propuesta en esta tesis cuenta con un fundamento teórico. Se basa fundamentalmente en los principios de la Terminología basada en Marcos (§3.1.2.1), toma nociones de la Gramática del Papel y la Referencia (§3.1.2.2), del Modelo Léxico Construccional (§3.1.3.4), así como de FrameNet (§3.1.3.2).
4. La entrada propuesta en esta tesis está orientada tanto para fines de codificación como de descodificación, lo que queda de manifiesto por las varias opciones que se ofrecen en EcoLexicon para acceder a la información fraseológica.
5. No se emplea un metalenguaje complicado. Tal y como se ha mencionando, nociones como las de *rol*, *macrorrol* o *categoría conceptual* se usaron en el

análisis de la información, pero no se codificaron en las plantillas que se despliegan en EcoLexicon.

6. En consonancia con el punto 4, el módulo de información fraseológico permite al usuario buscar y acceder a las colocaciones de varias maneras, tanto a través de la base, como a través del colocativo.
7. No hay sobrecarga de información en la microestructura de cada entrada y además se utilizan diferentes colores en el despliegue de la información para facilitar en la medida de lo posible el acceso ágil y rápido al usuario.
8. Se proporciona una clasificación semántica de las colocaciones basada en su significado. Esto se materializa en la microestructura de las entradas, que clasifica las colocaciones verbales en función del dominio léxico, y seguidamente atendiendo al marco activado.
9. En la definición de cada marco y en el apartado de notas de cada entrada, se ofrece una descripción detallada de los patrones sintácticos y semánticos asociados con el significado de cada colocación.
10. Se ofrecen varios ejemplos de uso para cada colocación verbal, que muestran el verbo en contexto.

El módulo fraseológico diseñado en esta investigación integra los aspectos positivos del análisis de los recursos lexicográficos, terminógraficos y centrados en la descripción del verbo, para proporcionar una representación basada en marcos que está en consonancia con las premisas de la Terminología basada en Marcos. Está configurado de manera que resulte útil tanto para la comprensión como para la producción de textos, y cumple de esta forma con el principal objetivo de esta investigación: el diseño de una plantilla para la codificación y descripción de información fraseológica en EcoLexicon.

Son varias las líneas de investigación que nos gustaría proseguir en un futuro. En primer lugar, tal y como se ha mencionado, el grupo de categorías semánticas establecidas para el subdominio de los RIESGOS NATURALES o EVENTO EXTREMO constituye solo una clasificación preliminar. En este sentido, nuestro objetivo sería redefinir el inventario de categorías propuesto, y establecer al mismo tiempo un inventario de categorías completo para el dominio del MEDIO AMBIENTE. Esta categorización resultaría de inmensa utilidad para la anotación de los argumentos, y en consecuencia, para la automatización del estudio del comportamiento verbal. Asimismo,

esta categorización podría ser de utilidad para la implementación de la incipiente ontología que se está desarrollando en EcoLexicon.

Otro aspecto en el que nos gustaría seguir trabajando es en el establecimiento de correspondencias directas entre las colocaciones para distintas lenguas. Tal y como se ha puesto de manifiesto, al final de este estudio, las correspondencias verbales entre inglés y español se ofrecen en EcoLexicon, pero no de forma directa. El usuario debe, por tanto, acudir al término en la lengua de destino y desplegar sus colocaciones para acceder a la colocación que está buscando. En un futuro próximo la idea es que los marcos de las distintas lenguas estén vinculados para que la correspondencia pueda ser directa.

En consonancia con la idea anterior, otro objetivo sería el de extender la metodología propuesta en esta tesis al resto de lenguas de EcoLexicon y comprobar en qué medida los marcos establecidos funcionan igual de bien para lenguas tan dispares como el ruso y el español, igual que se ha constatado que ocurre para lenguas más próximas como es español y el inglés. Si este fuera el caso, la correspondencia entre marcos se podría generalizar para el conjunto de lenguas disponibles en EcoLexicon, y si no fuera así, se podrían establecer las restricciones pertinentes a las lenguas que sí lexicalizan y activan el mismo tipo de marcos, para poder establecer correspondencias directas entre las colocaciones en las distintas lenguas.

Asimismo, nuestra idea es también extender esta metodología, no sólo al resto de lenguas, sino al resto de subdominios y eventos de EcoLexicon. Esta tarea no creemos que supondrá muchas dificultades, dado que en el espacio privado reservado para el grupo de investigación, disponemos de un apartado en el que el corpus se ha ido introduciendo manualmente y está etiquetado con metaetiquetas que informan, entre otras cosas, sobre las palabras clave o el dominio o dominios que activan los textos. Es por esto por lo que corpus fiables y de calidad centrados en un dominio específico dentro del MEDIO AMBIENTE pueden compilarse de forma automática en fracciones de segundo desde nuestro espacio privado.

Por último, nos gustaría ampliar el perfil de usuarios potenciales que precisan de información fraseológica, y proporcionar diferente información de acuerdo con el perfil de usuario. En este sentido, mientras que para un traductor, nociones como las de *rol*, *macrorrol* o *categoría semántica* pueden resultar opacas, para un usuario con conocimientos en Lingüística más profundos sí le sería de utilidad tener acceso a este tipo de metalenguaje.

0. Introduction

There is general consensus among terminographers, lexicographers, and the users of their resources that phraseological information in terminographic resources is necessary. However, few specialized dictionaries and databases actually include word combinations (L'Homme and Leroyer 2009: 260). Although the situation is gradually improving and the representation of phraseological information in specialized dictionaries is becoming more and more frequent, those specialized dictionaries that do include collocations are not consistent in their method of listing them and representing them in entries (Montero and Buendía 2012). In fact, there is still no general agreement as to which word combinations should be included in dictionaries or how they should be described and classified. In addition, in a large number of terminographic resources that encode some kind of phraseology, the focus is on the description and analysis of noun + noun or noun+ adjective collocations to the exclusion of combinations of other categories of language, such as verbs, despite the fact that verbs are considered to be the most important category of language (L'Homme 1998).

This thesis was developed within the research group Lexicon (http://lexicon.ugr.es). The head of the research group, Professor Pamela Faber, developed a new cognitive approach to the study of Terminology known as *Frame-based Terminology* (FBT). The practical application of FBT is the environmental knowledge base *EcoLexicon* (http://ecolexicon.ugr.es). EcoLexicon represents the conceptual structure of the specialized domain of the Environment in the form of a visual thesaurus in which environmental concepts are configured in semantic networks. The various terminological designations for a concept are provided in six languages: Spanish, English, German, French, Russian, and modern Greek. This resource currently contains more than 3,500 concepts and more than 18,500 terms. For each concept it gives the definition of the concept, graphical resources that illustrate the concept, as well as the whole set of terms that designate the concept in the various languages. In addition, for each term, EcoLexicon provides information concerning its grammatical category (noun, noun phrase, verb, adjective or adverb); term type (main term, synonym, geographical variant, and acronym); gender (masculine, feminine, and neuter); and contexts of use. However, despite the fact that EcoLexicon provides a great quantity of linguistic information for each specialized knowledge unit, it does not provide any phraseological information. This is a problem since phraseology is a

necessary part of linguistic knowledge, and is useful for a wide range of users, especially for translators. In fact, the results of a questionnaire completed by translation students at the University of Granada (López, Buendía, and García 2012) (§1.5) reflected that phraseology along with usage contexts were considered to be the most useful information for text production in the translation process.

Accordingly, this research proposes the inclusion of a phraseology module in EcoLexicon, which provides collocational information for each term. For this purpose, we decided to begin with verb phrases associated to terms since recent studies have highlighted the importance of verbs in specialized texts and terminology (Lorente 2007; López 2007; Buendía 2012), and there are currently few terminographic resources that incorporate them. To this end, within the general domain of the environment, we focused our study on the subdomain of NATURAL HAZARDS also referred to as EXTREME EVENT, in consonance with the conception of event in Frame-based Terminology (§1.4.3). The reason lies in the growing interest in this subdomain since the Earth is suffering an increasing number of disasters, both natural and human-induced. This has made the environment a vital area of study for people all over the world.

0.1 Hypothesis

This research is based on the following hypothesis:

- There is a need to include phraseological information in terminographic resources basing the design of phraseological entries on premises from cognitively-oriented Terminology theories and on meaning-based linguistic theories. This signifies that the representation of phraseological information can be designed as a network of interlinking frames that in turn, stem from basic conceptual categories, such as MOVEMENT, CHANGE, EXISTENCE, etc. It is thus possible to establish templates that generalize phraseological meaning for entire semantic domains.

0.2 Objectives

The general objective of this study was the following:

- To design a phraseological template for encoding and describing phraseological information in EcoLexicon.

To this end, the following operational objectives were established:

- To define the user profile and the communicative situations to which templates can be applied.
- To establish the theoretical foundation for the phraseological template proposed in this thesis.
- To formulate the guidelines for an ideal terminological entry in a specialized dictionary.
- To compile and analyze a corpus of texts on the environment, and more specifically on NATURAL HAZARDS by combining methodologies of the *web for corpus (WfC)* and of the *web as corpus (WaC).*
- To design a protocol for the evaluation of online resources which assures the reliability and quality of the texts in our corpus.
- To define a set of semantic categories and semantic roles relevant for the EXTREME EVENT frame.
- To establish the subframes activated in the EXTREME EVENT so as to link each frame to a template that encodes the requirements and restrictions of the frame.

0.3 Outline of the thesis

The design of a template for encoding and describing noun + verb and verb + noun collocations in EcoLexicon required on the one hand, an overview of the theories of Terminology (§1) as well as a description and discussion of the linguistic theories dealing with predicate argument structure (§3). This analysis enabled us to lay the theoretical foundations for our verb analysis. On the other hand, an analysis of the most representative terminographic resources with phraseological information (§2), along with the description of the most important meaning-based resources for verbs (§3.2) was also considered necessary. Accordingly, our verb template is based on the conclusions drawn from this analysis of the positive and negative aspects of terminographic and verb meaning-based resources.

In this regard, the first chapter of this thesis outlines the cognitive theories of Terminology that are the foundation of this work. It includes a brief introduction that discusses the differences between Lexicology and Lexicography and Terminology and

Terminography (§1.2). This is followed by a summary of the Function Theory of Lexicography (§1.3). Section 1.4 discusses the cognitive theories of Terminology upon which this thesis is based, namely, Sociocognitive Terminology Theory (§1.4.2) and, more specifically, Frame-based Terminology (§1.4.3). In addition, a large part of the chapter is devoted to the explanation of the information contained in EcoLexicon, and how it can be enhanced with phraseological information. Finally, we explain the usefulness of EcoLexicon for translators, and describe the needs of the potential users, which have been taken into account for the design of the phraseological template, based on the Function Theory of Lexicography in mind (§1.5).

Chapter two focuses on the description of phraseology in Lexicography and Terminography. First, a review of phraseology and phraseological unit is given (§2.1.2), as well as a survey of the various approaches to the study of collocations (§2.1.3): (i) *the semantically-based approach*, which assumes a particular meaning relationship between the constituents of a collocations (§2.1.3.1); (ii) *the frequency-oriented approach*, in which collocations are evaluated on the basis of how often they appear in texts (§2.1.3.2). Within the semantic approach, a significant part of the section is devoted to Mel'čuk and his Meaning-Text Theory and the Explanatory Combinatorial Dictionary since it has been and still is the most influential combinatorial dictionary for any language. Section 2.1.4 describes our approach to the concept of collocation. A distinction between collocations and the rest of word combinations, such as free phrase, idioms or compounds is also provided. Section 2.2 describes how English and Spanish lexicographic resources encode collocations in their entries. Dictionaries are thus categorized as follows: (i) monolingual general dictionaries; (ii) bilingual general dictionaries; (iii) collocations or combinatory dictionaries. Section 2.3 does the same for terminographical resources. Finally based on the positive and negative aspects of the lexicographic and terminographic resources analyzed in this chapter, a set of guidelines for an ideal phraseological entry in a specialized dictionary is specified in 2.4.

Chapter three describes the linguistic theories that can be applied to Terminology, and more specifically, those whose premises were particularly useful for our study. As such, our study incorporates insights from the Lexical Grammar Model (§3.1.2.1), Role and Reference Grammar (§3.1.2.2), the Lexical Constructional Model (§3.1.3.4), as well as certain premises from FrameNet (§3.1.3.2). In addition, a description of the most well known lexical resources with verb information is included, i.e. the English resources, WordNet (§3.2.1), VerbNet (§3.2.2), PropBank (§3.2.3) and the Spanish

resources, ADESSE (§3.2.4) and SenSem (§3.2.5). We explain in detail which aspects of these resources were interesting for our verb entry proposal.

Chapter four is the chapter devoted to the materials and methods used to carry out our study. It first specifies the boundaries and the scope of our study within the subdomain of the EXTREME EVENT (§4.1). Subsequently, it explains the design, compilation, and description of the corpus used to analyze the data. To that end, a brief revision of the concept of corpus is provided (§4.2.1.1) as well as a succinct description of the typology of corpora (§4.2.1.2). In addition, the methods and criteria for corpus compilation are described (§4.2.2) since corpus compilation and design significantly influenced and determined the results of our research analysis. In this regard, we designed a protocol for the evaluation of online resources which assures the reliability and quality of the texts in our corpus (§4.2.2.1.2), and which can be applied to any type of corpus.

The two methods of corpus analysis were also described: (i) *web for corpus (WfC)*, in which the web is used as a source of texts in digital format for the subsequent implementation of an offline corpus; (ii) *web as corpus (WaC)*, which uses the web directly as a corpus. Finally, section 4.3 describes the software application used, namely, the term extractor, TermoStat, and the corpus analysis tool, WordSmithTools. TermoStat was mainly used to select the specialized verbs activated within the EXTREME EVENT and, thus, the candidate verbs to be analyzed in this thesis. Regarding WordSmithTools, the *Concordance* option was extremely useful to analyze the behavior of the verbs selected for this study.

Chapter five describes how the analysis of data was performed, as well as the results obtained with the analysis. As shall be seen, arguments as well as predicates were the focus of our analysis. The results of this analysis are encoded in templates (§5.4.3), which helped to subsequently implement the phraseological module in EcoLexicon (§5.5). Finally, the implementation of this module is explained so that the reader can visualize the results of this study. We describe the macrostructure, namely, the ways of accessing collocations in EcoLexicon, and also give a detailed description of the microstructure of a phraseological entry within EcoLexicon. Section 5.5.4 provides evidence that the template proposed in this thesis for encoding verb phraseological information satisfies the criteria listed in 2.4 for the design of an ideal phraseological entry in a specialized resource.

Finally, chapter six lists the conclusions that can be derived from this thesis and discusses future lines of research to be pursued.

1. Cognitive-based Theories of Terminology

1.1 Introduction

The first chapter of this thesis presents the theoretical framework and methodological context of the research carried out. As such, it outlines the cognitive theories of Terminology that are the foundation of this work[1]. The first section of the chapter is a brief introduction that differentiates Lexicology from Lexicography, and Terminology from Terminography (§1.2). This is followed by a summary of the Function Theory of Lexicography (§1.3), which quite aptly states that no resource can be designed without taking the specific information needs of the targeted user group into account. Section 1.4 discusses the cognitive theories of Terminology upon which this thesis is based namely, Sociocognitive Terminology Theory (§1.4.2) and, more specifically, Frame-based Terminology (§1.4.3). Even though an in-depth analysis of Terminology theories from Wüster to the present is not provided, a general overview of their main assumptions is given in §1.4.1.

Since this study is largely based on the principles of Frame-based Terminology (§1.4.3), a significant portion of this chapter explains the basic tenets of this model and describes EcoLexicon, an environmental knowledge base that is the practical application of FBT (§1.4.3.2). As shall be seen, we explain how knowledge is extracted for the concept TROPICAL CYCLONE (§1.4.3.1) and also how the terminological entry for HURRICANE is designed (§1.4.3.2). This justifies the need to enhance EcoLexicon with phraseological information. Finally, we explain the usefulness of EcoLexicon for translators, and we describe the needs of the potential users, which have been taken into account for the design of the phraseological entry (§1.5).

1.2 Distinction between Lexicology/Lexicography and Terminology/Terminography

To differentiate Terminology from Lexicology, most authors begin by establishing a correspondence between Lexicology and Lexicography, on the one hand, and Terminology and Terminography, on the other. For example, Pérez (2002: ch.3.3) states that Lexicology is the discipline that studies and describes the lexicon of a language,

[1] Linguistic theories applied to Terminology, and more specifically to this study, are explained in chapter 3.

whereas Lexicography is Lexicology applied to the compilation of general language dictionaries. In the same way, Terminology is concerned with the theoretical and methodological description of specialized language, whereas Terminography is Terminology applied to the elaboration of specialized dictionaries.

Although this distinction between the theoretical and the practical is very well established for Lexicology and Lexicography, this is not the case for Terminology and Terminography. As is well known, *terminology* can refer to the following (Nkwenti-Azeh 1998: 137): (i) a theory that explains the relationship between concepts and terms; (ii) the job of compiling, describing and presenting terms (i.e. in this sense, it is a synonym of *terminography*); (iii) the vocabulary of a specialized domain. In this sense, in order to differentiate between the various meanings of *terminology*, when it refers to the theory that studies specialized language, the word *terminology* begins with a capital *T* (Temmerman 2000a: xiii). However, when it refers to its other senses, it begins with a lower case *t* (i.e. 'terminology'). This means that quite often, the words *terminology* and *terminography* are indiscriminately used when referring to the practical application of Terminology (Montero, Faber, and Buendía 2011: 21)[2]. One possible reason is that *terminography* was not officially adopted by the International Organization for Standardization until 1975 to refer to that part of Terminology in charge of the production of specialized dictionaries.

However, the boundary between Terminography and Lexicography is far from being clear-cut. Since specialized dictionaries are "dictionaries that treat specialized fields of knowledge (e.g. business, chemistry, law)" (Bowker 2010: 156), thus focusing on specialized language, many authors refer to terminography as *specialized lexicography* (Bergenholtz and Tarp 2010; Montero 2002, *inter alia*). As Bergenholtz and Tarp (2010: 27) point out: "The dividing line between specialized lexicography and terminography is nonexistent" (Bergenholtz and Tarp 2010: 29). We also agree that specialized lexicography and terminography basically refer to the same type of activity. However, throughout this research, we will use the term *terminography*.

Since the difference between lexicography and terminography largely resides in the nature of the language that each deals with, it is necessary to differentiate between general and specialized language. In the words of Cabré (1999a: 59):

[2] In this thesis, we always use the dichotomy 'lexicographic/terminographic resource', but, as shall be seen, we refer to EcoLexicon as a 'terminological knowledge base', because it contains terms.

> The set of rules, units and restrictions that form part of the knowledge of most speakers of a language constitutes the common or general language. The units of the general language are used in situations we call 'unmarked'. In contrast, we speak of special or specialized languages to refer to a set of subcodes (that partially overlap with the subcodes of the general language), each of which can be 'specifically' characterized by certain particulars such as subject field, type of interlocutors, situation, speakers' intentions, the context in which a communicative exchange occurs, the type of exchange, etc. Situations in which special languages are used can be considered as "marked".

More specifically, for language to be considered *specialized language*, it should meet the needs of specialized user groups either because of its topic, level of expertise, or field of use. It should consist of specialized knowledge units with interrelated characteristics, which are activated in discourse with a predominantly communicative function (Cabré 1993: 135):

> Una definición de lenguaje de especialidad: a) se trata de conjuntos «especializados», ya sea por la temática, la experiencia, el ámbito de uso o los usuarios; b) se presenta como un conjunto con características interrelacionadas, no como fenómenos aislados; c) mantienen la función comunicativa como predominante, por encima de otras funciones complementarias.

Bergenholtz and Tarp (2010: 28–29) underline the difference between lexicographers who deal with general language and terminographers who deal with specialized language:

> (i) Terminographers make tools for experts for text production, lexicographers for laymen for text reception, and (ii) Lexicographers make dictionaries with an alphabetical macrostructure, terminographers use a systematic one.

In reference to the first part of the above statement, these authors also affirm that the real difference between Terminography and Lexicography is not that clear since lexicographers normally create entries conceived for multifunctional dictionaries, rather than taking user needs as their starting point (Bergenholtz and Tarp 2002, 2003, 2004, 2010; Tarp 2008). As for the second part, it is valid for many paper general language dictionaries, but not for all of them. Nor they are certain that most terminographic work has resulted in tools with a systematic macrostructure, as shall be seen in section 2.3.

1.3 Lexicology: the Function Theory of Lexicography

Over the last 15 years, the Function Theory of Lexicography (Bergenholtz and Tarp 2002, 2003, 2004, 2010; Tarp 2008) has been developed at the Centre for Lexicography at the Aarhus School of Business (Denmark). According to this theory, lexicographic resources should be designed to meet the specific information needs of certain user groups. As Bergenholtz and Tarp (2010: 29) state:

> The needs, by definition, are related not only to a specific type of user, but also to the specific type of social situation where this type of user may have a specific type of lexicographically relevant need that may lead to dictionary consultation.

Accordingly, a *lexicographic function* can be defined as the satisfaction of the specific types of lexicographically relevant needs that may arise in a specific type of potential user, in a specific type of extra-lexicographical situation (Tarp 2008: 81). This definition includes four basic elements:

- *satisfaction*, i.e. the assistance the dictionary can provide for a potential user, including easy and quick access to data;
- *specific types of lexicographically relevant needs*, i.e. what a dictionary can be used for;
- *a specific type of potential user*, i.e. *who* may benefit from using a dictionary;
- *a specific type of extra-lexicographical situation*, i.e. *when* a dictionary can be used.

This tendency towards user needs can also be extended to terminographic products (§2.3). In this sense, Nielsen very aptly describes the functionality of specialised translation dictionaries for learners (Nielsen 2010: 69):

> Dictionaries are utility products that are designed to help specific types of users in specific types of situations to solve specific types of problems. This means that the type of dictionary that is relevant in this context is one whose function is to help learners solve specific types of problems encountered when translating subject-field specific texts into a foreign language.

Therefore, in order to be able to talk about the user needs of any lexicographical or terminographic resource, it becomes necessary to specify the types of user situation

as well as the types of user. Experience has shown that the type of situation (§1.3.1) is more important than the type of user (§1.3.2) when determining needs (Tarp 2006). As such, the criteria for establishing the user typology are to some extent determined by the types of user situation. As such, the starting point in determining the functions of a specific dictionary is first an analysis of the various types of situations where lexicographic needs may arise, and subsequently, the types of users.

1.3.1 Types of user situation

As previously mentioned, user situations are the most important elements of lexicographic functions. In the words of Bergenholtz and Tarp (2010: 30):

> No user has specific needs unless they are related to a specific type of situation. Consequently, it is not enough to define which types of users have which needs, but also the types of social situations in which these needs may arise should be defined. However, not all such situations are relevant for lexicography; only situations in which needs may arise that can be satisfied by consulting dictionaries.

The situations where lexicographical and terminographic resources provide assistance to users can be divided into *cognitive* (§1.3.1.1) and *communicative* (§1.3.1.2) situations, although a third type, referred to as *operational*, has also been proposed (§1.3.1.3) (c.f. Tarp 2008b).

1.3.1.1 Cognitive situations

Cognitive functions are motivated by the need to acquire new knowledge or verify existing knowledge about a certain topic or a language for specific purposes (LSP). In cognitive user situations, lexicographically and terminographically relevant needs are thus needs of encyclopaedic knowledge related to language, specialised language, culture or any specific subject field (L'Homme and Leroyer 2009: 269). Based on Tarp (2007), cognitive-oriented user situations can be said to include the following:

- the acquisition of information about a special subject field in the native language and/or in a foreign language;
- the acquisition of information about the native LSP and/or the foreign LSP;
- the comparison between the special subject field in the native and foreign culture; and between the native and foreign LSP.

1.3.1.2 Communicative situations

Communicative functions and communicative situations are determined by the need to obtain dictionary assistance when the user is engaged in some textual activity, such as reading or revising a text, translating a source text into a target language or writing a text in the mother tongue or in a foreign language (L'Homme and Leroyer 2009: 270). These situations include the following (Bergenholtz and Tarp 2010: 31):

- the production of text in the mother tongue;
- the reception of text in the mother tongue;
- the production of text in a foreign language;
- the reception of text in a foreign language;
- the translation of text from the mother tongue into a foreign language;
- the translation of text from a foreign language into the mother tongue;
- the translation of text from one foreign language into another.

Communication can be considered to be the mediating element by which the information retrieved from the data contained in dictionaries can be transformed into language knowledge (Tarp 2010: 47). Figure 1 shows this possible relationship between a dictionary consultation and language skills, where communication works as a linking element (Tarp 2010: 48):

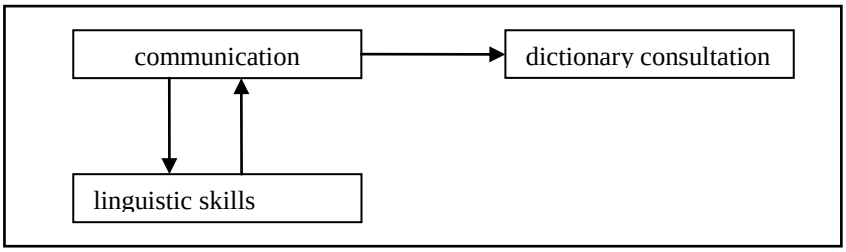


Figure 1. Relationship between dictionary consultation, communication, and linguistic skills

1.3.1.3 Operational situations

Operational situations refer to practice, i.e. practical exercises and training, whose nature should be determined for each particular subject field (Tarp 2010: 49). As Tarp (2010: 49) underlines, the learning of practical skills can be done in one of two ways: (i) directly, i.e. by providing information that can be used to clarify doubts related to practical exercises and the training process in terms of observing, evaluating, and interpreting the situation and taking the corresponding operative action; (ii) indirectly,

by transmitting knowledge along with a systematic study of the subject field in question, a knowledge that could subsequently be used in practical exercises and training. Accordingly, the knowledge and information extracted either directly or indirectly from lexicographic products and used to assist the practical exercises and training process can be assimilated as improved practical skills, as shown in Figure 2 (Tarp 2010: 49):

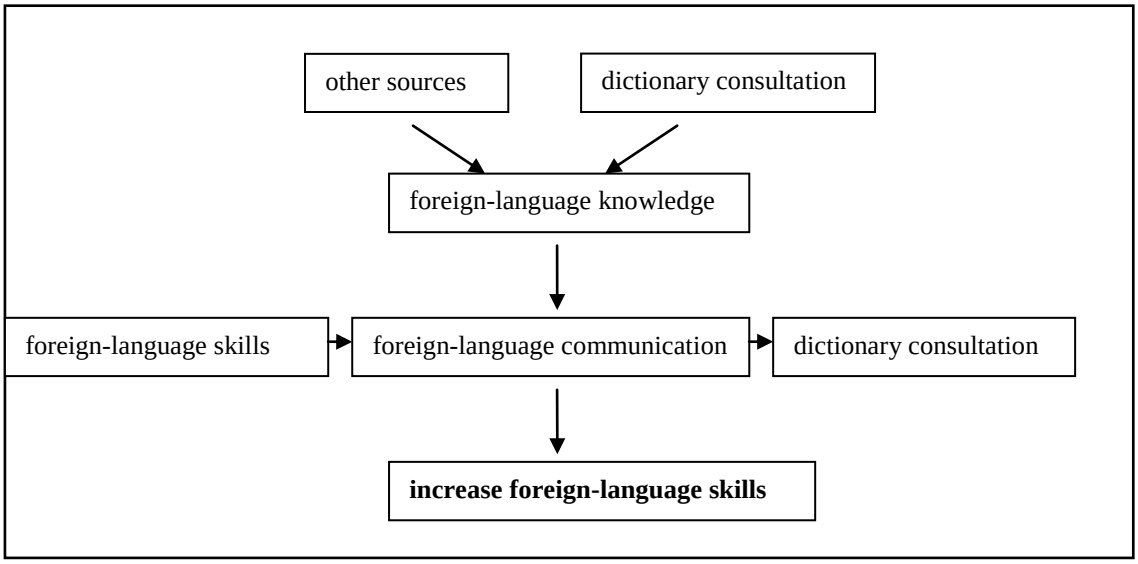


Figure 2. The learning of practical skills

Tarp (2008b: 126) mentions reference works, such as handbooks, manuals, and how-to guides, which are designed to give directions and instructions on how to proceed in specific situations. These works thus can be said to have an operational function. Despite the fact that there is currently no dictionary geared to assist the user in operational situations, he is very hopeful that the situation will change in the near future (Tarp 2010: 50). As shall be seen in 1.5, in EcoLexicon, operational situations are not taken into account.

1.3.1.4 Summary of user situations

Succinctly put, the user situations in the context of a lexicographic or terminographic resource are of two types: (i) the need to acquire knowledge (cognitive situation); (ii) the need for skills, which can be either linguistic (communicative situation) or practical (operational situation). As shown in Figure 3 (Tarp 2010: 44), the acquisition of knowledge can take place as a systematic or as a sporadic process. Communicative or

linguistic skills refer to the learning of abilities that enable users to engage in the various phases of the communication process, the most important of which are text reception, text production, and translation. Finally, practical skills refer to the learning of manual or intellectual skills, i.e. brick-laying, the operation of a machine, swimming, etc. All of these skills have two separate but interdependent components, i.e. the ability to interpret the situation by means of observation (interpretative skills) and the ability to take action when necessary (operative skills):

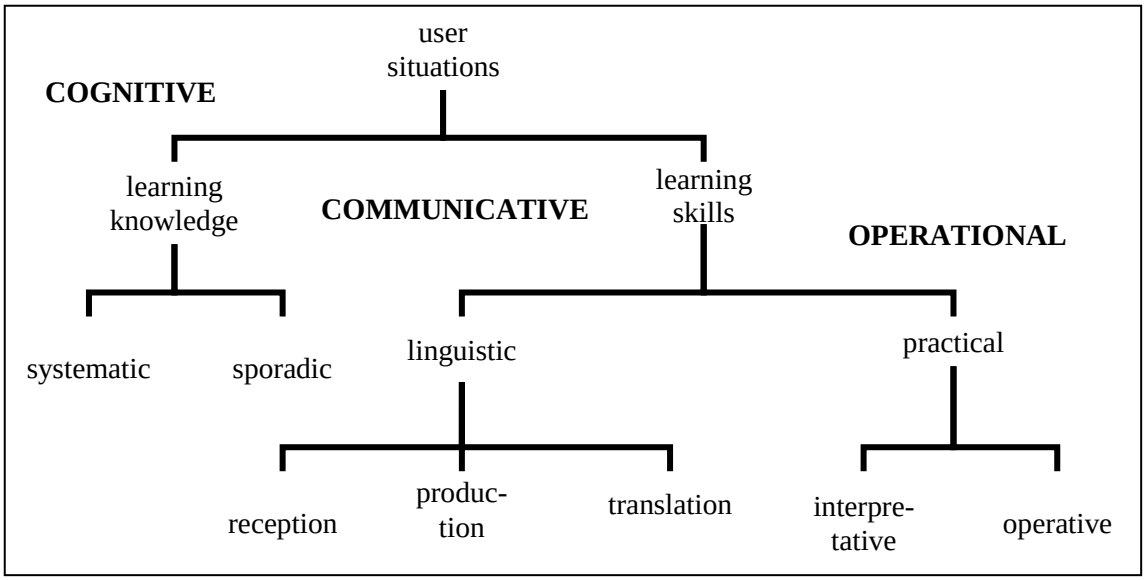


Figure 3. Summary of the user needs that dictionaries should satisfy. Adapted from Tarp (2010: 44)

1.3.2 Type of users

When designing a profile of potential dictionary users, a number of criteria should be taken into account (Bergenholtz and Tarp 2010: 31):

- What is the mother tongue of the users?
- To what extent do they master their mother tongue?
- To what extent do they master a specific foreign language?
- To what extent do they master a specific LSP in their mother tongue?
- To what extent do they master a specific LSP in a foreign language?
- How much experience in translation do they have?
- How great is their general cultural knowledge?
- How great is their knowledge of culture in a specific foreign language area?
- How much do they know about a specific subject or science?

By combining the user typology resulting from the answers to these questions with the user situations described above (§1.3.1), it is possible to determine the corresponding information needs of a certain lexicographic or terminographic resource. In other words, on the basis of these nine criteria, it is possible to establish a user typology related to a specific lexicographic/terminographic resource.

1.4 Terminology: Cognitive-based Theories of Terminology

1.4.1 From Wüster to the cognitive-based Theories of Terminology: brief summary

Terminology as a discipline of study is relatively new. As Cabré (2000: 37) points out, Terminology emerged from the need of specialists to unify concepts and terms of their specialized fields to enhance communication and the transfer of knowledge. Its starting point can be traced back to the 1930s to Eugen Wüster, who is considered the father of Terminology and the founder of the first theoretical model in Terminology: the General Terminology Theory (GTT).

Wüster was the author of *The Machine Tool, an Interlingual Dictionary of Basic Concepts* (Wüster 1968), a French-English dictionary of standardized terms with a German supplement. This dictionary laid the foundations for the GTT (Wüster [1979] 1998), a prescriptive approach, whose theoretical and practical premises were developed by Wüster's disciples of the School of Vienna, and materialized in the publication in 1979 of *Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie* [Introduction to the General Terminology Theory and Terminographic Lexicography].

The GTT's main purposes were the following (Cabré 2003: 173): (i) to eliminate ambiguity in technical language standardizing its terminology with a view to transforming specialized languages into efficient tools of communication; (ii) to convince specialized language users of the benefits of standardized terminology; (iii) to establish Terminology as a discipline and transform it into a science.

Given Wüster's main objects, it is hardly surprising that the GTT put a great deal of effort into differentiating specialized language from general language, and by extension, terms from words. In this regard, the GTT states that if in general language there is polysemy and synonymy, in specialized language, univocity and mono-

referentiality prevail (Montero, Faber, and Buendía 2011: 38). As it is well known, *univocity* means that "one concept is referred to by one term (no synonymy) and one term can only refer to one concept (no polysemy)" (Temmerman 1997: 54–55), whereas *monoreferentiality* signifies that one term only designates one concept (Cabré 1993: 213). In other words, the GTT initially proposed an idealized vision of specialized communication based on the notions of univocity and monoreferentiality that silenced variation and diversity (Cabré 1999b: 105). In the words of Faber (2009: 112):

> Even the most cursory examination of specialized language texts shows that terminological variation is quite frequent, and that such variation seems to stem from parameters of specialized communication, such as the knowledge and prestige of the speakers, text function, text content, etc. The same concept can often be designated by more than one term, and the same linguistic form can be used to refer to more than one concept. Furthermore, terms have distinctive syntactic projections, and can behave differently in texts, depending on their conceptual focus.

Today, few people would deny that synonymy and polysemy are also present in specialized language. Therefore, the premises of univocity and monorefentiality means that the GTT is unable to account for the real use of specialized language, especially in activities such as Translation.

In addition, the GTT conceived Terminology as exclusively synchronic and within its framework, concepts were the starting point in the description and organization of terminological information. In this sense, a concept was conceived as an abstract cognitive entity representing an object, and a term was merely its linguistic label, and concepts were regarded as being independent from terms (Felber 1984: 103):

> The concept is defined as an element of thinking […]. It is a mental construct representing a material or immaterial individual object. The concept consists of an aggregate of characteristics which we can cognize as being common to a number of individual objects, and which we use as means for mental ordering and for communication. […] The concept […] exists independently of the term, the meaning of which it is. A term is assigned deliberately to a concept after due consideration whether this term corresponds to the concept in question.

Accordingly, since Terminology was regarded as mainly denominative, not enough attention was paid to the communicative function of language. Syntax and pragmatics were not considered at all (Faber 2009: 112):

> Since Wüster believed that the function of Terminology was to create and standardize names for concepts, syntax was not regarded as falling within the scope of Terminology.

As shall be seen, subsequent theoretical approaches that emerged as a reaction to the GTT did take into account the communicative function of Terminology and analyzed terms as they actually behaved in real discourse.

Another basic premise of the GTT was its rather rigid notion of conceptualization. It only considered hierarchical relations, i.e. *is_a* and *part_of* relations, and did not describe non-hierarchical ones. This made it impossible to account for the dynamicity of terms (Faber 2011; Buendía 2012, *inter alia*) and the multidimensionality of concepts (Bowker 1997; Kageura 1997; Meyer, Eck, and Skuce 1997; Rogers 2004; León 2009).

The importance of the GTT lies in the fact that it transformed Terminology into a discipline and paved the way for further research. However, as previously mentioned, there were many aspects of specialized language that the GTT could not explain. In the 1990s, new theories emerged as a reaction to the GTT. These theories were descriptive instead of prescriptive and were social, communicative, and cognitive-based approaches to the study of terms and specialized language.

Centered on a more social and communicative approach to Terminology, the most representative theories are Socioterminology (Gaudin 1993, 2003; Boulanger 1991, 1995; Gambier 1991, 1993) and the Communicative Theory of Terminology (Cabré 1993, 1999, 2001a, 2001b); and from a more cognitive-based perspective of Terminology, Sociocognitive Terminology (§1.4.2), and Frame-based Terminology (§1.4.3). As previously mentioned in the introduction, this research adopts a cognitive-based approach to Terminology, mainly based on the premises of Frame-based Terminology.

Since an analysis of the various approaches to Terminology from Wüster onwards can be found in recent works (León 2009; Fernández 2011; Sanz 2011, *inter alia*), it is not relevant to describe these approaches in great detail. However, an overview of certain theoretical models is necessary in order to understand the context that led to Frame-based Terminology and the reason why we selected it for our research.

Socioterminology (Gaudin 1993, 2003; Boulanger 1991, 1995; Gambier 1991, 1993) applied sociolinguistic premises to Terminology. In reaction to the GTT, it favored a more descriptive and diachronic study of language, and did not cut Terminology off from Linguistics (Boulanger 1995: 198). In addition, Socioterminology affirmed the existence of terminological variation, and consequently, of polysemy and synonymy in specialized language. This was amply proven by studying terms in real usage contexts (Boulanger 1995: 204–205):

> La socioterminologie fait surgir de l'ombre le concept d'«usage» et le ramène dans l'environnement du terme. Qui dit usage dit aussi somme de discours dans lesquels s'enchâssent les unites lexicales que des interactions spécfiques agitent, comme des atomes, pour reveler la vraie nature de la terminologie

The Communicative Theory of Terminology (CTT) (Cabré 1993, 1999, 2001a, 2001b) is a communicative theory that describes the complexity of terms (or specialized knowledge units) in real situational contexts (Cabré 2003: 164). To this end, the Theory of the Doors, or metaphor of the polyhedron, is used to describe the various ways of accessing terminological units. Accordingly, Cabré compares a specialized knowledge unit to a polyhedron with three dimensions (cognitive, linguistic and communicative), through which it can be accessed. This is the basis for the multidimensionality of such units.

In this sense, the cognitive dimension is the door for the description of concepts and their hierarchical and non-hierarchical relations. The linguistic dimension describes a specialized knowledge unit as it is encoded in language. Finally, the communicative dimension describes how the unit is activated in different communicative situations. As Cabré (2003: 188) underlines, the CTT approaches specialized knowledge units through the linguistic door:

> To be specific we approach the units through the door of language and thus we intend to account for them from the viewpoint of a theory of natural language.

However, it is important to emphasize here that the choice of one door does not mean the exclusion of the other two, which only recede into the background (Cabré 1993: 187).

In consonance with this, and in contrast to the GTT, a basic premise within the CTT is the existence of denominative variation in specialized languages (Freixa 2002, 2006, *inter alia*), defined as follows (Freixa 2006: 51):

> Denominative variation can be defined as the phenomenon in which one and the same concept has different denominations; this is not just any formal variation (variation between a term and a periphrasis, or a definition, for example), but is restricted to variation among different denominations, i.e., lexicalized forms, with a minimum of stability and consensus among the users of units in a specialized domain.

Unlike the GTT, which established clear-cut boundaries between words and terms, the CTT conceives terminological units as general linguistic units by default, and states that it is the context that gives a particular unit the status of *terminological* or *specialized*. In the words of Cabré (2003: 189–190):

> […] a lexical unit is by itself neither terminological nor general but that it is general by default and acquires special or terminological meaning when this is activated by the pragmatic characteristics of the discourse. […] Any lexical unit would thus have the potential of being a terminological unit. […] The condition of 'terminological unit' applied to a lexical unit does not exist prior to its usage in a specific communicative context.

The CTT has made an enormous contribution to the study of Terminology and has afforded valuable insights into specialized language and discourse. However, though recognizing its incalculable contributions to Terminology, Faber (2009: 115–116) lists the following negative aspects of the CTT: (i) it is not based on any linguistic theoretical model; (ii) although its semantics are said to be based on conceptual representation, it does not explain how conceptual representations are created, what they look like, and their main constraints would be; (iii) although it specifies that a lexical unit acquires specialized meaning in discourse, it does not offer a clear explanation of specialized meaning and its components.

More recently, in consonance with recent trends in Linguistics, new cognitive-based theories of Terminology have emerged, which are more focused on meaning and conceptual structure. This has led a growing number of authors to affirm that a *cognitive shift* is currently taking place in Linguistics and Terminology (Evans and Green 2006; Faber 2009, 2011, *inter alia*):

> […] Terminology is essentially a linguistic and cognitive activity. In this sense, terms are linguistic units which convey conceptual meaning within the framework of specialized knowledge texts. In the understanding of the nature of terms, this process of meaning transmission is as important as the concept or concepts that they designate. Terminological units are thus subject to linguistic analysis. Since this type of analysis can be carried out in a number of ways, it is necessary to choose the linguistic approach most in consonance with the object of study. Such an approach should be lexically-centered and usage-based. It should also have its primary focus on meaning and conceptual representation. […] such is the case of theoretical approaches based on Cognitive Linguistics (Faber and López 2012: 11).

As shall be seen, cognitive-based terminological theories share some of the premises of CTT and Socioterminology. For example, they are descriptive rather than prescriptive, and base their analysis on the term (i.e. semasiological approach), instead of the concept (onomasiological approach). However, what makes cognitive-based terminological theories different from the CTT and Socioterminology is the importance that they confer to categorization and conceptual organization. Accordingly, they incorporate premises from Cognitive Linguistics (§3.1.3) and Psychology in order to describe concepts and category structure. The most representative contributions within this framework are Sociocognitive Terminology Theory (§1.4.2) and Frame-based Terminology (§1.4.3).

1.4.2 Sociocognitive Terminology Theory

Temmerman (2000, 2001) developed a theory of descriptive terminology, referred to as *Sociocognitive Terminology Theory* (STT) based on sociocognitive principles as a reaction to the limitations of the GTT. This theory emerged thanks to the results obtained in the study of the terminology of the life sciences (Temmerman 1997, 1998a, 1998b, 2000a, 2000b), which provided evidence that for scientific disciplines, standardization was of minor importance, and that the premises of GTT were not applicable to real terminological work. The following paragraphs summarize Temmerman (2000a: 219–237) who compares the STT to the GTT.

First of all, the GTT conceived the concept to be the focus of study in Terminology. Terminological work was thus encouraged to start from the concept (onomasiological approach). Temmerman underlined that the traditional notion of concept as a "unit of thought constituted through abstraction on the basis of properties

of a set of one or more objects" (ISO/CD 1087-1 1995 as quoted in Temmerman 2000a: 223) was too restrictive and that few concepts really exist objectively. In contrast, instead of concepts, the STT prefers to talk of *units of understanding* (UoU), most of which have a prototypical structure. In fact, those UoU that are prototypically structured are referred to as *categories*. Temmerman reserves *concept* for those UoUs which are not prototypical, and therefore could be described in terms of the GTT. Figure 4 summarizes the idea of UoU as outlined by the STT (Geentjens et al. 2006: 144):

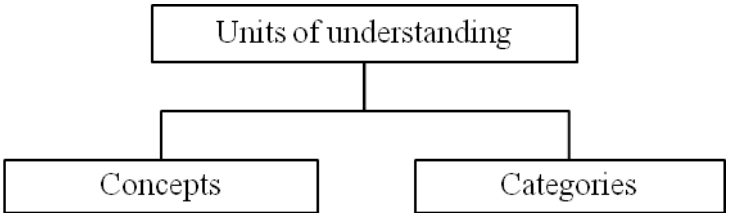


Figure 4. Types of units of understanding in the STT

In the STT, the representation of relations between UoUs are in the form of idealized cognitive models (ICMs) of the type proposed by Cognitive Linguistics (Lakoff 1987) (§3.1.3). In the words of Lakoff (1987: 68):

> […] we organize our knowledge by means of structures called *idealized cognitive models,* or ICMs, and that category structures and prototype effects are by-products of that organization.

As Ruiz de Mendoza (1999: 9) points out, although Lakoff has not provided a precise definition of what an ICM is, it can be understood as a conventional conceptual representation of how we perceive and organize reality. In this sense, ICMs do not exist objectively in nature, but are created by human beings. They are *idealized* because they are the result of a particular kind of regularity extracted from the characteristics of many common experiences; *cognitive* since they are construed in the mind; and *models* since, although not being real, they try to represent reality. This model of categorization is based on Rosch's Prototype Theory (§3.1.3) (Rosch 1973a, 1973b; Rosch and Lloyd 1978) which asserts that in order to understand the world, humans organize entities into a set of categories with a prototype which represents the features that are the most typical of the category member. Consequently, the entities within a category are defined based on the differences and similarities they present *vis-a-vis* the prototype.

According to Temmerman, units of understanding exist because of texts that express ways of understanding certain categories within certain ICMs, which can differ from those of other authors and from the interpretations of readers. Accordingly, she believes that the term, not the concept, should be the starting point for terminological description, in line with the other approaches that emerged in reaction to the GTT (Temmerman 2000a: 224):

> As terminology can only be studied in discourse it is better to accept that it is the **term** which is the starting point in terminological description rather than what was traditionally called the *concept*.

The second main assumption of the STT is that categories are the ideal structure to describe UoUs, most of which do not have clear-cut boundaries and thus, cannot be clearly defined. As Temmerman points out "each category is understood as existing within cognitive models" (Temmerman 2000a: 225). As such, in line with the first assumption, in the STT, a category's existence is dependent on the language. Succinctly put, in contrast to the meaning approach followed by the GTT, the STT focuses on an understanding approach by means of categorization.

The third basic assumption regarding STT is that, unlike the GTT which supported that terminographic definition should be intensional and reflect the position of the concept within the conceptual system, Temmerman claims that the definition of a UoU is not static, but rather depends on various parameters, such as the type of UoU being described, the knowledge level of the text sender and receiver, and the profile of the termbase user (Temmerman and Kerremans 2003). In this sense, one can say that they consider term definitions to be dynamic.

In line with the other approaches, the STT also believes in the existence of synonymy, polysemy, and figurative language in specialized texts. In this regard, the analysis of the field of life sciences in Temmerman (1997, 1998a, 1998b) provides evidence that, although there are some concepts with clear-cut boundaries, the vast majority have a prototype structure, as reflected in the synonymy, polysemy, and figurative language that were an integral part of their naming history (i.e. 'blotting' and 'biotechnology') (Temmerman 1997: 88):

> We realized that it is possible to distinguish between two types of concepts: those which are fit for a description in accordance with the univocity principle of traditional Terminology and those which are not. Concepts of the first type appeared to be clear-cut and not to show characteristics of prototype structure. For categories which do show prototype structure the principle of univocity is impossible to apply. The reason is that the principle is not at all in accordance with the facts of conceptualisation, categorization and naming in LSP. For several of the categories we studied, the principle of univocity appeared to be useless as polysemy, synonymy and figurative language showed to have their function in the coming into existence and the further development of these categories.

Finally, unlike the restrictive synchronic position of the GTT, the STT defends the diachronic study of language. In the words of Temmerman (2000a: 230):

> In order to understand categorization and naming, an analysis of the history of categories is essential. It proves that naming in science is hardly ever arbitrary and that categorization is a process in time.

This is closely related to the previous idea of polysemy and synonymy (Temmerman 1997: 67):

> Polysemy is functional in LSP discourse, it is a consequence of meaning evolution. The constant discussion over how to name and what words mean is in the discourse of a community and has a time aspect. Polysemy is the result. What is univocal at one time may grow into polysemy depending on the type of category and how it is understood.

We agree with Faber (2009: 118) in that the focus of terms and concepts from a diachronic perspective is indeed one of the most valuable contributions of Sociocognitive Terminology Theory. In recent years, the STT has evolved and now incorporates methods and guidelines of ontological analysis (Fernández, Gómez-Pérez, and Jurista 1997; Staab and Studer [2003] 2009). This combination of Terminology and ontologies is now known as *termontography* (Temmerman and Kerremans 2003; Kerremans, Temmerman, and Tummers 2003).

As is well known, the most widely accepted definition of ontology is that of Gruber (1993: 908), who states that an ontology is "[…] a formal, explicit specification of a shared conceptualization". *Formal* signifies that it should be machine executable; *explicit* refers to the fact that all elements of an ontology should be explicitly defined;

shared means that it should capture consensual knowledge of a group of people (Studer, Benjamins, and Fensel 1998: 175). Based on Genesereth and Nilsson (1987), *conceptualization* is in the words of Gruber (1993: 199):

> […] an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly.

The notion of conceptualization underlies that of knowledge base. Even though database and knowledge base are often used indistinctively to refer to types of computational representation system, there are clearly differences between the two. A database stores large quantities of data by organizing them in a way that enhances their updating, storage and retrieval. In contrast, a knowledge base not only stores data, but represents knowledge. It is thus not surprising that they are closely linked to the notion of ontology.

Regarding specialized language, Meyer (c.f. Meyer, Bowker, and Eck 1992, 1997) was one of the first authors to recognize that databases or termbanks would indeed be greatly enriched if they were organized in a way that resembled the organization of concepts in the mind (Meyer, Bowker, and Eck 1992: 159):

> […] term banks would be more useful, and useful to a wider variety of people eventually even machines, if they contained a richer and more structured conceptual component than they do at present.

As can be inferred from her words, when databases become knowledge bases, data are enhanced because they account not only for conventional *is_a* and *type_of* relations, but also for non-hierarchical relationships. This paves the way to the representation of the multidimensionality of concepts.

Boguraev and Pustejovsky (1996: 9) offer a definition of *lexical knowledge base*, which could be extended to specialized language:

> […] the term "lexical knowledge base" […] [refers] to a large-scale repository of lexical information, which incorporates more than just static descriptions of words, e.g., by means of clusters of properties and associated values. A lexical knowledge base would state: constraints on word behavior, dependence of word interpretation on context, and distribution of linguistic generalizations. It is essentially a dynamic object, as it

> incorporates, in addition to its information types, the ability to perform inference over them and thus induce word meaning in context.

As previously mentioned, the termontography can be regarded as the practical application of the STT. In termontography, first, a framework of domain-specific categories and intercategorial relations is developed with the help of experts in the field. This is followed by the manual and semi-automatic extraction of knowledge from a corpus of specialized texts. This extraction enriches the framework by incorporating more and better-described conceptual relations (Kerremans, Temmerman, and Tummers 2003: 665). The information derived from this analysis is stored in a terminological knowledge base, referred to as *termontological database* (Temmerman and Kerremans 2003).

Temmerman's team is currently working on new implementations of the Termontography methodology, and is developing a set of natural language processing modules, especially for translators, interpreters and other specialists. One of these new tools, referred to as *Interlexor*, is especially interesting for our research and subsequent studies in phraseology, since it extracts terms, variants, collocations, and definitions (Domínguez, Kerremans, and Temmerman 2012). See more information concerning the Termonto Platform tools in <http://taalkunde.ehb.be/cvc/languagetechnology/tools>.

The STT has significantly contributed to the cognitive shift in the study of specialized language. In addition, it has adapted to the new advance in the 21st century, by elaborating a new methodology centered on ontologies, and where ICMs or prototypes have a reduced importance since they are not susceptible to computational representation.

Nevertheless, as pointed by Faber (2009: 119–120), similarly to the TCT and Socioterminology, the STT does not account for syntax. One reason is that Terminology has always been divorced from the study of syntax. In addition, Faber (2009: 120), underlines the subjectivity of conceptualization in STT: "it is impossible to define the exact nature of the center of prototypical categories or explain how degrees of prototypicality can be objectively measured". A further consideration is the fact that this kind of representation is totally unconstrained since it is based on an open-ended inventory of conceptual relations.

1.4.3. Frame-based Terminology

Frame-based Terminology (Faber 2009, 2011, 2012), henceforth FBT, is a recent cognitive approach to Terminology. It shares many of the assumptions of the CTT and STT, such as the *continuum* between words and terms, and the need to study specialized knowledge in real texts. It combines certain aspects of Corpus Linguistics, the Lexical Grammar Model (Faber and Mairal 1999) (§3.1.2.1), and Frame Semantics (Fillmore 1976, 1982, 1985; Fillmore and Atkins 1992) (§3.1.3.2) in order to structure specialized domains and create non-language-specific representations. Such configurations are the conceptual meaning underlying specialized texts in different languages, the specification of which facilitates specialized knowledge acquisition.

In contrast to STT, in which knowledge is organized in ICMs based on prototypes, the FBT organizes knowledge in frames. As is well known, frames are a type of cognitive structuring device based on experience that provide the background knowledge for the words in a language, as well as the way that those words are used in discourse. According to Frame Semantics (§3.1.3.2), in order to understand the meanings of words in a language, it is first necessary to have knowledge of the semantic frames or conceptual structures that underlie their usage.

Unlike ICMs, frames have the advantage of making explicit both the semantic and syntactic behavior of specialized language units, which necessarily includes a description of conceptual relations as well as a term's combinatorial potential (Faber 2009: 123). In addition, ICMs in STT encode the multiple ways of categorizing the same reality. However, this is largely dependent on the circumstances of the conceptualizer and seems to use a totally unconstrained typology of conceptual relations. In contrast, frames do not depend on the intuition of the terminologist or the reader, but are regarded as a more objective way of conceptualizing reality by means of a closed set of hierarchical and non-hierarchical relations. Finally, unlike ICMs, frames can be more easily processed by computers.

FBT mainly focuses on: (i) conceptual organization, based on frames or events; (ii) the multidimensional nature of terminological units, by accounting for both hierarchical and non-hierarchical relations (§1.4.3.2.1); (iii) the extraction of semantic and syntactic information through the use of multilingual corpora and dictionaries (§1.4.3.1). One of the basic premises of this approach is that the configuration of specialized domains is based on the events that generally take place in them, and can be

represented accordingly (Grinev and Klepalchenko 1999). Therefore, each knowledge area can be said to have its own event template.

Accordingly, for the domain of the environment, the Environmental Event (EE) (Figure 5) was derived from corpus and dictionary analysis (§1.4.3.1). The EE is a representation of the most generic conceptual level of the environmental field and underlies the organization of all concepts in the domain. The prototypical domain event or action-environment interface (Barsalou 2003: 513; Faber, Márquez, and Vega 2005; Faber 2011) configures the most generic or base-level categories of a domain, providing a template applicable to all levels of information structuring. In other words, it establishes a frame for the organization of more specific concepts. The specific concepts within each category are organized in a network in which they are linked by both hierarchical and non-hierarchical relations (§1.4.3.2.1).

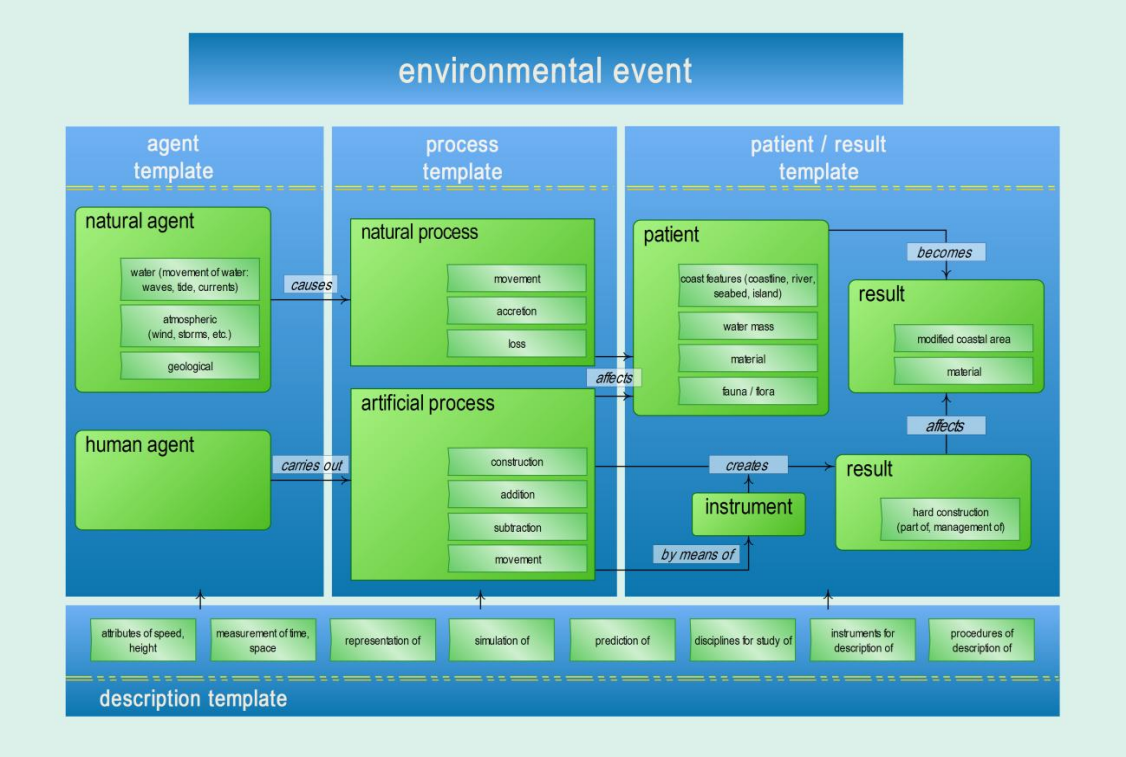


Figure 5. The environmental event (EE)

As shown in Figure 5, the EE has two types of AGENT that can initiate processes: inanimate (natural forces) or animate (human beings). NATURAL AGENTS, such as earth movements and atmospheric phenomena, cause NATURAL PROCESSES in a geographic area. These processes affect other entities or PATIENTS which as a RESULT, may suffer changes. HUMAN AGENTS can also use INSTRUMENTS to implement ARTIFICIAL

PROCESSES (e.g. constructions), which can generate or prevent EFFECTS normally caused by natural processes. AGENT, PATIENT, RESULT and INSTRUMENT are the most characteristic semantic roles of this specialized domain and the EE represents their relationships. However, there are also peripheral categories that include concepts used for the measurement, analysis, and description of the processes in the main event. Accordingly, each subdomain within the event is characterized by a template with a prototypical set of conceptual relations.

This kind of relational representation based on an event facilitates knowledge acquisition for textual processing and production, as stated by Faber (2011: 16):

> Proof of the usefulness of event knowledge can be found in written communication since a comprehender's knowledge of events plays a central role in sentence processing. […] Evidently, terms, whether they designate objects or processes, are powerful cues for the wider event knowledge targeted. In this regard, the choice of a specific term is enough to generate expectations and predictions that constrain the range of likely events.

1.4.3.1 Knowledge extraction in Frame-based Terminology

The FBT methodology used to design EcoLexicon specifies the conceptual system of the domain under study by means of an integrated top-down and bottom-up approach, which inherits the knowledge-based terminology approach of the OncoTerm project (Faber, López, and Tercedor 2001; OncoTerm 2002). The top-down approach includes the information provided by specialized dictionaries and other reference materials, together with the help of experts in the field. The bottom-up approach consists of extracting information from a corpus of texts in various languages, specifically related to the domain.

Dictionary analysis

The information in dictionaries constitutes a lexical-conceptual network that is in direct relation to the knowledge expressed. However, to extract conceptual parameters that can be used to create frames for the definition of conceptual categories and their members, it is necessary to use various dictionaries, and compare definitions in terms of the conceptual relations activated in them (Faber 2002). Since the analysis in this thesis focuses on the environmental subdomain of NATURAL HAZARDS or EXTREME EVENTS in

its sense of natural disaster (§4.1), the concept TROPICAL CYCLONE is used here as an example of the methodology applied to extract knowledge in FBT[3].

TROPICAL CYCLONE is an atmospheric concept known for having an important cause-damage component. In order to extract conceptual information regarding TROPICAL CYCLONE, the definition of TROPICAL CYCLONE was extracted from the following dictionaries: *Encyclopaedia Britannica*[4], *Dictionary of Geophysics, Astrophysics and Astronomy*[5], *The Concise Oxford Dictionary of Earth Sciences* (Allaby and Allaby 1990), *MacMillan Dictionary of the Environment* (Allaby 1988), *Weather Channel Glossary*[6], *Encyclopedia of World Climatology*[7], and *International Meteorological Vocabulary*[8]. The tagged definitions of TROPICAL CYCLONE are shown in Table 1. Each dictionary was assigned a letter in order to compare the information in the definitions more easily:

***Encyclopaedia Britannica* (A)**
Severe atmospheric disturbance **[IS_A//TYPE_OF]** in the tropical oceans between latitudes of approximately 5º and 30º in both hemispheres **[LOCATION]**. These storms are characterized by very low atmospheric pressures **[ATTRIBUTE_OF]** in the calm, clear centre of a circular structure of rain **[ATTRIBUTE_OF]**, cloud **[ATTRIBUTE_OF]**, and very high winds **[ATTRIBUTE_OF/CAUSE]**. In the western Atlantic and the Caribbean they are called hurricanes **[TYPE_OF]**; in the western Pacific, typhoons **[TYPE_OF]**; and in the western Australia, willy-willies **[TYPE_OF]** (if the surface winds exceed 117 kilometres (73 miles) per hour **[ATTRIBUTE_OF]**).
***Dictionary of Geophysics, Astrophysics and Astronomy* (B)**
A large low pressure system **[IS_A/ TYPE_OF]** that originates over the tropical oceans **[LOCATION]**; including tropical depressions **[PHASE_OF]**, tropical storms **[PHASE_OF]** and hurricanes (cyclones, typhoons) **[TYPE_OF]**, with winds sometimes up to 300 km/h **[ATTRIBUTE_OF].**
***The Concise Oxford Dictionary of Earth Science* (C)**
A generally fairly small but intense, closed low-pressure system **[IS_A/ TYPE OF]** which

[3] In this thesis, concepts are written in small capitals and conceptual relations are offered in italics.
[4] <http://www.britannica.com/> [12/01/2012].
[5]<http://www.deu.edu.tr/userweb/emre.timur/dosyalar/Dictionary%20of%20Geophysics,%20Astrophysics%20and%20Astronomy.pdf> [12/01/2012].
[6] <http://www.weather.com/glossary/> [12/01/2012].
[7]<http://www.springerreference.com/docs/navigation.do?m=Encyclopedia+of+World+Climatology+(Earth+and+Environmental+Science)-book34> [12/01/2012].
[8] <http://www.wmo.int/e-catalog/detail_en.php?PUB_ID=402> [12/01/2012].

develop over tropical oceans **[LOCATION]**. Wind speeds of at least 33 m/s (force 12 on the Beaufort scale, 64 knots or more) **[ATTRIBUTE_OF]** define such storms and distinguishing them from less intense systems, e.g. tropical depressions **[PHASE_OF]** (of twice or more than twice the diameter) or tropical storms **[PHASE_OF]**.
***The Weather Channel Glossary* (D)**
A warm core low pressure system **[IS_A/ TYPE_OF]** which develops over tropical, and sometimes subtropical waters **[LOCATION]**, and has an organized circulation **[ATTRIBUTE_OF]**. Depending on sustained surface winds, the system is classified as a tropical disturbance **[IS_A]**, a tropical depression **[PHASE_OF]**, a tropical storm **[PHASE_OF]**, or a hurricane **[TYPE_OF]** or typhoon **[TYPE_OF]**.
***Encyclopedia of World Climatology* (E)**
Warm-core, non-frontal low-pressure systems of synoptic scale **[IS_A]** that develop over tropical or subtropical oceans **[LOCATION]**.
***International Meteorologic Vocabulary* (F)**
Cyclone **[IS_A/ TYPE_OF]** of tropical origin of small diameter (some hundreds of kilometres) with minimum surface pressure in some cases less than 900 mb, very violent winds **[ATTRIBUTE_OF]**, and torrential rain **[ATTRIBUTE_OF]**; sometimes accompanied by thunderstorms **[ATTRIBUTE_OF]**. It usually contains a central region known as the 'eye' of the storm **[PART_OF]**, with diameter of the order of some tens of kilometres, and with light winds and more or less lightly clouded sky.
***MacMillan Dictionary of the Environment* (G)**
A violent storm **[IS_A/TYPE_OF]**, with a very small area of low pressure **[ATTRIBUTE_OF]** at the centre, around which the isobaras are almost circular, very close together and the winds extremely violent **[ATTRIBUTE_OF/]**. The term cyclone **[TYPE_OF]** is used in the Indian Ocean and Bay of Bengal, hurricane **[TYPE_OF]** in the Caribbean, typhoon **[TYPE_OF]** in the China Sea and Willy willy **[TYPE_OF]** in the Western Australia.

Table 1. Tagged definitions of TROPICAL CYCLONE

The definitions are then compared so as to craft a well-structured definition that encodes the relations and attributes of the concept in question. The relations in each of the definitions are listed in Table 2:

	type_of/ is_a	*location*	*attribute_of*	*has_type*	*phase_of*
A	SEVERE ATMOSPHERIC DISTURBANCE	TROPICAL OCEANS	*-low atmospheric pressure* *-rain* *-cloud* *-high winds:+117km/h*	-HURRICANE -TYPHOON -WILLY-WILLIES	
B	LARGE LOW PRESSURE SYSTEM	TROPICAL OCEANS	*winds up to 300 km/h*	-HURRICANE -TYPHOON	-TROPICAL DEPRESSION -TROPICAL STORM
C	CLOSED LOW PRESSURE SYSTEM	TROPICAL OCEANS	*winds at least 33 m/s*		-TROPICAL DEPRESSION -TROPICAL STORM
D	WARM CORE LOW PRESSURE SYSTEM// TROPICAL DISTURBANCE	-TROPICAL WATERS -SUBTROPICAL WATERS	*organized convection*	-HURRICANE -TYPHOON	-TROPICAL DEPRESSION -TROPICAL STORM
E	WARM-CORE NON-FRONTAL LOW PRESSURE SYSTEMS	-TROPICAL OCEANS -SUBTROPICAL OCEANS			
F	CYCLONE		*-small diameter* *-minimum surface pressure* *-violent winds* *-torrential rain* *-thunderstorms*		
G	VIOLENT STORM		*-small area of low pressure* *-violent winds* *-torrential rain* *-thunderstorms*	-CYCLONE -HURRICANE -TYPHOON -WILLY-WILLY	

Table 2. Comparative definitions of TROPICAL CYCLONE

As can be observed in Table 2, all the definitions of TROPICAL CYCLONE contain a *genus*, i.e. the generic term that designates the category membership of a concept. This *genus* is indicative of the *type_of* relationship and points to an ATMOSPHERIC DISTURBANCE or a LOW PRESSURE SYSTEM. Five of the seven definitions refer to its *location* (i.e. where a TROPICAL CYCLONE takes place), which is in TROPICAL WATERS. Half of the definitions include *phase_of* and *type_of* relations. The *phase_of* relation refers to how a TROPICAL CYCLONE develops, which begins as a TROPICAL DEPRESSION, and then often evolves into a TROPICAL STORM. The *has_type* relation (the inverse of *type_of*) highlights the various kinds of TROPICAL CYCLONE, which are principally

HURRICANE and TYPHOON. The *attribute* encoded in all of the definitions is *high winds*, whose speed must be at least *33 m/s = 117 km/h*. This analysis creates a definitional template for TROPICAL CYCLONE and its subtypes, which can be further enriched with the information derived from corpus analysis.

Corpus analysis

In the same way as the analysis of dictionary definitions, corpus analysis offers the possibility of using linguistic clues to extract conceptual information regarding semantic relations. This information is semantically classified and analysed so that the concept can be placed in the underlying conceptual framework of a knowledge domain event (Faber et al. 2007).

Before analysing concordances to obtain conceptual information, a word list is obtained from the corpus[9]. As Bowker and Pearson (2002) state, word lists enable terminologists to investigate patterns that might not otherwise have been observed. According to the list, the most frequent lexical units are 'hurricane', 'typhoon', 'cyclone', and 'depression', all of which play a prominent role in the frame of TROPICAL CYCLONE. Other related terms with a high frequency are 'storm', 'tornado', 'flood', 'flooding', 'rain', 'precipitation', 'tsunami', 'rainfall', 'landfall', 'surge', and 'thunderstorm', which are some of the possible results or consequences of TROPICAL CYCLONE. Words such as 'velocity', 'speed', 'mph', and 'knots' are also important since they indicate the conceptual relation *has_velocity*/*has_intensity*. The relatively high frequency of 'season', and certain months (i.e. 'July' or 'September') refer to the fact that a TROPICAL CYCLONE generally occurs at a particular time of the year. Words such as 'scale', 'category', and 'measurement' underline the fact that a TROPICAL CYCLONE can vary in intensity. In addition, the frequency of ocean names (i.e. 'Pacific' and 'Atlantic') and cardinal points ('east', 'south', 'northern', and 'southern') and regional indicators ('subtropical' and 'extratropical') indicate that this phenomenon takes place at a certain location.

The concordances are then analysed, and the corpus is searched for the string *tropical cyclone** in order to find knowledge-rich contexts. These contexts provide conceptual information referring to the relations between the entity and its attributes,

[9] See §4.3.2 for a fine-grained description of the information that can be extracted from a WordList. In addition, section 5.2 displays the wordlist with the 50 most frequent units of our corpus in the field of natural hazards as well as a complete description of the corpus used and the steps to follow in order to obtain reliable results.

which make up the categorisation structure of TROPICAL CYCLONE. Meyer (2001: 279) defines a knowledge-rich context as follows:

> a context indicating at least one item of domain knowledge that could be useful for conceptual analysis. In other words, the context should indicate at least one conceptual characteristic, whether it be an attribute or relation.

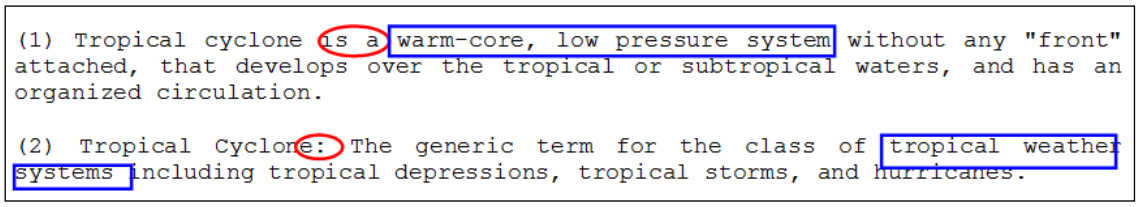
As is well known, conceptual relations manifest themselves in certain predictable recurring patterns or as lexical markers in the text. A *knowledge pattern* (KP) (Meyer 2001; Barrière 2004; Barrière and Agbago 2006) refers to the explicit domain-independent knowledge patterns, metalinguistic information regarding terms, and their conceptual structure. Such KPs generally take the form of linguistic markers that help the reader understand the meaning of a concept and how this concept relates to others, according to the relation conveyed. On the basis of the typology of lexical patterns proposed by Bowker and Pearson (2002: 219) (Table 3), we then show how knowledge has been extracted from our corpus in order to describe the relations and attributes of TROPICAL CYCLONE.

Lexical patterns	**Type of conceptual knowledge provided**	**Example**
is a, kind of, type of, includes	Describe generic-specific relations	The tabor is a type of drum.
has a, contains, consists of, includes	Describes part-whole relations	A snare drum has a batter head and a snare head.
used for, used to, employed to	Describes the function of an item	A wooden stick is used to strike the drum head.
causes, produces, produced by, results from	Describes relations of cause and effect	Striking the drum head causes the snares to vibrate.
also called, also known as, sometimes referred to as	Indicates possible synonymy	The tambourine, also known as the tambourin provençale, is the largest of all the tabors

Table 3. Lexical patterns and the possible knowledge-rich contexts that they can activate (Bowker and Pearson 2002: 219)

The first relation explored was *type_of*, which was marked by *is_ a* and *are*, as well as by the colon (see examples 1 and 2, Fig. 6). In the concordances displayed

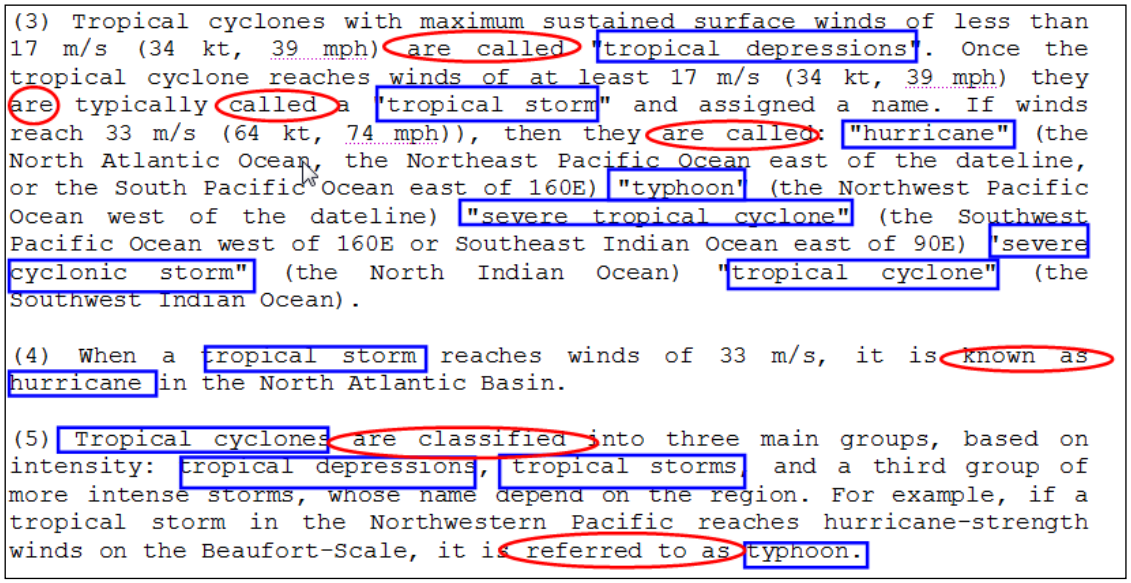
henceforth, the KPs are circled in red, whereas the information codified by the KPs is enclosed in blue rectangles.


(1) Tropical cyclone is a warm-core, low pressure system without any "front" attached, that develops over the tropical or subtropical waters, and has an organized circulation.

(2) Tropical Cyclone: The generic term for the class of tropical weather systems including tropical depressions, tropical storms, and hurricanes.

Figure 6. *type_of* concordances of TROPICAL CYCLONE

As shown in Figure 6, the possible genus of TROPICAL CYCLONE extracted from the corpus is quite similar to the generic terms extracted from dictionary definitions, and indicates that a TROPICAL CYCLONE is a low pressure system or tropical weather system. The lexical markers in text concordances that were also used to retrieve the different *types* and *phases_of* a TROPICAL CYCLONE were the following: *are called*, *known as*, *referred to as*, *are classified* (Figure 7):


(3) Tropical cyclones with maximum sustained surface winds of less than 17 m/s (34 kt, 39 mph) are called "tropical depressions". Once the tropical cyclone reaches winds of at least 17 m/s (34 kt, 39 mph) they are typically called a "tropical storm" and assigned a name. If winds reach 33 m/s (64 kt, 74 mph)), then they are called: "hurricane" (the North Atlantic Ocean, the Northeast Pacific Ocean east of the dateline, or the South Pacific Ocean east of 160E) "typhoon" (the Northwest Pacific Ocean west of the dateline) "severe tropical cyclone" (the Southwest Pacific Ocean west of 160E or Southeast Indian Ocean east of 90E) "severe cyclonic storm" (the North Indian Ocean) "tropical cyclone" (the Southwest Indian Ocean).

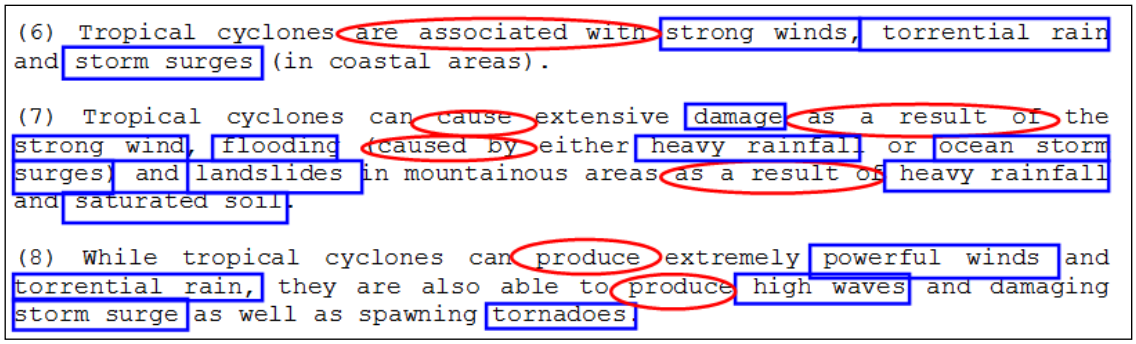
(4) When a tropical storm reaches winds of 33 m/s, it is known as hurricane in the North Atlantic Basin.

(5) Tropical cyclones are classified into three main groups, based on intensity: tropical depressions, tropical storms, and a third group of more intense storms, whose name depend on the region. For example, if a tropical storm in the Northwestern Pacific reaches hurricane-strength winds on the Beaufort-Scale, it is referred to as typhoon.

Figure 7. *type_of and phase_of* concordances of TROPICAL CYCLONE

The information extracted from the concordances in (3), (4), and (5) revealed lexical units linked to TROPICAL CYCLONE by the following relations: *type_of* (HURRICANE, TYPHOON, SEVERE TROPICAL CYCLONE, SEVERE CYCLONIC STORM), and *phase_of* (TROPICAL DEPRESSION and TROPICAL STORM), both of which respond to the same KP.

The consequences or results potentially produced by a CYCLONE are identified by the markers *cause, caused by, produce, produced by, as a result of,* and *associated with,* which indicate the *cause* relation:


(6) Tropical cyclones are associated with strong winds, torrential rain and storm surges (in coastal areas).

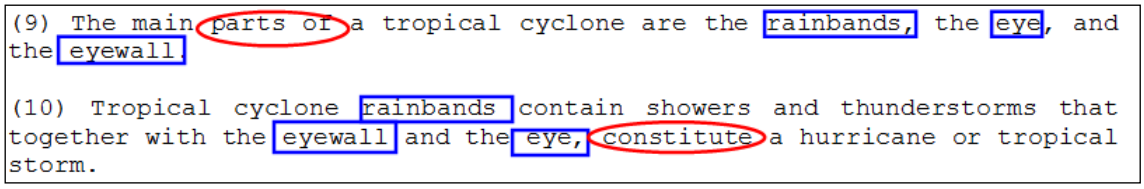
(7) Tropical cyclones can cause extensive damage as a result of the strong wind, flooding (caused by either heavy rainfall or ocean storm surges) and landslides in mountainous areas as a result of heavy rainfall and saturated soil.

(8) While tropical cyclones can produce extremely powerful winds and torrential rain, they are also able to produce high waves and damaging storm surge as well as spawning tornadoes.

Figure 8. *cause* concordances of TROPICAL CYCLONE

The concordances in Figure 8 show that a TROPICAL CYCLONE *causes* WIND, TORRENTIAL RAIN, STORM SURGE, FLOODING, RAINFALL, LANDSLIDE, and TORNADO.

Typical KPs that denote the conceptual relation *part_of* are *part(s) of, contain* or *constitute,* as can be seen in (9) and (10). In this way, it was possible to identify the different parts of a TROPICAL CYCLONE: EYE, EYEWALL and RAINBAND (Figure 9):


(9) The main parts of a tropical cyclone are the rainbands, the eye, and the eyewall.

(10) Tropical cyclone rainbands contain showers and thunderstorms that together with the eyewall and the eye, constitute a hurricane or tropical storm.

Figure 9. *part_of* concordances of TROPICAL CYCLONE

Other units formalize different types of relation between entities and processes. Examples include the AGENTS that produce/cause TROPICAL CYCLONE as reflected in the KPs *come from* or *form from*. As shown in Figure 10, there are four AGENTS that can generate a TROPICAL CYCLONE: EASTERLY WAVE (also called TROPICAL WAVE), WEST AFRICAN DISTURBANCE LINE (WADL), TROPICAL UPPER TROPOSPHERIC TROUGH (TUTT), and OLD FRONTAL BOUNDARY:

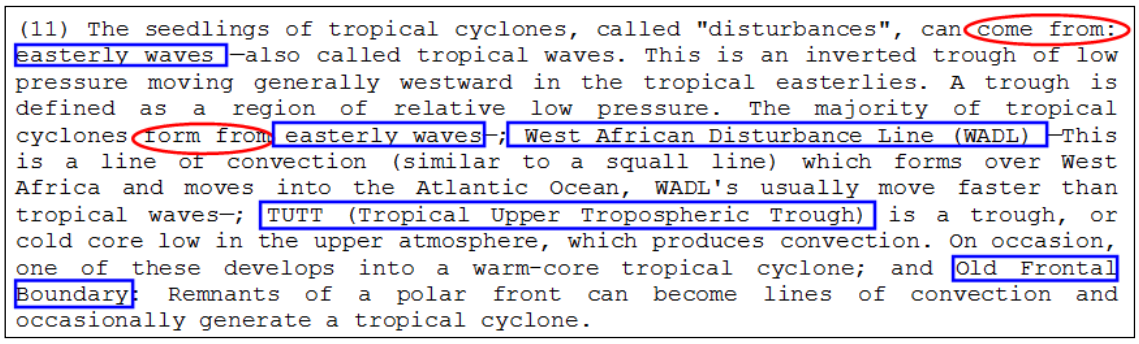
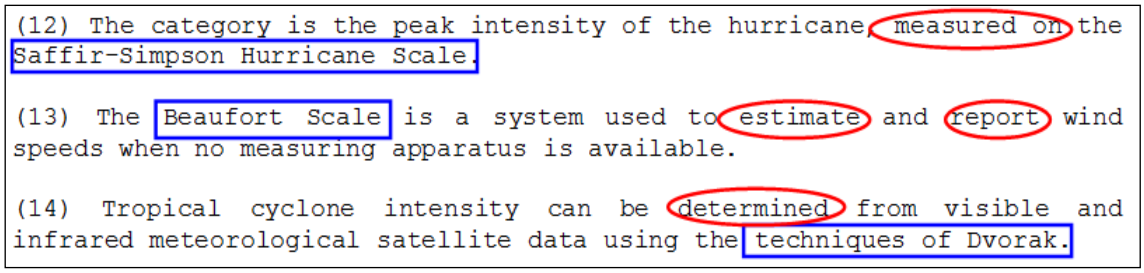

(11) The seedlings of tropical cyclones, called "disturbances", can come from: easterly waves —also called tropical waves. This is an inverted trough of low pressure moving generally westward in the tropical easterlies. A trough is defined as a region of relative low pressure. The majority of tropical cyclones form from easterly waves—; West African Disturbance Line (WADL) —This is a line of convection (similar to a squall line) which forms over West Africa and moves into the Atlantic Ocean, WADL's usually move faster than tropical waves—; TUTT (Tropical Upper Tropospheric Trough) is a trough, or cold core low in the upper atmosphere, which produces convection. On occasion, one of these develops into a warm-core tropical cyclone; and Old Frontal Boundary: Remnants of a polar front can become lines of convection and occasionally generate a tropical cyclone.

Figure 10. *has_origin* concordances of TROPICAL CYCLONE

The scales and techniques for measuring the intensity of a TROPICAL CYCLONE were also identified by means of the lexical markers *estimate*, *report*, *determine*, *and measure*. As shown in Figure 11, the intensity of a TROPICAL CYCLONE is determined by the SAFFIR-SIMPSON HURRICANE SCALE and the BEAUFORT SCALE, mainly by means of DVORAK TECHNIQUES:


(12) The category is the peak intensity of the hurricane, measured on the Saffir-Simpson Hurricane Scale.

(13) The Beaufort Scale is a system used to estimate and report wind speeds when no measuring apparatus is available.

(14) Tropical cyclone intensity can be determined from visible and infrared meteorological satellite data using the techniques of Dvorak.

Figure 11. *measured_by* concordances of TROPICAL CYCLONE

The KPs *form*, *formation area*, and the relative pronoun *where*, show that a TROPICAL CYCLONE takes place at certain locations. Examples (15) and (16) in Figure 12 point to seven different basins, namely, *north Atlantic Ocean*, *eastern Pacific Ocean*, *western Pacific Ocean*, *southwestern Pacific Ocean*, *southwestern Indian Ocean*, *southeastern Indian Ocean* and *northern Indian Ocean*, which are all TROPICAL and SUBTROPICAL WATERS:

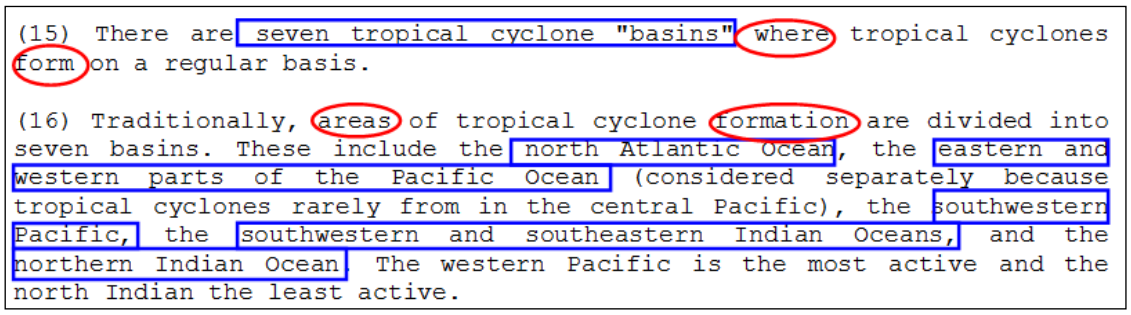


Figure 12. *has_location* concordances of TROPICAL CYCLONE

In addition, a TROPICAL CYCLONE occurs at a certain time of year, known as TROPICAL CYCLONE SEASON or HURRICANE SEASON (17) and (18). The patterns or lexical markers that help retrieve this information are *period* and *occur*. As can be seen, TROPICAL CYCLONE SEASON varies, depending on the basin involved. The basins are underlined in Figure 13:

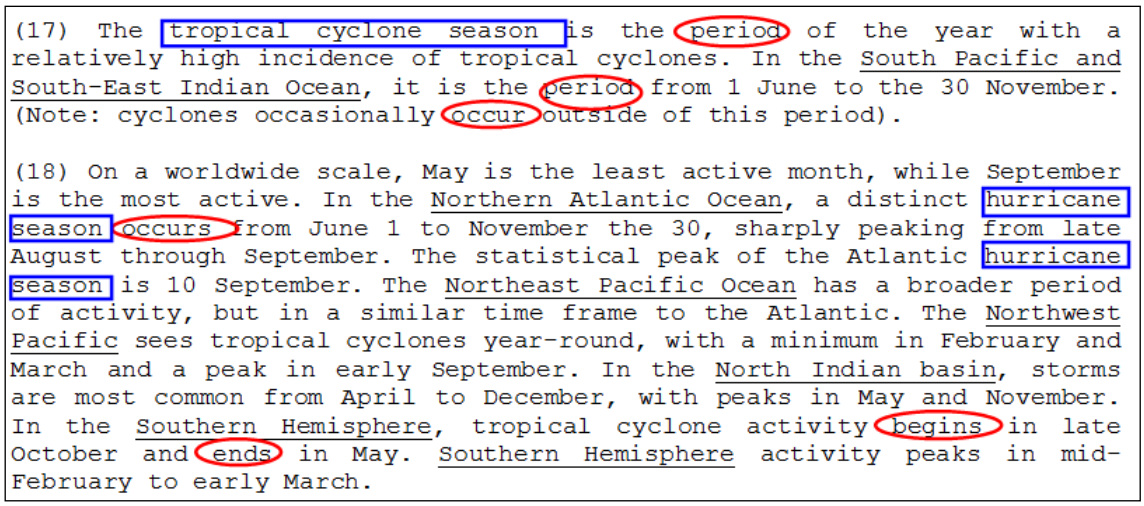


Figure 13. *has_time* concordances of TROPICAL CYCLONE

The analysis of the concordances along with the analysis of dictionary definitions permitted the extraction of the conceptual relations and attributes that were activated in TROPICAL CYCLONE and which are listed in Table 4:

TROPICAL CYCLONE	
type_of	TROPICAL DISTURBANCE
has_origin	EASTERLY WAVE WEST AFRICAN DISTURBANCE LINE TUTT OLD FRONT BOUNDARY
attribute (has_intensity)	winds of 33m/s (64kt, 74mph, 118km/h)
measured_by (instrument)	SAFFIR-SIMPSON HURRICANE SCALE BEAUFORT SCALE
measured_by (technique)	DVORAK TECHNIQUE
causes	TORRENTIAL RAIN STORM SURGE FLOODING TORNADO TSUNAMI LANDFALL
has_location	TROPICAL WATERS SUBTROPICAL WATERS
has_time	HURRICANE SEASON
has_type	HURRICANE TYPHOON
has_phase	TROPICAL DEPRESSION TROPICAL STORM
has_part	EYE EYEWALL RAINBANDS

Table 4. TROPICAL CYCLONE conceptual relations and attributes

As shown, the combined analysis of dictionaries and corpus are the basis for the specification of the TROPICAL CYCLONE frame in EcoLexicon Following this methodology the various frames within the general Environmental Event were specified.

1.4.3.2. Practical application: EcoLexicon

The practical application of FBT is the environmental knowledge base EcoLexicon (http://ecolexicon.ugr.es). EcoLexicon represents the conceptual structure of the

specialized domain of the Environment in the form of a visual thesaurus in which environmental concepts are configured in semantic networks. The various terminological designations for a concept are offered in six languages: Spanish, English, German, French, Russian, and Greek. In EcoLexicon it is assumed that up to a certain level, its potential users are familiar with scientific language and its usage in English or Spanish at least, since these are the interface languages. Potential users should thus possess a good command of any of the six languages in the knowledge base, as well as a minimum of scientific knowledge (López, Buendía, and García 2012: 62).

In consonance with this, EcoLexicon is designed to meet the needs of different user types, such as a student of science wishing to acquire specialized knowledge about a certain concept, a translator seeking translation correspondences in a language, or a specialist interested in text production. As such, following Sager (1990: 197-199), Pearson (1998: 35-39), and Bergenholtz and Tarp (2010: 34–35), there are three user profiles: laypeople, semi-experts and experts.

Laypeople are potential dictionary users who have a basic knowledge of general science obtained in high school (Bergenholtz and Tarp 2010: 34). In other words, people using this type of communication are not assumed to have subject-specific knowledge, only a good command of the language in which the communication is written. This approach is common in science magazines, such as the *New Scientist* (Pearson 1998: 38).

Semi-specialized communication occurs when experts communicate with others who have some knowledge of the field, but do not have the same level of expertise. Term density is likely to be lower than in specialized communication as the exchange includes explanations of the terminology used (Pearson 1998: 37). Within the environmental domain, semi-experts may be the following: (i) experts from other related fields, working in the public or private sectors, who are confronted daily with environmental information; (ii) general advisors on science and technology, whose daily job involves advising politicians at local level; (iii) journalists, who are exposed to this subject frequently; (iv) politicians and other decision makers who work with environmental issues and thus, have become so familiar with environmental matters, that they can be regarded as semi-experts (Bergenholtz and Tarp 2010: 35). Bergenholtz and Nielsen (2002: 6) suggest that advanced learners, such as translation students and professional translators, are also considered to be semi-experts since they “have

acquired substantial factual knowledge as a result of their education, training and work with LSP".

Specialized or expert communication is the type of communication with the highest density of specialized terms. Therefore, the language used differs substantially from general language. Specialized language discourse does not offer any explanation of the terminology unless it is to redefine an existing concept or to coin a new term (Pearson 1998: 36). Nevertheless, user categorisation is not as simple as it may first appear since according to Bergenholtz and Nielsen (2002: 6), experts within a subject field are likely to have considerable linguistic and textual knowledge in their own language, but considerably less knowledge of the foreign language. On the other hand, experts may be experts within one or two subject fields, but laypersons or semi-experts in rest of fields. As Bergenholtz and Tarp (2010: 35) state:

> A true expert will have no reception problems within his own field. He may have to acquire new knowledge, but he is not likely to find this in any lexicographical dictionary.

In EcoLexicon, searches can be carried out either by concept or by term. In addition, searches can be constrained to a specific contextual domain (i.e. climatology, ecology, meteorology, oceanography, zoology, etc.)[10]. Figure 14 shows the entry for HURRICANE as displayed in EcoLexicon. As can be observed, on the right-hand side of the screen, there is a conceptual network that links the search concept to all related concepts. On the left side, the modules *Definition, Terms, Resources,* and *Conceptual Category* supply linguistic, graphical, and conceptual information regarding the search concept:

[10] Cf. León and San Martín (2011: 175) for a detailed description of all the domains included in EcoLexicon.

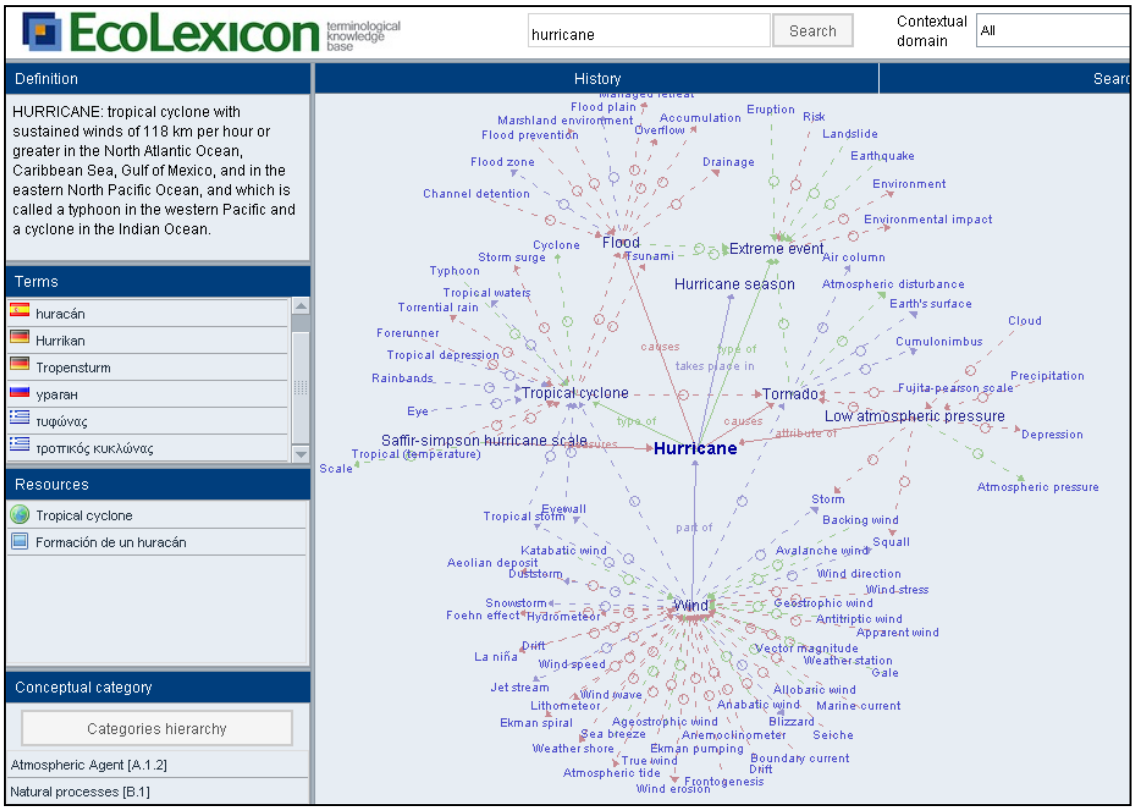


Figure 14. Entry for HURRICANE in EcoLexicon

1.4.3.2.1 Conceptual information

Conceptual information is reflected in EcoLexicon in three ways: (i) the semantic network for each concept, which is based on a closed inventory of semantic relations and which is also lexically represented in the definition; (ii) the association to a *conceptual category*, which makes it possible to access the classes of the ontology to which the search concept belongs. For example, thanks to this incipient ontology[11], the user knows that a HURRICANE can be conceptualized either as an ATMOSPHERIC ENTITY that initiates a process, or as the PROCESS itself; (iii) *resources* that are pictorial representations of the concept.

Semantic relations and definitions

In EcoLexicon, each concept type is related to other concepts by a set of conceptual relations, which are hierarchical *(is_a, type_of part_of)* and non-hierarchical (*made_of, phase_of, delimited_by, located_at,* and *attribute_of*). As previously mentioned, knowledge patterns facilitate the extraction of conceptual information regarding

[11] For more information about the ontology behind EcoLexicon, see León and Magaña (2010).

semantic relations. Table 5 lists some of the most important patterns for the conceptual relations in the specialized domain of the environment (León and Reimerink 2010: 20):

Conceptual relation	**Knowledge pattern**
is_a	such as, rang* from, includ*
part_of	includ*, consist* of, formed by/of
made_of	consist* of, built of/from, constructed of, formed by/of/from
located_at	form* in/at/on, found in/at/on, tak* place in/at, located in/at
result_of	caused by, leading to, derived from, formed when/by/from
has_function	designed for/to, built to/for, purpose is to, used to/for
effected_by	carried out with, by using

Table 5. Knowledge patterns and their conceptual relations

The definition of the search concept can be visualized either by placing the mouse over it in the conceptual and dynamic network on the right in the previous figure (Figure 14), or in the left-hand side section, where there is a special section for definitions. A HURRICANE is defined as follows:

> tropical cyclone [*type_of*] with sustained winds of 118 kms per hour or greater [*attribute_of*], in the North Atlantic Ocean, Caribbean Sea, Gulf of Mexico, and in the eastern North Pacific Ocean [*has_location*], and which is called 'typhoon' [*type_of*], in the western Pacific and 'cyclone' [*type_of*], in the Indian Ocean.

Apart from the relations activated in the definition, more knowledge about HURRICANE is also included. For example, a hurricane *takes_place_in* the HURRICANE SEASON and it *causes* TORRENTIAL RAIN, STORM SURGE, FLOODING, LANDSLIDE, or TORNADO. The EYE, EYEWALL, and RAINBANDS are *part_of* a HURRICANE. Regarding the attributes of the concept, it is a LOW ATMOSPHERIC PRESSURE AREA, and its intensity is *measured_by* the SAFFIR-SIMPSON HURRICANE SCALE.

Graphical information

Another important aspect of Frame-based Terminology is that it includes images to represent specialized concepts. It clarifies how the linguistic and graphical description of specialized entities are linked and can converge to highlight the multidimensional nature of concepts as well as the conceptual relations within a specialized domain

(Faber et al. 2007). Graphical information is selected with the aim of highlighting the content of the definition. Depending on the type of knowledge activated, three different kind of images are distinguished, namely, iconic, abstract, or dynamic images (Prieto 2009; Prieto and López 2009).

Iconic images resemble the real-world object represented through the abstraction of conceptual attributes in the illustration. This kind of image is especially relevant in the representation of generic-specific relations since its similarity to the entity in the real world allows the user to identify the object through the inference of its basic characteristics. Figure 15 shows two examples of iconic images of HURRICANE: (i) the possible consequences of a HURRICANE; (ii) the place where this atmospheric phenomenon usually happens.

HURRICANE

- TORRENTIAL RAIN, STORM SURGE, FLOODING, TORNADO, LANDFALL, TSUNAMI: *cause*

- TROPICAL AND SUBTROPICAL WATERS: *location*

Figure 15. Iconic images of HURRICANE[12]

Abstract images generally focus on a certain aspect or aspects of the concept. Abstraction refers to the cognitive effort made by the perceiver to recognize the concept thus represented (Levie and Lentz 1982; Park and Hopkins 1993; Rieber 1994 as quoted in Prieto and López 2009: 184). Figure 16 shows two abstract images of HURRICANE: the image on the left focuses on the parts of a HURRICANE: EYE, EYEWALL, and RAINBANDS, whereas the one on the right reflects the time period during which these phenomena take place. As can be seen in the graph, the HURRICANE SEASON lasts from 1 June until 30 November. The period of the greatest activity is in the middle of August and the end of October.

[12] The photo on the left was retrieved from: <http://earthobservatory.nasa.gov/IOTD/view.php?id=471> [15/06/2012], and that of the right is courtesy of NOAA, Wikimedia Commons: <http://beyondweather.ehe.osu.edu/issue/getting-warmer/consequences-of-climate-change-lessons-about-water-availability-and-extreme-weather> [15/06/2012].

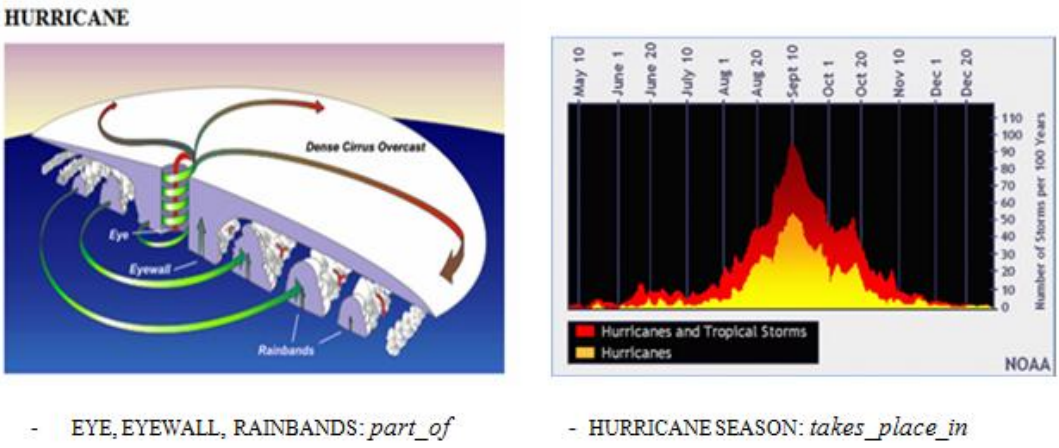


Figure 16. Abstract images of HURRICANE[13]

Finally, dynamic images represent either spatial or temporal movement. Figure 17[14] shows the different phases of HURRICANE, which goes from being a TROPICAL DEPRESSION to a TROPICAL STORM, until it finally becomes a HURRICANE.

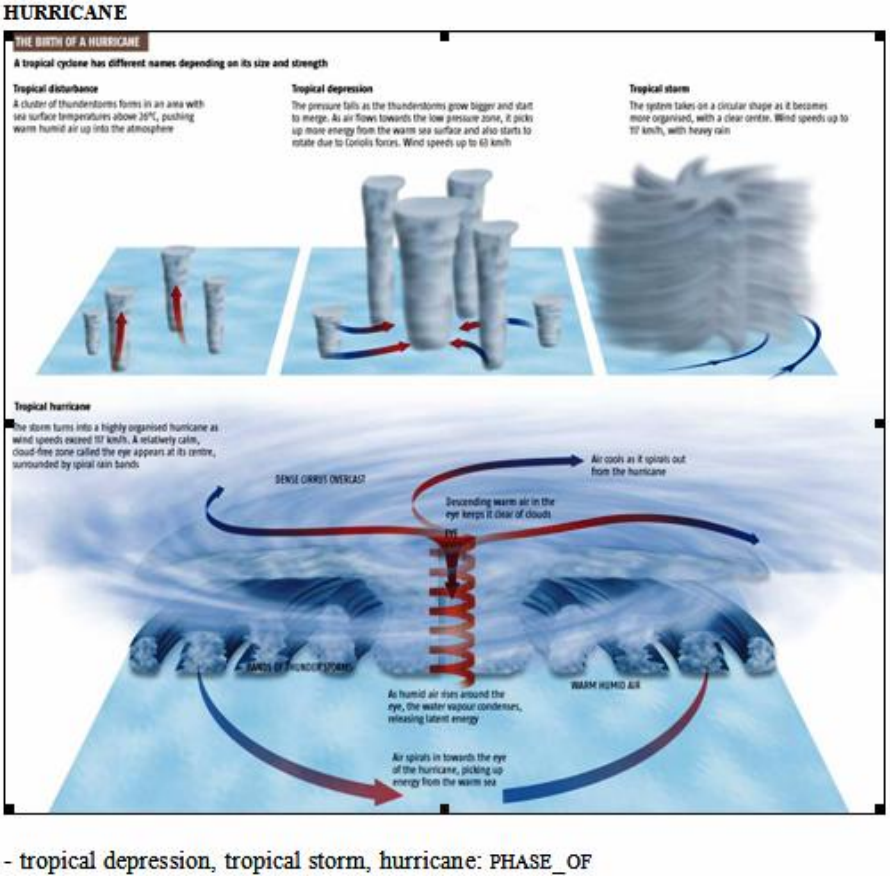


Figure 17. Dynamic image of HURRICANE

[13] The picture on the left has been taken from (Roussy 2006) and that of the right from (Tisdal 2012).
[14] The picture has been taken from:
<http://www.newscientist.com/data/images/archive/2528/25281301.jpg> [08/05/2012].

1.4.3.2.2 Linguistic information

The terms associated with HURRICANE are offered in the section *Terms* on the left side of the screen (Figure 13): 'huracán', 'hurricane', 'Hurrikan', 'Tropensturm', 'ураган', 'τυφώνας', 'ouragan'. By clicking on each term, a new window is displayed which provides users with the following linguistic information (Figure 18): (i) morphosyntactic information regarding grammatical category (noun, verb, adjective or adverb); term type (main term, synonym, geographical variant, and acronym); gender (masculine, feminine, and neuter); (ii) contexts of use.

As shown in Figure 18 in the new window displayed for the term, 'hurricane' is described as a noun and the main entry term. No gender is specified because the term is in English. A set of usage contexts can be visualized through the txt file link (i.e. *hurric3a.txt*) (see Figure 19 below).

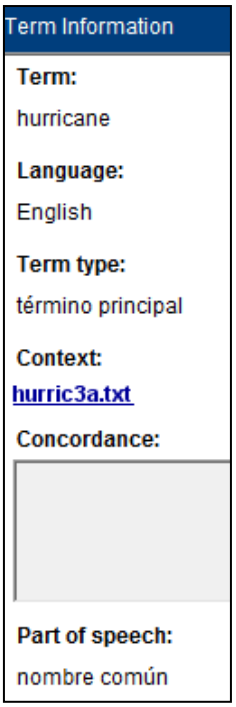


Figure 18. Linguistic information in EcoLexicon: 'hurricane'

As shown, EcoLexicon provides a great quantity of linguistic information for each specialized knowledge unit, but does not provide any phraseological information. This is a problem since phraseology is extremely important, especially for translators. In fact, the results of a questionnaire completed by students of translation at the University of Granada (López, Buendía, and García 2012) (§1.5) reflected that phraseology along

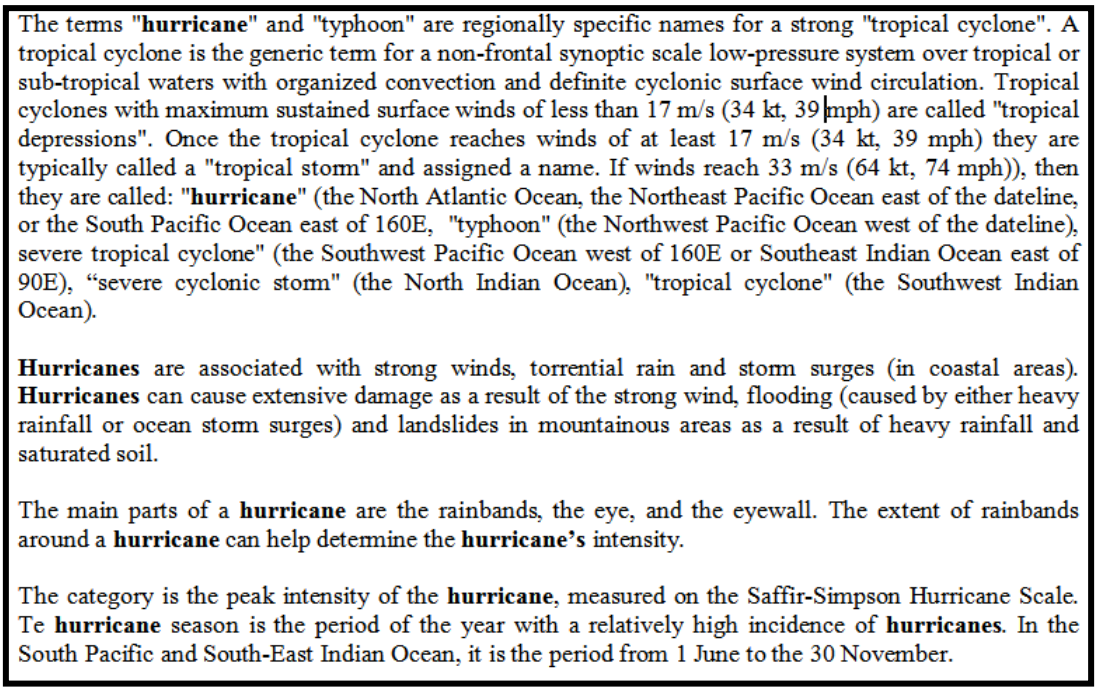
with usage contexts were considered to be the most useful information for text production in the translation process.

Accordingly, this thesis proposes the inclusion of a phraseology module within the term information window contained in Figure 18, which provides the collocational information for each term. Even though we plan to include the most representative phraseological collocations of each specialized knowledge unit in EcoLexicon, we decided to begin with verb phrases for each term because verbs are regarded as the most important lexical and syntactic category of language (Fellbaum 1990; Hanks 2008). Furthermore, there are currently few terminographic resources that incorporate them. This is mainly because terminology has generally focused on noun phrases and has played down the role of other grammatical categories, such as verbs (L'Homme 1998; López 2007; Lorente 2007).

As mentioned in the introduction, the design of a template for encoding and describing collocations in EcoLexicon requires an analysis of the characteristics and limitations of the most representative terminographic resources that include phraseological information (§2), along with the main linguistic theoretical models dealing with predicate argument structure (§3).

Usage contexts

Contexts complete the information given in the definition and facilitate knowledge acquisition. According to Meyer (2001: 279), *knowledge-rich contexts* should contain at least one conceptual item, either an attribute or a relation. However, some contexts are richer in knowledge than others. In this sense, (Reimerink, García de Quesada, and Montero 2010) distinguish between *meaningful context*, a context which includes at least one knowledge element, and a *defining context*, a context which includes all or most of the elements necessary to understand a concept. In EcoLexicon, for every term, various usage contexts are displayed. Figure 19 shows the usage contexts for the term 'hurricane':



The terms "**hurricane**" and "typhoon" are regionally specific names for a strong "tropical cyclone". A tropical cyclone is the generic term for a non-frontal synoptic scale low-pressure system over tropical or sub-tropical waters with organized convection and definite cyclonic surface wind circulation. Tropical cyclones with maximum sustained surface winds of less than 17 m/s (34 kt, 39 mph) are called "tropical depressions". Once the tropical cyclone reaches winds of at least 17 m/s (34 kt, 39 mph) they are typically called a "tropical storm" and assigned a name. If winds reach 33 m/s (64 kt, 74 mph)), then they are called: "**hurricane**" (the North Atlantic Ocean, the Northeast Pacific Ocean east of the dateline, or the South Pacific Ocean east of 160E, "typhoon" (the Northwest Pacific Ocean west of the dateline), severe tropical cyclone" (the Southwest Pacific Ocean west of 160E or Southeast Indian Ocean east of 90E), "severe cyclonic storm" (the North Indian Ocean), "tropical cyclone" (the Southwest Indian Ocean).

Hurricanes are associated with strong winds, torrential rain and storm surges (in coastal areas). **Hurricanes** can cause extensive damage as a result of the strong wind, flooding (caused by either heavy rainfall or ocean storm surges) and landslides in mountainous areas as a result of heavy rainfall and saturated soil.

The main parts of a **hurricane** are the rainbands, the eye, and the eyewall. The extent of rainbands around a **hurricane** can help determine the **hurricane's** intensity.

The category is the peak intensity of the **hurricane**, measured on the Saffir-Simpson Hurricane Scale. Te **hurricane** season is the period of the year with a relatively high incidence of **hurricanes**. In the South Pacific and South-East Indian Ocean, it is the period from 1 June to the 30 November.

Figure 19. Usage contexts for 'hurricane' in EcoLexicon

These contexts (see Figure 19) allow the user to access information regarding the places and time period in which a HURRICANE can occur, its phases, attributes, components, consequences, and the scale on which its intensity is measured.

1.5 The translation process and the user needs of EcoLexicon

As shown in this chapter, the new trend in Lexicography and Terminograhy, based on Bergenholtz and Tarps' Function Theory of Lexicography, highlights the need to consider specific information needs when designing a lexicographic or terminographic tool. In this sense, any resource aimed at assisting users during the translation process should facilitate the transfer of a message from one language into another within a particular subject field. In other words, its functions involve helping translators meet their specific needs when translating within a specific subject field. In consonance with this, in order to be able to design an effective template for including phraseological information within each term entry in EcoLexicon, the specific needs of translators must be previously examined. It is thus necessary to analyze the various phases within the Translation process.

According to Tarp (2007: 241), the translation process can be divided into three phases, namely, preparation, translation, and revision (Figure 20):

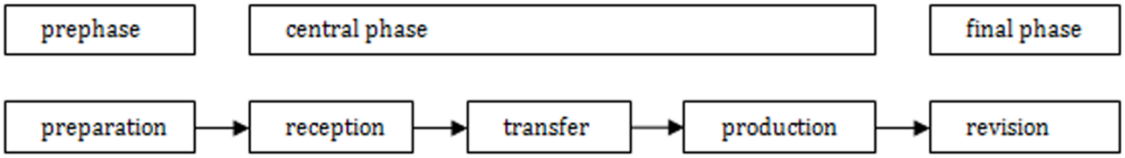


Figure 20. The translation process

The first phase is preparation. In order to carry out a translation assignment, translators need factual competence (i.e. basic knowledge about the specialised field related to the text to be translated). When translators do not have this knowledge, they frequently consult parallel texts[15] on the subject area. In this phase, it is helpful for the translator to be able to access the conceptual structure of the domain as well as its most prominent conceptual relations. EcoLexicon fulfils this need since it highlights conceptual information and provides data which help users to understand the conceptual information activated by the definition (§1.4.3.2.1) and contexts of use (§1.4.3.2.2). Besides specifying the usage contexts of a terminological unit, such data complement and complete the information given in the definition.

Secondly, the translation process itself focuses on understanding the source language (SL) text and on translating the knowledge structures encoded in its terminology. To this end, translators require a monolingual or bilingual solution, which includes explanations and/or potential translation correspondences. In EcoLexicon, the definitions of concepts can be found both in English and Spanish, along with the terms associated with each concept in the six languages represented in EcoLexicon.

The transfer phase is not the real translation, but rather an interpretation of the linguistic features of the SL text and a comparison with possible target language (TL) equivalents. In this phase, translators may need a bilingual solution, based on the terms and collocations in the SL as well as their equivalents in the TL. In the case of two or more possible equivalents, translators may need to be able to access meaning distinctions that allow them to choose the best equivalent. In this case, EcoLexicon has the advantage of providing all the terms associated with a concept in the SL as well as in the TL. Contextual information is also provided for each term.

[15] It becomes necessary to highlight here that the notion of *parallel text* within the domain of Translation differs from that of *parallel corpus* (§4.2.1.2). In Translation, a *parallel text* refers to a text either in the source or target language that is comparable to the text to be translated in terms of subject matter or text type.

The final phase is the production of the TL text. In this phase, the translator may need grammatical and syntactic information concerning terms, including collocations in the TL. The more collocations a translation dictionary contains, the better it can fulfil its function (Bergenholtz and Tarp 2010: 33). EcoLexicon provides contextual information for each term as well as information regarding the morphosyntactic features of each terminological unit (grammatical category, term type, gender). As previously mentioned, it currently lacks a systematic presentation of collocational information, which is only available via the contexts of some terms. This research remedies this deficiency by proposing a template for information regarding the verb phrase constructions associated with each term in EcoLexicon. As shall be seen in 4.1, the subdomain of Natural Hazards was chosen to test the methodology proposed in this thesis. This methodology will gradually be extended to the rest of subdomains within the Environment domain in the near future.

Since translation is composed of several phases and requires both cognitive and communicative skills, determining the possible user needs during the translation process is an extremely complex issue:

> […] boundaries between both cognitive and communicative uses are rarely clear. In order for dictionaries to be suited to particular types of users, their micro and macrostructural design should be oriented towards the cognitive-functional uses that particular user groups make of dictionaries. Evidently, making a dictionary for translators involves describing the meaning of words, their use in context, and their possible correspondences in other languages. It also entails making their position explicit (at least at some level) in the configuration of the mental lexicon. This involves considering cognitive and functional criteria in a continuum since the concepts of situation (as a set of knowledge acquisition needs) and linguistic context are intertwined (Tercedor, López and Faber 2012: 182).

In our opinion, an effective way of examining translators' needs is obviously to ask students of Translation themselves. Nielsen (2010: 70–71) highlights that this kind of study has the following shortcomings: (i) the respondents are themselves learners and, thus, have only limited experience and knowledge regarding specialized translation; (ii) the findings of each study refers to a specific text type and subject field and not to specialised texts in general and specialized translation as a whole; (iii) the number of participants is normally very small, so the results risk of not being representative.

Despite these potential drawbacks, we decided to design a questionnaire to evaluate and validate EcoLexicon for translation (López, Buendía, and García 2012). The questionnaire was completed on-line by 44 students in their 3rd year of the Degree in Translation and Interpreting of the University of Granada. The purpose of the questionnaire was to elicit their opinion about the following: (i) the contribution of EcoLexicon to the acquisition of environmental knowledge; (ii) the usefulness of EcoLexicon in the translation of specialised texts on the environment; (iii) the usability of the interface.

The results of this analysis helped us to specify the needs of the potential user consulting the phraseology section in EcoLexicon. Since in this thesis we only focus on English and Spanish, our target audience consisted of both English and Spanish native speakers, who might find themselves immersed in a number of social situations where terminographically relevant needs might occur. These include the following:

(i) cognitive situations:
   - information on subjects related to the environment;
   - information about Spanish and English environmental language usage;

(ii) communicative situations, which are the most important situations addressed:
   - reception of Spanish environmental texts;
   - production of Spanish environmental texts;
   - reception of English environmental texts;
   - production of English environmental texts;
   - translation of Spanish environmental texts into English;
   - translation of English environmental texts into Spanish

More specifically, as shall be seen, the encoding of collocations proposed facilitates their retrieval in two communicative situations:

(i) the translation of a specific collocation from L1 to L2, such as the translation of 'tropical cyclone blow out' into Spanish;

(ii) the production in L2 of a collocation in which the collocate conveys a specific meaning, i.e. the meaning "to cease to exist" in combination with 'tropical cyclone' ('blow out').

As can be deduced from these examples, a phraseological template should be multilingual, or at least bilingual, with correspondences between the languages

involved. It should include a semantic description of collocations. In addition, EcoLexicon should provide various ways of accessing collocations.

However, it is also true that very frequently the same user can perform different roles depending on the context. For example, a translator can sometimes behave as a teacher or as a linguist. In the words of Nuccorini (2003: 367):

> [...] it must be borne in mind that often the dictionary-intended addressees do not coincide with the actual users and that, on the other hand, different roles are often performed by the same individual (for example a teacher and an advanced user, a linguist and a translator) who might adopt different perspectives.

Although we agree with Nuccorini, in order to facilitate the access to information the template proposed in this thesis requires no prior linguistic knowledge. Therefore, no difficult metalanguage is used to encode and describe collocations in the template. In the future, there are plans to provide various types of information in consonance with the profile of the user.

2. Phraseology in Lexicography and Terminography

2.1 Phraseology

> [...] learners of any language, have traditionally devoted themselves to mastering words – their pronunciation, forms, and meanings. However, if they wish to acquire active mastery of English [or any language], that is, if they wish to be able to express themselves fluently and accurately in speech and writing, they must learn to cope with **the combination of words** into phrases, sentences, and texts (Benson, Benson, and Ilson 2009: xiii) (emphasis added).

> Becoming a fluent speaker involves learning an enormous inventory of expressions larger than words [...] (Langacker 2008: 19).

2.1.1 Introduction

It seems that about 80% of the words in discourse are chosen according to the co-selection principle rather than for purely syntactic or grammatical reasons (Sinclair 2000: 197). Thus, the analysis of how words co-select or combine with other words is a necessary focus of study for any linguist and, more specifically, for any translator wishing to create a text that is as natural and linguistically correct as possible. As shall be seen (§2.1.2), the combination of words is indeed the object of study of phraseology. In the words of Benson, Benson, and Ilson (2009: vii):

> To use language you must be able to combine words with other words to form phrases and to combine words into grammatical patterns to form clauses and sentences. Traditionally, the combination of words with words has been called collocation and its result has been called phraseology.

Bally ([1909] 1951) in his work *Traité de stilistique française* can be regarded asthe father of phraseology in that he was the first to distinguish between *locutions phraséologiques* [phraseological units in the broadest sense] and *unités phraséologiques* [idioms in the strictest sense].

Palmer and Hornby were the first linguists to approach phraseology from a pedagogical perspective. Palmer, as a teacher of English in Japan in the 1930s, set up a research project to collect and classify a wide range of multiword units, which he called

comings together-of-words or *collocations* (Palmer 1933: 13, as quoted in Cowie 1998c: 211):

> It is not so much the words of English or the grammar of English that makes English difficult, but that vague and undefined obstacle to progress in the learning of English consists for the most part in the existence of so many odd comings-together-of-words.

Palmer's theoretical views appear in his introduction to the *Second Interim Report on English Collocations* (1933). He defined *collocation* as a "[...] a succession of two or more words that must be learned as an integral whole, and not pieced together from its component parts" (Palmer 1933: 5). He differentiates *collocations* as a general category from what he referred to as *free phrases* or *free combinations* which combine by common rules of grammar (Palmer 1933: 5). Palmer and Hornby used the term *collocation* for the whole spectrum of word combinations. Since Palmer did not recognize a gradation of idiomaticity, he did not differentiate between the more and less idiomatic cases with the subsequent limitations of this assertion (Cowie 1998c: 211).

Although very few authors would now apply the term *collocation* to the whole range of multiword units (see §2.1.2 for the various classifications/denominations of phraseological units), Palmer and Hornby's approach greatly influenced the treatment of multiword units in learners' dictionaries of the 1930s and 1940s (Cowie 1990), and their work provided the basis for the treatment of collocations by future generations of phraseologists.

Phraseology, however, was formally established as a discipline by the Russian Vinogradov (1947). As pointed out by Cowie (1998b: 2):

> Classical Russian theory, with its later extensions and modifications, is probable the most pervasive influence at work in current phraseological studies and is unrivalled in its application to the design and compilation of dictionaries.

Since the late 1960s, the flow of ideas in phraseology has been almost entirely from East to West (Klappenbach 1968; Weinreich 1969; Lipka 1974; Aisenstadt 1979; Cowie 1981; Gläser 1986, as quoted in Cowie 1998c: 209). In Spain, interest in phraseology arose somewhat later in the 1980s, thanks to Zuluaga (1980). However, in recent decades, the scientific production in phraseology has increased dramatically. In 1981, the first international meeting on phraseology took place in Mannheim

(Germany), organized by German and Slovenian phraseologists. In 1999, the European Society of Phraseology (EUROPHRASS[16]) was created in Zurich (Switzerland) to promote scientific exchange and international cooperation within the field of phraseology.

Initially, studies in phraseology were mostly diachronic, and only focused on certain phraseological units. However, from the 1990s a new synchronic perspective arose, which better defined the discipline, established its boundaries, and provided a classification or systematization of those units. New research lines appeared, thanks to Corpus Linguistics, more powerful corpus analysis tools, and the Internet. As such, studies on comparative phraseology started to appear, as well as studies centered on the close relation between phraseology and specialized languages (Tercedor 1999; López 2000; Montero 2002, *inter alia*).

2.1.2 The notion of phraseology and phraseological unit

Generally speaking, *phraseology* is the discipline that studies phrases, "where 'phrases' means any multi-word expression up to sentence level" (Pawley 2001: 122). There have been many attempts to categorize 'phrase' (e.g. *terminological phrase*, *LSP phrase*, *phraseme*, *phraseological unit*, *phraseological term*) (Thomas 1993: 57). Other frequent names include: *multiword units; multiword lexemes; multiword lexical units; multi-word lexical phenomena; phrasemes; conventional expressions, formulae, prefabs, composites, fixed expressions, set expressions, set phrase, word combinations, phrasal lexemes*, etc. (Corpas 1997, Cowie 1998a, Pawley 2001, as quoted in Montero 2002). In continental Europe, the most widely used term is *phraseological unit.* Wray (2000: 465) gives a summary of the terms used in the literature to describe and define phraseological language (Table 6, following page).

Instead of *phraseological language*, Wray (2000: 465) uses *formulaic language* and proposes *formulaic sequence* as a term for phraseological units, defined as follows:

> A sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

[16] <http://europhras.org/> [12/01/2012].

amalgams	gambits	preassembled speech prefabricated routines and patterns
automatic	gestalt	ready-made expressions
chunks	holistic	ready-made utterances
clichés	holophrases	recurring utterances
co-ordinate constructions	idiomatic	rote
collocations	idioms	routine formulae
composites	irregular	schemata
conventionalized forms	lexical(ized) phrases	semi-preconstructed phrases that constitute single choices
F[ixed] E[xpressions] including I[dioms]	lexicalized sentence stems	sentence builders
fixed expressions	multiword units	stable and familiar expressions with specialized subsenses
formulaic language	non-compositional	stereotyped phrases
formulaic speech	non-computational	stereotypes
formulas/formulae	non-productive	stock utterances
fossilized forms	non-propositional	synthetic unanalysed chunks of speech
frozen metaphors	petrifications	
frozen phrases	praxons	

Table 6. Set of terms for phraseological language

This thesis uses the term *phraseological unit.* The question now is what a phraseological unit actually is and if it can be divided into subtypes. Cowie (1998b: 7) summarizes typologies of phraseological units as proposed by phraseologists from Vinogradov (1947) to Howarth (1996), and also includes his own classification:

Author	**General category**	**Opaque, invariable unit**	**Partially motivated unit**	**Phraseologically bound unit**
Vinogradov (1947)	Phraseological unit	Phraseological fusion	Phraseological unity	Phraseological combination
Amosova (1963)	Phraseological unit	Idiom	Idiom (not differentiated)	Phraseme, or Phraseloid
Cowie (1981)	Composite	Pure idiom	Figurative idiom	Restricted collocation
Mel'čuk (1988)	Semantic phraseme	Idiom	Idiom* (not differentiated)	
Gläser (1988)	Nomination	Idiom	Idiom (not differentiated)	Restricted collocation
Howarth (1996)	Composite unit	Pure idiom	Figurative idiom	Restricted collocation

Table 7. Subcategories of phraseological units (Cowie 1998b: 7)

As Cowie (1998b: 4) highlights, these typologies merely reflect the fact that there are various research initiatives, and that there is still no generally accepted methodology:

> In phraseology, as in other fields within linguistics, it is not uncommon for individual scholars to apply different terms to the same category (or the same term to different categories).

The Russian School (i.e. Vinogradov 1947) uses *phraseological unit*, defined as "non-motivated word-groups that cannot be freely made up in speech but are reproduced as ready-made units" (Ginzburg et al. 1979: 74). Accordingly, the ISO defines a *phraseological unit* as "any group of two or more words that form a unit, the meaning of which frequently cannot be deduced based on the combined sense of the words making up the phrase" (ISO12620:1999)[17]. According to Corpas (2003: 134), *phraseological unit* is defined as follows:

> Por unidad fraseológica se entiende una combinación estable de al menos dos palabras, que, en virtud de las distintas corrientes, tendrá como límite superior el sintagma o la oración compuesta y presentará como rasgos inherentes la fijación o la idiomaticidad por si solas, o bien una combinación de ambos criterios.

According to this definition, a *phraseological unit* is composed of two or more words, which can form a phrase or even a sentence. This combination is stable with varying level of idiomaticity. Therefore, a *phraseological unit* is characterized by the following criteria (Corpas 1998: 167):

(1) **polylexicality** (i.e. a length of two or more words);
(2) **high frequency**;
(3) **familiarity**, by reproducing a certain combination, the speakers of a language recognize it as familiar and treat it as a unit (Corpas 2001c: 44);
(4) **fixedness:** According to Penadés (2012: 7-8), there are four levels of fixedness: (i) word order (e.g. 'black and white'); (ii) singular/plural (e.g. 'raining cats and dogs', instead of a cat and a dog; (iii) number of elements in the phrase (e.g. 'kick the bucket', but not 'kick away the wooden bucket'); (iv) word form (e.g. 'mirada fija en lo alto', but not 'en la altura'). In this sense, there are degrees of

[17] <http://www.isocat.org/rest/dc/339> [12/01/2012].

fixedness since the form of certain units can be modified to a certain extent. According to Penadés (2012: 8), this degree of formal fixedness is reflected in the fact that it is possible to do the following: (i) change the order of the components of certain phraseological units (e.g. 'por un oído te entra y por otro te sale' or 'te entra por un oído y te sale por otro'; (ii) make certain elements singular or plural (e.g. 'Este hotel es del año de la pera' / 'Han utilizado un vehículo de los años de la pera'); (iii) change the number of components in the phrase (e.g. 'de narices'/ 'de tres pares de narices'); (iv) modify certain components in the phrase (e.g. 'siempre será el ojo/ojo derecho de tu padre');

(5) **idiomaticity or semantic fixedness:** the combinatorial constraints and semantic specialization of the components of a phraseological unit. Zuluaga (1980: 122) defines *idiomaticity* as a characteristic of a phrase, whose sense cannot be deduced from the meaning of its components (e.g. 'pull the wool over sb's eyes'). When this happens, the phraseological unit is idiomatic. In the same way as formal fixedness, there are also degrees of semantic fixednesss (Penadés 2012: 9);

(6) **potential variations** or linguistic variations and discursive manipulations (e.g. 'ask a question', 'address a question', 'pose a question').

Nonetheless, the boundaries of phraseological units are extremely fuzzy. For instance, as pointed out by Blasco (2000), it is very difficult to distinguish between verbal periphrases and phraseological units formed by a verb plus a noun. In line with this, there has been much discussion about the limits of the discipline. The School of Prague proposed a center-periphery model for determining the limits of inclusion of phraseological units. According to Fleischer (1997), this model has been widely used for separating extremely fixed and idiomatic phraseological units (the center) from those that only have a certain degree of fixedness (the periphery). This initially led to two different approaches to phraseology: (i) a broader conception of phraseology; (ii) a more restricted conception of phraseology (cf. Hundt 1994: 37–43; Ruiz-Gurillo 1997; Corpas 2003: 39–66).

The more restricted views regards phraseological units as fixed combinations of at least two words which appear together and function as a noun, verb, adverb, or preposition. As a result, only idiomatic expressions are contemplated (Corpas 2003: 45). In contrast, the broader approach to phraseology regards phraseological units as all

combinations of words with a certain degree of stability. Accordingly, within the center-periphery model of the School of Prague, a more restricted view of phraseology only focuses on elements in the center (i.e. sayings and proverbs), whereas the broader view also focuses on more peripheral elements, such as collocations (Corpas 2003: 44-45).

Since this distinction can be controversial (Corpas 1998), boundaries often vary depending on the author. For example, Rossenbeck (1989) states that phraseological units can only be idioms and excludes lexical phenomena at the sentence level. Zuluaga (1997) also considers that collocations are outside the scope of phraseology since collocations are merely lexical solidarities.

From a wider perspective, some authors consider phraseological units to be all word combinations that have certain stability. This approach was adopted by Hausmann (1984, 1985, 1989) and Gläser (1986), who believe that collocations are central to the study to phraseology. In fact, Gläser (1994/95: 45) proposes a generalized approach in which phraseology includes the following: (i) the inventory of phraseological units[18]; (ii) the linguistic discipline in charge of investigating the properties of phraseological units from a theoretical perspective, classifying them according to their constituent structure and codifying them in dictionaries. In a parallel way to the distinction between lexicology/lexicography, terminology/terminography (§1.3), Gläser considers phraseology both as a theoretical discipline of study as well as its practical application to phraseological units.

In line with Montero (2002), we believe it is not necessary to establish a new theoretical framework for the study of phraseological units, since this framework can be found in lexicography or terminography. Therefore, this study is limited to phraseology in its sense of word combinations rather than in its more theoretical sense. This idea is also supported by Sager (1992: 586):

> There is also a danger in attempting to set up phraseological studies as a new fashionable pursuit in contrast to terminology, in a similar way as a lot of energy was wasted at one time on proclaiming the independent status of terminology. Phraseology is not another "ology"; it is not even another practice [...]; it is simply a collective noun for a group of phrases.

[18] Gläser extended the restrictive conception of phraseological units to include not only combinations equivalent to words, but also all combinations with a certain formal stability, regardless of whether they had a certain semantic stability at the phrase or sentence level.

In the wider view of phraseology in this study and in line with Cowie's summary (see Table 7), the following combinations are considered to be *phraseological units*: (i) *collocations* (the focus of study in this thesis); (ii) *idioms*; (iii) *compounds*. Free combinations are also compared with the rest of categories.

2.1.3 Approaches to the study of collocations

Collocations are now regarded as useful information that should be included in both general language dictionaries (Benson, Benson, and Ilson 1986; Hausmann 1979; Mel'čuk et al. 1984-1999) (§2.2) and specialized dictionaries (§2.3). However, it is still a matter of debate how this should be done. For example, there is still no consensus as to which collocations should be included, where they should be placed, or how they should be classified. Not surprisingly, there are almost as many approaches to collocations as authors or types of study. However, different approaches can be divided into two groups: (i) *the semantically-based approach*, which assumes a particular meaning relationship between the constituents of a collocations; (ii) *the frequency-oriented approach*, in which collocations are evaluated on the basis of how often they appear in texts.

2.1.3.1 Semantically-based approach

As previously mentioned, the semantically-based approach assumes a certain meaning relationship between the constituents of a collocation. Its leading advocates are Mel'čuk (Mel'čuk et al. 1984-1999), Hausmann (1989), Benson, Benson, and Ilson (1986, 2009), *inter alia*. Generally speaking, these researchers conceive collocations as binary combinations, where there is a semantically autonomous *base* or *node* and a semantically dependent *collocate*. In addition, they assume directionality in the components of the collocation. Consequently, the selection of the collocate is contingent on the prior selection of the base (Hausmann 1989). This is the reason why the resources developed within semantically-based approaches classify collocations in terms of their base (§2.1.3.1).

Benson

Benson was clearly influenced by the Russian tradition since he defines collocations in terms of other word combinations. His view of collocation is based on the BBI (§2.2.3),

which he compiled along with Benson and Ilson. As such, these authors divide word combinations into the following categories (Benson, Benson, and Ilson 1986: 252–256):

(i) free combinations, defined in the BBI as "elements that are joined in accordance with the general rules of English syntax and freely allow substitution" (Benson, Benson, and Ilson 2009: xix);

(ii) idioms conceived as "relatively frozen expressions whose meanings do not reflect the meanings of their component parts" (Benson, Benson, and Ilson 1986: 252), (e.g. 'to kill two birds with one stone', i.e. to achieve two aims with one action);

(iii) collocations that are placed "between idioms, on one hand, and free combinations, on the other" (Benson, Benson, and Ilson 1986: 253), (e.g. 'commit murder'). According to the authors, 'commit murder' is not an idiom in the sense that the meaning of the whole reflects the meaning of the parts. Nor is it a free combination in that it is a frequent combination whose synonymy is restricted. This is the reason why these authors also refer to collocations as *fixed combinations* or *recurrent combinations*. As shall be seen in the description of the BBI (§2.2.3), collocations can be further subdivided into grammatical and lexical collocations. *Grammatical collocations* consist of a dominant word (a noun, adjective/participle, verb) and a preposition or a grammatical structure, such as an infinitive or clause; whereas *lexical collocations* are formed by constructions such as: verb + noun, adjective + noun, noun + verb, noun + noun, adverb + adjective, adverb + verb (Benson, Benson, and Ilson 2009: xiii);

(iv) transitional combinations, which are more transparent than idioms but less variable than collocations (e.g. 'to catch one's breath');

(v) compounds, which include many adjective + noun combinations (e.g. 'floppy disk'), noun + noun combinations (e.g. 'White House', 'night owl'), or simple verb + one or two adverbs or prepositions (e.g. 'hand in', 'carry out'). This last group also receives the name of *compound verbs* or *phrasal verbs* (Benson, Benson, and Ilson 1986: 254). The main difference with respect to collocations is that they "are completely frozen and no variations at all are possible".

According to Benson, Benson, and Ilson (1986: 254-255), the only group that should not be included in dictionaries is the group of free combinations: "free combinations should ordinarily be included in dictionaries only when they are needed to exemplify the meaning of a word, especially if it is polysemous". More specifically, they argue that the only compulsory collocations that should be included in lexicographic resources are what they call *CA collocations*, i.e. collocations that consist principally of transitive verbs denoting *creation* and/or *activation* + a noun/pronoun (e.g. 'come to an agreement') (Benson, Benson, and Ilson 2009: xxxi), and *EN collocations*, i.e. collocations which include verbs meaning *eradication* and/or *nullification* (e.g. 'reject an appeal') (Benson, Benson, and Ilson 2009: xxxii) (§2.2.3 for more details concerning their typology).

Hausmann

Hausmann was also one of first linguists to seriously study collocations. First of all, he distinguished fixed from non-fixed combinations. Fixed combinations are idioms and compounds, whereas non-fixed combinations include the following:

(i) co-creations, which are free combinations that are creatively combined by the speaker (e.g. 'une maison agréable');

(ii) collocations, which are not creatively combined but come together because of some convention, (e.g. 'ton péremptoire');

(iii) counter-creations, which are words that do not usually combine, but which are mainly found in literature and advertisements to create a special effect, (e.g. 'jour fissuré') (Hausmann 1984, as quoted in Nesselhauf 2004: 16).

According to Hausmann (1989: 1010) collocations differ from free combinations in that the word combinations in collocations are more constrained. However, they differ from idioms in that they are not fixed and their meaning is transparent. A collocation is thus defined as follows:

> La combinaison caractéristique de deux mots dans une des structures suivantes: a) substantif + adjective (épithète) b) substantif + verbe c) verbe + substantif (objet) d) verbe +adverbe e) adjective + adverbe f) substantif + (prép.) + substantive.

As such, Hausmann distinguishes between the following combinations for collocations:

(1) noun + adjective (epithet) (e.g. 'confirmed bachelor');
(2) noun + verb (e.g. 'his anger falls');
(3) verb + noun (object) (e.g. 'to withdraw money');
(4) verb + adverb (e.g. 'it is raining heavily');
(5) adjective + adverb (e.g. 'seriously injured');
(6) noun + preposition + noun (e.g. 'a gust of anger').

In line with the semantic approach to collocations, Hausmann's most remarkable contribution to the study of collocations is the fact that he conceives collocations as a binary phenomenon in which the two elements in the combination do not have the same role. As previously mentioned, the *base* (the noun in the case of verb + noun or adjective + noun), is semantically autonomous, whereas the other component (the *collocate*) is chosen with regards to the base, and thus derives its meaning from its combination with the base (Hausmann 1989: 1010):

> Dans la collocation, le status des deux partenaires combinés n'est pas égal. L'un des partenaires, par ex. *le célibataire*, est autonome sur le plan sémantique. L'autre partenaire *(endurci)* ajoute une caractérisation qui ne modifie pas l'identité du caractérisé. On appellera *base* de la collocation le partenaire caractérisé (...) et *collocatif* le partenaire caractérisant qui ne reçoit son identité sémantique que par la collocation. Le rapport base-collocatif est l'orientation de la collocation.

Hausmann (1989: 1010) also supports directionality and the fact that collocations should be listed under the base. The inclusion of the collocation under the collocate helps to test a hypothesis. However, placing the collocation in the base entry allows users to discover a collocation that they might not know or might have forgotten. According to Hausmann (1989: 1010):

> Dans l'article du collocatif (…), la mention de la collocation est plus ou moins indispensable pour compléter la définition, c'est-á-dire pour la fonction de réception. A l'article de la base (...), la mention de la collocation est sans importance pour la compréhension du mot traité. En revanche, elle est utile pour la production de textes, car le locuteur organise ses textes en allant de la base aux collocatifs et non en sens inverse. (...). Dans l'optique de la production de textes, mentionner la collocation à l'article du collocatif

> permet seulement à l'utilisateur de vérifier une hypothèse sur la collocation, (...), mais ne lui permet pas de trouver une collocation inconnue ou de retrouver une collocation oubliée.

Mel'čuk: The Meaning-Text Theory and the Explanatory Combinatorial Dictionary

The Meaning-Text Theory (MTT) (Mel'čuk and Zholkovskij 1970) originated in Russia in the 1960s. Contrary to the predominant linguistic trend of that time, the Meaning-Text Theory maintained that semantics had priority over syntax. Its basic premise is that any act of linguistic communication involves content (meaning), an oral or written signal (text), and a mapping (a set of correspondences between meaning and text) (Faber and Mairal 1999: 11). In other words, it provides the necessary information to go from an idea (the sense) to its realization in a given language (the text), specifying the semantic, syntactic, morphological and phonetic levels of linguistic description.

The Explanatory Combinatorial Dictionary (ECD) (Mel'čuk et al. 1984-1999; Mel'čuk, Clas, and Polguère 1995) is the lexical component of the MTT. The ECD encodes all the previously mentioned levels of information. Its objective is to provide a systematic and formal description of the entire set of linguistic properties of lexical units, referred to as *lexemes* or *phrasemes* which convey a specific meaning (L'Homme and Leroyer 2009: 271). The MTT also proposes an inventory of *lexical functions* (LFs), a formal system for encoding collocations within the ECD. LFs are specified for each lexical unit (L) in the dictionary. They are part of a system designed to represent a large set of lexical relations. According to Mel'čuk (1996: 39), a lexical function *f* is a correspondence that associates a given lexical unit (L) (the argument or keyword, i.e. the base of the collocation) with a set of lexical items (L_1) (the collocate) which express a specific meaning associated with *f*. It can be represented by the following formula: *f* (L) = L_1. It should be highlighted that lexical functions are not lexical units of the language, but rather correspond to metalexies.

The first ECD was created for Russian by Mel'čuk and colleagues in the sixties (Apresjan, Mel'čuk, and Žolkovsky 1969). Its purpose was the automatic generation of texts in Russian. Although, it was originally conceived for general language dictionaries, other authors, such as Frawley (1988) suggest that this model is ideal for the compilation of specialized resources since the ECD format guarantees the inclusion of all relevant information.

Mel'čuk's work continues today, primarily at the University of Montreal, where an ECD has been compiled for French (*Le dictionnaire explicatif et combinatoire du français contemporaine: Recherches lexico-sémantiques,* Mel'čuk et al. 1984-1999, four volumes, henceforth DEC). Other resources include its electronic version (the DiCo)[19] and the *Lexique Actif du Français* (Mel'čuk and Polguère 2007), a simplified version in which lexical functions are paraphrased. The *Dictionnaire fondamental de l'informatique et de l'Internet* (DiCoInfo), and the *Dictionnaire fondamental de l'environnement* (DiCoEnviro) are specialized resources that are also largely based on Explanatory and Combinatorial Dictionary models.

The ECD is "semantics-and paraphrase-based" (Mel'čuk 1998: 49). Set phrases or phrasemes (following Mel'čuk's terminology) can be found either as headwords or listed in the entries. As shall be seen, his classification of collocations also belongs to the semantic approach. In the words of Cowie (1998b: 6):

> Mel'čuk recognizes that the meaning of a two-word collocation includes 'intact' the meaning of one its constituents, but that the other component of its meaning is expressed by an element contingent on the first. This formulation, with its echoes of contextual 'binding', places Mel'čuk firmly within the classical Russian tradition.

Mel'čuk distinguishes between *free phrases* and *set phrases* or *phrasemes*. In the MTT, a collocation is considered to be a type of phraseme. His inventory of phrasemes distinguishes between the following: (i) *pragmatic phrasemes*, which include pragmatemes and idioms; (ii) *semantic phrasemes*, which include collocations and quaisi-idioms.

A *pragmateme* is when the meaning 'S' corresponds to the sum of the meanings of 'A' and 'B', but whose use depends on the pragmatic situation, e.g., when in English one says 'my emphasis', in French in order to express the same sense, the phrase is 'c'est moi qui souligne', whose literal translation into English is 'It's me who emphasizes' (Mel'čuk 1995: 178). In more formal terms, a pragmateme is defined by Mel'čuk (2006: 178) as follows:

> A pragmateme AOB of L is a set phrase composed of two lexemes A and B such that its signified 'AOB" is not unrestrictedly — although regularly — constructed on the basis of

[19] The DiCo is accessible through its search interface called DiCouèbe: <http://olst.ling.umontreal.ca/dicouebe/index.php> [20/04/11].

> the given ConceptR (of an extralinguistic situation SIT that the speaker wants to verbalize) out of the signifieds 'A" and 'B" of the lexemes A and B of L. ['AOB" is a regular sum of 'A" and 'B" but it cannot be replaced by any (fully or partially) equivalent signified 'X", which in principle can be constructed for SIT by rules of L; 'AOB" is determined, or bound, by ConceptR(SIT).]"

A *full phraseme* or *idiom* is when the meaning of 'X' is independent of the meanings of its constituents 'A' and 'B' (e.g. 'to rain cats and dogs') (Mel'čuk 1995: 180):

> An idiom AB of L is a semantic phraseme such that its signified 'X" does not include either of the signifieds 'A" and 'B" of A and B in a dominant position.

A *semi-phraseme* or *collocation* is given when the meaning 'X' includes the sense of one of its components (Mel'čuk 1995: 180):

> A collocation AB of L is a semantic phraseme of L such that its signified 'X" is constructed out of the signified of the one of its two constituent lexemes — say, of A — and a signified 'C" [so that 'X" = 'AOC"] such that the lexeme B expresses 'C" contingent on A".

The lexeme A, which keeps its signified intact in the dominant position within the signified of the collocation and determines the expression of C by B, is the argument of the corresponding LF. It is called the keyword of the LF or of the collocation (see §2.1.3.1 for more information regarding lexical functions). As a general rule, the components of a collocation are related by an LF. Some collocations, however, go beyond the range of LFs, namely, those in which an actant of the keyword L is expressed in a phraseologically bound way (Barrios 2010: 50–55). For instance, in 'sick leave', (i.e. leave because of illness, or in 'maternity leave', the lexical units 'sick' and 'maternity' are phraseologically bound expressions of an actant of the noun 'leave'. All such collocations are covered not by LFs, but by the government pattern of the keyword.

According to the MTT definition of collocation, units such as 'eat fish', 'play a game', 'fire burns', 'fly on a plane' are not regarded as collocations since each unit ('eat', 'fish', 'play', 'game') keeps its meaning, and 'fish' or 'game' are not contingent on the prior selection of 'eat' or 'play'. However, they are considered to be collocations

within the MTT since their components are related by an LF. Generally speaking, in the MTT, syntagmatic lexical functions are associated with collocations. However, there are also collocations that are regarded as collocations by the MTT and are not covered by LFs, but rather by the government pattern of the keyword. This is the case of those in which an actant of the keyword is expressed in a phraseologically bound way (e.g. 'sick leave' or 'maternity leave', i.e. leave because of illness or child bearing (Mel'čuk 1995: 185).

Finally, a *quasi-phraseme* or *quasi-idiom* is coined when the meaning 'X' includes the meanings of its components 'A' and 'B', but neither 'A' nor 'B' are the semantic nucleus (e.g. 'give the breast', whose nucleus is neither give nor breast) (Mel'čuk 1995: 181-186):

> A quasi-idiom AB of L is a semantic phraseme such that it satisfies simultaneously the following two conditions:
> 1. Its signified 'X" includes the signifieds 'A" and 'B" of the two constituent lexemes.
> 2. a. Either 'X" includes a further signified 'C" different from 'A" and 'B"; b. or 'X" includes just the signifieds 'A" and 'B", but the one in the dominant position corresponds to the syntactically dependent lexeme.

The six main formal properties of an Explanatory Combinatorial Dictionary are the following: (1) it is a *theoretical dictionary*, which is elaborated within a coherent linguistic theory with semantic, syntactic, and morphological information, and which highlights the importance of the lexicon; (2) it is an *active dictionary* in that it aims at production; (3) it is a *semantic dictionary*, based on semantic representations of all the expressions it contains, and where the definitions are the major part of any lexical entry; (4) it is a *combinatorial dictionary*, allocating syntactic and lexical co-occurrences; (5) it is a *formalized dictionary*, as it can be regarded as a lexical database; (6) it is an *exhaustive dictionary,* in that each entry includes all a native speaker knows about the lexical unit in question (Mel'čuk 1998: 50).

Lexical entries are divided into three zones, namely, the *semantic zone*, the *syntactic zone,* and the *lexical combinatorial zone* (Mel'čuk 1998: 50-53). *The semantic zone* formulates the definition, which is in propositional form with variables for semantic actants and which is the result of a strict decomposition of meaning. Generally speaking, *semantic actants* are the arguments or participants associated with the predicate. Predicates are verbs and adjectives. Nouns, especially deverbal and de-

adjectival nouns, can also be predicates (L'Homme 2010: 142). Mel'čuk (2004: 8-9) defines the notion of semantic actant[20] and lists the different variables used to represent them in ECD dictionaries as follows:

> Informally speaking, a semantic actant [SemA] of an LU L that has a predicative meaning {L(; ; . . . ;) corresponds to an argument of {L. […] In the lexicographic definition of L, which is a semantic decomposition of the meaning {L, all L's SemA-slots are represented by variables: X, Y, Z, etc.

In dictionaries, actants can appear in different parts of the entry. They are most frequently stated in the definition and in the representation of the syntactic structure of lexical units. Table 8 shows the representation of the predicate 'inviter' [to invite] in the DEC, i.e. the French version of the ECD (Mel'čuk et al. 1984-1999):

inviter, verbe.
1. X prie la Y de se déplacer…pour faire des actions [*Paul m'invite à aller au restaurant*]
2a. Dans le but d'inciter Y...X fait savoir à Y... [*Il nous invite au silence*]
2b. Fait X incite Y à Z-er... [*Ses promesses m'invitaient à redoubler de zèle*]
3. ...X cause que Y a envie de Z-er... [*Le calme de cette maison invite au repos*]

Table 8. Entry of 'inviter' in the DEC

However, actants can also be listed in other parts of the article. For example, in dictionaries dealing with collocations, actants are often stated both in the syntactic structure of collocations and in the explanation of their meaning (L'Homme 2010: 143).

The *syntactic zone* of an entry contains the government patterns, specified for each semantic actant of the headword and the corresponding syntactic actant. It also lists all surface means of expressing it in the text. As L'Homme (2010: 143) points out, most lexicographers agree that actants are a very useful way of providing users with information regarding the syntactic-semantic interface of the lexical unit, but the challenge is to present them in a user-friendly way.

Finally, the *lexical co-occurrence zone* specifies all the paradigmatic and syntagmatic associations of the entry in the form of *lexical functions*. The meaning of an

[20] For more information concerning actants, cf. Mel'čuk (2004).

LF is abstract and general, and can be associated with a large number of values (L'Homme and Leroyer 2009: 271). There is a set of approximately 60 standard lexical functions divided into paradigmatic and syntagmatic functions (Mel'čuk 1998; Mel'čuk, Clas, and Polguère 1995; Wanner 1996).

Paradigmatic lexical functions deal with *selection* (Mel'čuk 1998: 34). In other words, they describe the existing relations between lexical units connected in the same semantic paradigm and which can be interchangeable in certain contexts. More specifically, they model the following: (1) the main semantic relations, such as synonymy, antonymy, hyperonymy; (2) the syntactic derivatives, such as nominalisations; (3) the semantic derivatives.

Syntagmatic lexical functions deal with *combination* (Mel'čuk 1998: 34), and describe combinatorial relations. This is the reason why syntagmatic relations are of major interest for phraseology. They are normally presented according to part of speech and the syntactic role of the collocate. Some of the most common syntagmatic lexical functions are the following: **Magn**, **Ver** and **Bon**, for adjective and adverbial values, and $\mathbf{Oper_i}$, $\mathbf{Func_i}$ and $\mathbf{Labor_{ij}}$, for verbal ones (Orliac 2004: 46).

Magn is an intensifier lexical function which associates with a lexical unit (L), all the lexical units expressing intensification, i.e. the general sense of *very* (e.g. **Magn**(*s'étonner*) = *beaucoup*, *grandement*, *profondement*). **Ver** associates with a lexical unit (L) the lexical units that express the sense *as it must be made*. It is a kind of objective modifier of L (e.g. **Ver**(*haine*) =*justifiée*, *fondée*). **Bon** associates with a lexical unit (L) all the lexical units that express the meaning of *good*. They are subjective modifiers of L (e.g. $\mathbf{Bon_1}$(*invitation*) = *aimable*, *cordiale*).

$\mathbf{Oper_i}$, $\mathbf{Func_i}$ and $\mathbf{Labor_{ij}}$ model support verbs[21] (Gross 1981). The values of these three lexical functions are semantically empty verbs. The lexical unit contained is necessarily a noun, whose meaning is or includes a predicate, i.e. an action, an activity, a state, a property, a relation, etc. (Mel'čuk 1998: 37). Support verb lexical functions have a purely syntactic purpose. They link a predicative lexical unit (L) by means of a support verb to the semantic actants of L. Support verb lexical functions are differentiated by means of the syntactic role played by L and its actants. $\mathbf{Oper_i}$ associates with a predicative lexical unit (L), the support verbs that take the first actant of L as subject and L as first complement (e.g. $\mathbf{Oper_1}$(*analysis*) = carry out, which is

[21]See §2.1 for more information concerning support verbs.

realized by the sentence, '*John carries out the analysis of the phenomenon*[22]*'*). **Func$_i$** associates with a predicative lexical unit (L), the set of support verbs that take L as subject and which take the first actant of L as first complement (e.g. **Func$_1$**(*analysis*) = *is due*, which is realized by the sentence, *'The analysis of this phenomenon is due to John'*). **Labor$_{ij}$** associates with a predicative lexical unit (L) the set of support verbs that take the first actant of L as subject, the first actant of L as first complement, and L as second complement (e.g. **Labor$_{12}$** (*analysis*) = *submit*, in a sentence such as 'John submits this phenomenon to a (careful) analysis*'*).

As shown, not all LFs describe collocations. In addition, LFs can also be combined to account for more complex meanings. For instance, example 1 shows the relations of **Incep** (denoting the beginning), **Pred** (which means 'to be and X') and **Plus** (more):

(1) **IncepPredPlus**(*joie) = grandir, s'accroître, augmenter*

Regarding collocations, lexical functions encode three types of linguistic properties (L'Homme and Leroyer 2009: 271-272): (i) the syntactic relationship between the base and the collocate; (ii) the actantial structure of the base; (iii) the meaning of the collocate. This is better shown in the description of the *Dictionnaire fondamental de l'informatique et de l'Internet (DiCoInfo)* (§2.3.2), and in the *Dictionnaire fondamental de l'environnement (DiCoEnviro)* (§2.3.3).

Table 9 shows the entry for 'envie' [desire/urge] in the DEC. As shown, the different senses of the term are first displayed. Then, the actantial structure of the first sense is provided, followed by the lexical functions, and usage examples:

[22]All the sentences displayed for support verbs have been extracted from the examples contained in Mel'čuk (1998: 39).

ENVIE, nom, fém

1. Désir…[l'envie de Pierre de faire un voyage]

2. Émotion désagréable... [l'envie de Pierre envers son frère]

1. *Envie de X de Y-er* = Désir de X que Y ait lieu causé par le fait que X prend conscience de son besoin de Y.

Régime

1= X	2= Y
1. *de* N	1. *de* N
2. A_{poss}	2. *de* V_{inf}
3. A	3. *que* $PROP_{subj.}$

C1: l'envie de Pierre, son envie, l'envie populaire

C2: l'envie d'une nouvelle auto, l'envie de faire un voyage, l'envie que son fils réussisse

C1 + C2 : arrive au plus vite, l'envie populaire que le gouvernement change

Fonctions lexicales

Syn:	désir
$Syn_{\cap}$:	convoitise 1 ; goût, inclination, **litt** appétance
[...]	
Magn :	grande ; irrésisteible ; folle, furieuse, brülante ; terrible
AntiVer + Magn :	démesurée
$Magn^{temp}$:	constante
$AntiMagn^{temp}$:	courte, passagère
IncepPredPlus :	s'accroître, se développer, augmenter
CausePredPlus :	attiser, accroître [ART ~] [*Ces paroles ont attisé <accru> son envie de partir*]
IncepPredMinus :	diminuer, faiblir
$FinFunc_0$ ou IncepPredMinus :	se calmer
[...]	

Examples

La brûlante envie des distinctions imaginaires.[Ch. Baudelaire]. On se confie le plus souvent par vanité, par envie de parler [La Rochefoucauld]. J'ai envie de pleurer. D'habitude ses envies passaient très vite. Pourquoi me faire envie des choses que vous ne pouvez me donner? il lui venait une envie de se lever.

Table 9. Extract of the entry 'envie' from the DEC

Evidently, the ECD approach has many advantages in that it is both comprehensive, and systematic. However, it has certain shortcomings, as pointed out by Montero (2002: 184), who criticizes its metalanguage, which should be more user-friendly.

Lexical functions have also been used in machine translation and text generation. In fact, Mel'čuk (1998: 44) highlights the usefulness of LFs for the translation of collocations:

> Suppose that we have a system of Automatic Translation […] suppose furthermore that we are interested in the translation of collocations. In such a case, it suffices to reduce the source-language collocation to its LF representation, then translate the keyword only, and, finally, to select the value of the LF for the equivalent of the keyword in the target language (Mel'čuk 1998: 44).

Supporters of this view clearly assume that the most relevant aspects of the sense of a collocate are contained in the LF, and therefore, that the LF guarantees the accuracy of translations. However, as Montero (2002: 185) stresses, the problem is the excessive generality of LFs. This means that the meaning of a collocate is partly reduced to the meaning of the LF. This is not problematic if it is assumed that LFs always contain the same values. However, this is not always the case, as shown in example (2) for 'enthousiasme' [enthusiasm], where different possibilities are offered ('grand', 'bel', 'immense'). This means that it is difficult to differentiate between the possible realizations of an argument. Lexical functions thus do not provide the necessary information to select the best equivalent when several possibilities are given.

(2) **Magn**(enthousiasme) = grand, bel, immense, énorme; déchaîné, délirant …

Thirdly, we agree with Montero (2002: 185) who states that since LFs lack basic conceptual structure, they can be redundant. If a hierarchy were established, relations would be inherited. Therefore, the formalization of all possible conceptual relations would not be necessary. Mel'čuk (1998: 42) mentions the possibility of implementing lexical inheritance in an ECD in order to prevent redundancy:

> […] in many cases a given LF has the same values for quite a few different keywords, the reason often being semantic proximity: semantically related LUs can possess the same values for a given LF. This fact can be accounted for by following the general principle of lexical inheritance (Mel'čuk and Wanner 1996): PRINCIPLE OF LEXICAL INHERITANCE

> All lexicographic data shared by a family of semantically related LUs should be stored just once –under one LU of the corresponding vocable or under the generic LU of the corresponding semantic field, from where these data are 'inherited' in each particular case.

Lexical units that belong to the same semantic field may have the same values for the same LFs (Mel'čuk 1996: 77). Therefore, insofar as lexical inheritance is concerned, it is simply a question of specifying the semantic field, determining the generic lexeme, and introducing the semantic dimensions, based on shared semantic features. Once these features are extracted from the values of the LFs in the entries, they can be transferred to the lexical entry of the generic lexeme. Each element is thus enriched with semantic information justifying its use with the lexeme from which it was extracted. In addition, the values of the LF listed in the entries of the generic lexemes are inherited by all the lexemes in the semantic dimensions. Accordingly, it is necessary to reorganize the lexical entry of the generic lexeme by dividing it into two separate parts (Mel'čuk 1996: 78): (i) a *private subentry* that describes the lexeme's syntax and cooccurrence; (ii) a *public subentry* containing the shared features extracted.

The MTT has been and still is one of the most influential meaning-based theories in contemporary linguistics. A large number of lexicographic and terminographic resources on the market today are based on the methodology proposed by Mel'čuk. For example, for general language, *BBI Dictionary of English Word Combinations*, *Redes. Diccionario combinatorio del español contemporáneo*, *Diccionario combinatorio práctico del español contemporáneo*, or *Diccionario de colocaciones del español* (§2.2.3), and within specialized language (§2.3), *Lexique de cooccurrents—Bourse et conjuncture économique*, *Dictionnaire d'apprentissage du français des affaires (DAFA)*, *DiCoInfo* and *DiCoEnviro.* The Lexical Constructional Model (§3.1.3.4) also includes aspects of lexical functions within the Meaning-Text Theory.

2.1.3.2 Frequency-oriented approach

The frequency-oriented approach to collocations is represented by Sinclair and his followers. It has its roots in British contextualism as well as in the work of Firth, carried on by Halliday and Sinclair. In the words of Cowie (1998b: 3):

> Sinclair has played a leading role in developing corpus linguistics in Britain as a basis for both phraseological research and dictionary-making.

Generally speaking, authors within this approach conceive collocations as statistically significant co-ocurrences of two or more words. In other words, collocations within the frequency-oriented approach are regarded as such because of their frequency rather than for semantic reasons. Firth's concept of collocation is based on his *Theory of Meaning*, which distinguishes three levels of meaning: meaning by collocation, the conceptual or idea approach to the meaning of words, and contextual meaning. As such, he defines *meaning by collocation* as follows:

> […] an abstraction at the syntagmatic level and it is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night* (Firth 1957: 196).

Indeed, one of Firth's most revolutionary statements is that lexical relations are syntagmatic rather than paradigmatic. For example, part of the meaning of a word (e.g. 'night') is the fact that it collocates with another word. However, as Lyons (1977: 612) underlines, Firth apparently never proposed a definition for *collocation.* According to Nesselhauf (2004: 2–5), he uses the umbrella term *collocation* for naming four concepts: (i) co-occurrences of words in general (frequent or rare); (2) habitual (= frequent) co-occurrences of words; (3) a number of consecutive words; (4) an order of *mutual expectancy*. Although Firth (1957) does not deny that words have lexical meaning, his idea was to analyze meaning in context on the basis of word associations at the syntagmatic level. This idea of meaning is also supported by Wittgenstein (1968: 80) "the meaning of a word is its meaning in the language".

Halliday (1966) regarded collocations as syntagmatic associations of lexical items of all probabilities which occur at a certain distance from each other. He defines *probability* as "the frequency of the item in a stated environment relative to its total frequency of occurrence" (Halliday 1966: 156), but he does not clearly define the concept of *distance*. Halliday also follows Firth's (1968: 99) assertion that "collocation states the habitual company a key-word keeps". Since his main objective was to discover the combinations that were most frequent in a language, it became necessary to compile large corpora, which permitted the extraction of those combinations (§4.2.3). In this sense, the lexical approach initiated by Firth became a frequency-oriented or computational approach further developed by Sinclair (1991).

Within this context, basic notions in the study of collocations include the following: (i) the *node* or word under study; (ii) the *collocate*, or the word that can combine with the node; (iii) *collocational span* or the number of words to the right and to the left that separate the node from the collocate; (iv) the *collocational range* or the set of all the collocates that can appear in combination with the node. The only requirement is that the node and collocate should be syntactically related. When this happens, and when they appear more frequently together than chance would predict, then they are considered collocations. In addition, the notions of *node* and *collocate* are in consonance with the notions of *base* and *collocate* of the semantic approach to collocation. The basic tool in the study of collocations is thus the *concordance* or key word in context (KWIC).

It is true that these approaches have also been harshly criticized. Corpas (2001c: 50–51) summarizes their limitations as follows:

(i) there are combinations that are very frequent in a language, but which are not stable enough to be regarded as collocations;
(ii) there are collocations that are stable, but whose collocates are very infrequent words, and therefore, do not appear in a corpus;
(iii) there are collocations whose elements are separated over a large span, and therefore, cannot be extracted automatically;
(iv) statistical frequency cannot account for distinctiveness or salience of certain very stable combinations that are basic to a language;
(v) corpus analysis tools are not designed to detect collocations at the lexematic level, only at the wordform level;
(vi) statistical approaches do not have any instruments for the semantic analysis of a certain collocation;
(vii) the collocational node between two lexemes is sometimes inherited by a whole phrase that can contain a word with the same lexical meaning, but which belongs to a different grammatical category, or which can point to a collocate that can be elicited in discourse.

However, it is undeniable that the automatic extraction of collocations from a large corpus of texts has indeed revolutionized the study of collocations. Almost all lexicographic and terminographic resources compiled now rely on information extracted from corpora.

The results of this thesis are also based on corpus data. As shall be seen in 2.1.4, our approach combines the two approaches to collocation. In other words, we assume that there is a meaning relationship between the constituents of a collocation (semantically-based approach), but at the same time, collocations are also regarded as high-frequency word combinations.

2.1.4 Collocation: definition and access

> Why is collocation so important? Firstly, it is a central feature of language, and –whether you are speaking or writing –it is just as important as grammar. Getting the grammar right is an essential part of producing text which is free of errors. But selecting appropriate collocations is one of the keys to sounding natural and fluent. […] Secondly, collocation contributes to meaning. Most common words in English have more than one meaning, and we use the surrounding context to indicate (or work out) which meaning is intended. Collocations play a big part in this process (Rundell 2010: vii).

Collocations are extremely important for both encoding and decoding purposes. As Rundell states (2010: vii), collocations are as important as grammar and they are what makes speakers sound natural and fluent. Nevertheless, to date, there is no general agreement on what exactly a collocation is. Although it is widely acknowledged that collocations deal with co-selection, this phenomenon can also be applied to other word combinations, such as compounds or idioms, among others. As Fontanelle (1998: 191) observes:

> A vague definition would be to say that collocations are groups of words which frequently occur in combination with each other. This definition is totally unsatisfactory, since it says nothing about the number of elements involved, the degree of frequency of occurrence, or the classes of words which can combine. In the literature, it is generally agreed that collocations are different from idioms, although the latter could also be defined as 'groups of words which frequently occur in combination with each other.

Generally speaking, collocations tend to be *binary units* since the most common type of collocation is a combination of noun + verb. However, there are also three-word collocations (Siepmann 2005: 417):

> [...] collocations consisting of three items or syntactic 'slots' are in fact quite common. This is particularly true of collocations involving neither a human subject nor a human object, such as *experience +marquer+ vie*.

As such, defining a collocation as a binary unit is not totally accurate. It is preferable to say that collocations can be composed of two or more words.

As previously mentioned, this research integrates the semantically-based approach and the frequency-oriented approach. Accordingly, in line with Mel'čuk's MTT, it is our assertion that what distinguishes combinations such as 'eat fish' and 'drop the fish' are the following: (i) the definition of 'fish' (an animal which lives in water, is covered with scales, and which breathes by taking water in through its mouth, or the flesh of these animals eaten as food); (ii) the arguments allowed by 'eat' (to put or take food into the mouth, chew it, and swallow it[23]). As such 'fish', as an edible food, can appear with a verb of ingestion, whereas the argument of 'eat' is something that can be eaten, i.e. *food*. However, the definition of 'fish' makes no reference to verbs such as 'drop', for instance. Therefore, the combination 'drop the fish' is a free combination, whereas 'eat fish' is a collocation. Therefore, combinations, such as 'eat fish', which many authors consider free combinations, are regarded as collocations in our integrated approach.

In addition, our approach also includes a certain degree of compositionality in that each lexical unit in a collocation retains its meaning. As such, in the collocations 'hurricane damages' or 'hurricane hits', both elements in the collocation maintain their respective meanings. Nonetheless, they are not free combinations since the verb is imposed by the meaning of the noun, and at the same time, the verb selects its arguments. This means that such combinations have a high frequency in texts referring to certain frames of situations.

Based on the previous considerations and for the purposes of this study, a collocation is a combination of two or more words. In our case, the structure of such highly frequent combinations are noun + verb or verb + noun, where the noun is the *base* and the verb is the *collocate*. In this noun-centered type of collocation, the meaning of the collocate (the verb) is imposed by the meaning of the base (the noun), but at the same time, the collocate also constrains the arguments that can combine with it.

[23] The definition of 'fish' and 'eat' come from *Cambridge Dictionary Online:* <http://dictionary.cambridge.org/> [12/02/2012].

Another important question to be addressed is where collocations should appear in the micro and macrostructure of both general and specialized dictionaries so that information can be easily accessed by users. According to Fontanelle (1998: 295), the user should be able to access collocational information in various ways. However, most resources only allow users to access collocations in the entry of the base:

> Multiple access points enable the linguist to access information via the base (the keyword, or element in italics), the collocate (the dictionary headword), the lexical function, the translation, or the part of speech, all of which means that this type of collocational dictionary can fill a gap in lexicography: indeed, traditional collocational dictionaries usually provide one type of access key only or simply list potential collocates without any attempt at formalizing the main collocational description of a lexeme can then be viewed as a semantic network in which a node (the base) is linked to its collocates through arrows (the lexical functions) which are labels for specific lexical semantic relationships. Potential applications of such a database range from language teaching to information retrieval and machine translation and, more specifically, language generation, where selecting the appropriate word in context is of crucial importance (Fontanelle 1998: 295).

Evidence of the truth of Fontanelle's words are shown in sections 2.2 and 2.3. As shown, even though the various lexicographical and terminographic resources differ in their methods of listing and representing collocations, most of them coincide in only affording users to access phraseological information in the entry of the base. According to L'Homme (2009: 237), when there are various ways of accessing the information, this enhances language teaching, information retrieval, and language generation:

> The access to collocations […] can help solve problems related to the production of correct lexical units to be used in combination with a specific term or another lexical unit.

It is true that most authors write that collocations should be listed in the entry of the base term. This is indeed the approach followed by the MTT, and in many dictionaries, as shall be seen in 2.2 and 2.3. Accordingly, most authors within the MTT affirm that the speaker selects the base of the collocation, and this base determines the selection of the collocate. Others, such as Bosque (2004, 2006), author of REDES and PRACTICO, believe that predicates semantically constrain their arguments, and thus, it is the collocate that selects its base. Collocations should therefore be inserted in the corresponding collocate entry within the dictionary.

In our opinion, both positions are complementary. On the one hand, users normally select a collocate to go with a certain base term. However, at the same time, a predicate imposes certain restrictions on its arguments. In the collocation, 'fire burns', the predicate 'burn' only admits noun phrases designating combustible entities. In addition, 'fire' requires a verb designating a combustion process ('burn').

As such, in this thesis, the proposed phraseological module for EcoLexicon enables users to access collocational information both via the base, and collocate. For instance, 'hurricane strikes', can either be accessed in the corresponding entry for 'hurricane', or by searching for the verb 'strike'. This method of encoding phraseology is helpful both for the translation of a collocation from L1 to L2 (e.g. the Spanish translation for 'the hurricane strikes') and for the production in L2 of a collocation in which the collocate conveys a specific meaning, i.e. the meaning of "to come against sth with force" as applied to 'hurricane'.

2.1.4.1 Collocations vs. Free phrases

As previously mentioned, Palmer in the 1930s distinguished between *collocations*, as a general category, and what he referred to as *free phrases* or *free combinations*, which are combinations that "could be put together by dint of the application of the commonest and best-known rules of grammar" (Palmer 1933: 5). In line with this, Benson, Benson, and Ilson (2009: xix), conceive *free phrases* as "elements that are joined in accordance with the general rules of English syntax and freely allow substitution", because "in English a verb may be followed by adverbials (of time, place, and manner)". Thus, the resulting number of possible combinations is limitless. Therefore, free lexical combinations are the following:

> […] those in which the two elements do not repeatedly co-occur; the elements are not bound specifically to each other; they occur with other lexical items freely. Thus, a construction such as *condemn murder* is a free combination. The verb condemn occurs with an unlimited number of nouns: *they condemned – the abduction, abortion, abuse of power, the acquittal*, etc. In a similar manner, *murder* combines freely with hundreds of verbs: *abhor, accept, acclaim, advocate*, etc. (Benson, Benson, and Ilson 2009: xxxi).

As such, a *free phrase* or a *free combination* is a combination of words that follows the general rules of syntax and grammar of a language and whose elements allow substitution. In line with Hausmann (1989: 1010), we also believe that the two

premises that distinguish free phrases from collocations are: (i) the restricted commutability of their components; (ii) the frequency of co-occurrence.

The restricted commutability of the components of a combination refers to the fact that the elements cannot be free replaced by others. For example, in the collocation 'strong wind', 'wind' can only combine with a certain set of adjectives to express a wind of high intensity, i.e. 'fierce wind', 'stiff wind', 'terrible wind'. However, in free phrases like 'strong boy', both elements can be replaced by other words freely. Finally, the other feature that distinguishes collocations from free combinations is the frequency of appearance as was defended by the frequency-oriented approach.

2.1.4.2 Collocations vs. Idioms

Idioms are defined by Langacker (2008: 18) as "fixed expressions whose meanings are not predictable from their parts". Fontanelle (1998: 191) also gives the following definition of idiom:

> Idioms are part of the larger class of relatively fixed multiword units. They are frequently described as semantically opaque word-combinations-that is, combinations whose global meaning is different from the sum of the individual meaning of the constituent parts.

As such, an *idiom* is regarded as a semantically opaque combination of words, whose meaning cannot be inferred from the meaning of the words that make up the combination. Succinctly put, collocations can be differentiated from idioms by their semantic transparency, which contrasts with the semantic opacity of idioms. In the example of strong wind', the resulting meaning of the combination is a current of air ('wind') of great force ('high'). However, in 'take the wind out of sb's sails' (to make someone feel less confident or less determined to do something, usually by saying or doing something that they are not expecting), the meaning has nothing to do with the meanings of the various words in the combination.

As Fontanelle (1998: 191) argues, in a continuum, idioms are found at one extreme, free phrases at the other extreme, and collocations in the middle:

> [...] idioms are to be found at one extreme of a continuum ranging from totally free combinations of words to completely frozen, fixed multiword units. Collocations will be found in the fuzzy area half-way between free combinations and idioms.

Other authors (Fraser 1970; Michiels 1975, *inter alia*) state that another common feature that distinguishes idioms from collocations is the resistance of idioms to a syntactic manipulations (passivization, pronominalization, fronting, clefting, insertion of material, etc.). In line with this, Carter (1987) gives various examples of idioms (e.g. 'it's raining cats and dogs'), which are immutable insofar as they cannot be in passive voice; they do not allow insertion; and they have a fixed order.

2.1.4.2 Collocations vs. Compounds

Compounds are often defined as "one word (in the sense of lexeme) that is made up of two other words (in the sense of a lexeme)" (Bauer 1988: 65). In other words, compounds are thought to designate a single concept. As such, Sager (1997: 34) underlines that the meaning of a compound is independent of its components:

> A compound is a combination of two or more words into a new syntagmatic unit with a new meaning independent of the constituent parts.

In line with this, some of the basic characteristics of compounds identified by Sager (1997: 34-35) are the following: (i) when there are two elements in a compound, the first element normally determines the second, which is the *nucleus* (e.g. 'water load', 'canal bed', 'damp course'); (ii) compounds can be inserted into other combinations for new compounds (e.g. 'rock-type flood'); (iii) depending on the nature of the nucleus, there is a difference between compounds that can designate objects (e.g. 'concrete breaker'), processes (e.g. 'concrete casting'), and or properties (e.g. 'concrete stability').

According to Sager (1997: 35), the majority of compounds are nominal compounds, and thus can be used to build terminological conceptual systems:

> Noun compounds contribute to the building of terminological systems. The nucleus of the compound, usually the last element can then indicate the category to which the concept belongs and the determinant indicates the criterion for the subdivision of the category.

As a result, nominal compounds in English are either by noun + noun or adjective + noun. This is similar to collocations, which have a similar structure. For this reason, it is often difficult to distinguish between collocations and compounds. This has been the

subject of much debate, and as yet there is no consensus of opinion. In the words of Heid (2001: 791):

> We are not aware of any broadly agreed standard for distinguishing noun-noun and adjective-noun collocations from multiword terms. And often not the classification of the phenomena, but the additional lexical and terminological description is what really matters.

In this regard, there are even authors who question whether a distinction between collocations and compounds is even necessary. Within this scenario, Meyer and Mackintosh (1996: 3) coin the term *phraseme* to refer to both collocations and compounds:

> […] we will take phrasemes to include both collocations and compounds. We realize that these are different, in that normally a compound designates a single concept while a collocation does not. However, compounds and collocations are both realizations of terminological word combinations. Furthermore, they share important relations to the conceptual structure of domains.

In the verb entry proposal outlined in this thesis, the distinction between collocations and compounds does not entail any major difficulties since the noun + verb and verb + noun combinations are always collocations. However, the problem arises with noun + noun and adjective + noun combinations, which are not the object of our study.

2.2 Phraseology in lexicographic resources

Evidence that collocations are of paramount importance for lexicographers is their central role in monolingual and bilingual general dictionaries. There are also specific resources devoted exclusively to the description of collocations. This section describes how English and Spanish lexicographic resources encode collocations in their entries. Dictionaries are thus categorized as follows: (i) monolingual general dictionaries; (ii) bilingual general dictionaries; (iii) collocations or combinatory dictionaries.

The headwords ‘bed’, for English resources, and ‘cama’ [bed] for the Spanish counterpart are used as examples to describe each lexicographic resource. ‘Bed’ and ‘cama’ were selected since they are very common words in both languages, and appear

in the various resources described here. They also combine with a large number of verbs and participate in many word combinations.

Despite the fact that this thesis focuses on the description of verb phrases, the reason for choosing a noun instead of a verb for describing phraseological information in lexicographic resources is that the majority of resources do not offer phraseological information under verb headwords, but rather opt for including them under the base, i.e. under the noun or adjective with which they combine. One exception is *Redes*, the combinatory dictionary compiled by Bosque (2004). However, generally speaking, lexicographic and terminographic resources tend to place collocations in the noun or adjective entry and not in the verb entry.

2.2.1 Phraseology in monolingual general dictionaries

Monolingual dictionaries are mainly used for obtaining information about the meaning of a word and its pronunciation. They also inform users that their knowledge of a word is correct/incorrect (Montero 2002: 163). Entries in monolingual dictionaries are generally alphabetically organized, and the lemmas can be formed either by a monolexical or polylexical unit. For each entry, etymological, phonetic, and grammatical information is normally provided, along with a definition for each of the various senses of the word. Sometimes, usage examples are also included, which show the word in context.

Collocations are also provided in monolingual dictionaries. As shall be seen, collocations are either explicitly included in examples or deduced from the definitions. The Spanish examples come from the *Diccionario de la Real Academia Española* (DRAE) and the *Diccionario de Uso del Español* (DUE). The English examples are taken from the *Collins Cobuild Dictionary of English*[24] (COBUILD).

As is well known, the DRAE is the normative dictionary for Spanish compiled by the Spanish Royal Academy of Language *(Real Academia Española)*. The first issue dates from 1780, and the most recent is the 22nd edition (2001), which can be accessed online[25]. A new version is about to come up in 2013. In contrast, the DUE was compiled by María Moliner (1966-1967, 1998, 2007). It is not freely available online, but the last two issues include a CD-ROM.

[24] Cf. Calderón (1994) for a fine-grained analysis of collocations in monolingual dictionaries.
[25] <http://www.rae.es> [13/04/2012].

Although the DUE contains a similar number of words to the DRAE, its success resides in the fact that it is constantly being updated (Seco 1981: 36). Its novelty resides in the following three features (Seco 1981: 36): (i) it is both an encoding and decoding dictionary; (ii) its lexicon includes words, as well as usual and less usual expressions so that users can choose the expression that they prefer; (iii) it has extensively revised the definitions in the DRAE to eliminate archaic language and referential circularity (e.g. 'amparar' is defined as *favorecer, proteger*, and then 'favorecer' as *ayudar, amparar, socorrer*). In addition, María Moliner added usage examples of her own creation to each entry. However, the fact that the only source of examples in the DUE is the author herself is one of its main drawbacks. Figures 21 and 22 display the entry for 'cama' [bed] in both dictionaries and thus show how collocational information is displayed in Spanish monolingual dictionaries.

As shown in Figures 21 and 22, the phraseological information in these entries is offered as follows:

(i) a definition (e.g. the definition of sense 3 in the DUE states *sitio donde tienen costumbre de* ***acostarse*** *los animales*; and that of sense 1 in the DRAE, *mueble destinado a que las personas se* ***acuesten*** *en él*). Accordingly, it can be inferred that 'acostarse' collocates with 'cama'. In this regard, Penadés (2001: 62) affirms that indeed, in order to know whether a word collocates with another word, it is sufficient to see whether the candidate word is contained in the definition of the other word;

(ii) examples; the DUE provides examples within double quotation marks (e.g. in sense 1, almost all the verbs which collocate with 'cama' are listed: 'acostarse', 'echarse', 'meterse', 'tenderse', 'tirarse', 'tumbarse', 'yacer', 'levantarse', 'saltar', 'tirarse'). In contrast, the DRAE gives them in italics, as shown in sense 5 (*cama de liebres, de conejos, de lobos*). However, it should be highlighted that the DRAE does not normally include usage examples;

(iii) a section reserved for phrases at the end of the entry. Phraseological information appears in bold followed by a definition. First, noun phrases are given, followed by verb ones. Only noun phrases beginning with the headword are included (e.g. 'cama elástica', 'cama de matrimonio'). Combinations containing the headword in a different position can be found at the end of the entry in the form of cross-references (see bullet iv below). As shown in Figure 21, concerning verbal

phrases, the DUE includes ‘meterse en cama’ or ‘saltar de la cama’, information that was already provided in the examples in sense 1;

(iv) cross-references at the end of the entry. They are headed by ‘V’, which stands for *véase* [to see] (e.g. ‘salto de cama’, ‘sofá cama’, in the DUE, and ‘casa de camas’, ‘coche cama’, ‘colgadura de cama’, ‘cosido de la cama’, ‘salto de cama’, ‘sofá cama’, in the DRAE). As such, ‘salto de cama’ appears in the entry for ‘salto’, and ‘sofá cama’ in the entry for ‘sofá’.

cama[1] (del lat. hispánico "cama")

1 ("Acostarse, Echarse, Meterse, Tenderse, Tirarse, Tumbarse, Yacer, Levantarse, Saltar, Tirarse") f. Conjunto formado por una armazón y colchones, almohadas y ropas, donde las personas se acuestan para *dormir o descansar. Se emplea con "la" en vez de "una", aunque se trate de una indeterminada, cuando se menciona como sitio de acostarse o de estar acostado: "tumbarse en la cama, estar en la cama"; se emplea sin artículo en frases que expresan estado de enfermo o se refieren a ese estado: "ha pasado una semana en cama; está para meterse en cama". Armazón de madera o hierro que sostiene las demás cosas de ese conjunto, incluido el somier o sin incluirlo.

2 En un *hospital o sanatorio, plaza para un enfermo, o en un colegio interno, para un alumno.

3 Sitio donde tienen costumbre de acostarse los animales salvajes. *Cueva, *guarida, *madriguera.

4 *Capa de paja, hierba seca, etc., que se pone en los establos para que se acuesten los animales y para hacer estiércol.

5 *Camada (cría).*

6 *Suelo del carro.*

7 *En el melón y otros *frutos, parte que descansa en el suelo cuando están en la mata.*

8 (ant.) *Sepultura.*

9 Mar. *Hoyo que forma en la arena una embarcación varada.*

10 Cada una de las capas en que se dispone una cosa. Camada. *Capa de una vianda que se pone encima de otra para que se comuniquen el calor.*

11 AGráf. *Blandura formada con cartulina o papel con que se recubre el *tímpano de la prensa.*

Cama camera. La de tamaño intermedio entre la de una sola persona y la de matrimonio. *Cama* frailera.

C. elástica. Superficie muy tensa sujeta a un bastidor que permite dar grandes saltos; como la que hay en los parques de atracciones.

C. de matrimonio. La que tiene generalmente el destino que indica su nombre, casi tan ancha como larga.

C. nido. Conjunto de dos camas en un solo mueble, en que una de ellas se guarda debajo de la otra.

C. frailera. *Cama camera.*

C. redonda. 1 Aquella en que duermen varias personas. **2** Relación sexual en la que intervienen más de dos personas.

C. turca. *Diván sin brazos ni respaldo, que puede servir como cama.

Caer en cama. Acostarse por estar enfermo.

V. "*coche* cama".

En cama. Acostado, por estar enfermo. Decumbente.

Estar en cama. Estar en la cama por enfermedad.

Hacer la cama. Levantar las ropas después de haber dormido en ella y volver a colocarlas ordenadamente.

Hacer la cama a alguien. Trabajar en secreto para causarle *daño.

V. "*juego* de cama".

Levantar la cama. Levantar las ropas después de haber dormido en ella, para que se ventilen antes de volver a hacerla.

Meterse en cama. Acostarse.

Saltar de la cama. Levantarse después de haber dormido, para empezar la jornada.

V. "*salto* de cama, *sofá* cama".

Figure 21. Entry for ‘cama’ [bed] in the DUE

cama[1].(Del lat. de San Isidoro *cama*, por *camba*).
1. f. Mueble destinado a que las personas se acuesten en él, compuesto de un armazón, generalmente con patas, sobre el que se colocan un somier o tabla, un colchón, almohada y diversas ropas.
2. f. Dicha armazón sola.
3. f. Plaza para un enfermo en el hospital o sanatorio.
4. f. Plaza para un alumno interno en un colegio.
5. f. Sitio donde se echan los animales para su descanso. *Cama de liebres, de conejos, de lobos.*
6. f. Mullido de paja, helechos u otras plantas que en los establos sirve para que el ganado descanse y para hacer estiércol.
7. f. Tradicionalmente, suelo o plano del carro o carreta.
8. f. En el melón y otros frutos, parte que está pegada contra la tierra mientras están en la mata.
9. f. Capa de vianda que se echa extendida encima o debajo de otra.
10. f. **camada** (‖ conjunto de animales nacidos de un parto).
11. f. *Impr.* Capa de cartón, de papeles o mantilla que se coloca sobre el cilindro impresor para obtener una presión adecuada sobre el molde.
12. f. *Mar.* Hoyo que forma en la arena o en el fango una embarcación varada.
13. f. ant. **sepulcro.**
~ de galgos. 1. f. coloq. **cama** mal acondicionada y revuelta.
~ de matrimonio. 1. f. **cama** que tiene capacidad para dos personas.
~ de podencos. 1. f. coloq. **cama de galgos.**
~ elástica. 1. f. *Dep.* Lona sujeta con muelles a un bastidor sobre la que se hacen ejercicios gimnásticos.
~ mueble. 1. f. **cama** articulada que puede plegarse o recogerse para ahorrar espacio y que toma el aspecto de otro mueble.
~ nido. 1. f. Conjunto de dos **camas** que forman un solo mueble, en el que una se guarda debajo de la otra.
~ redonda. 1. f. **cama** en que duermen varias personas. **2.** f. Práctica de actos sexuales que realizan conjuntamente varias personas en el mismo lugar.
~ turca. 1. f. Especie de sofá ancho, sin respaldo ni brazos, que puede servir para dormir en él.
caer en ~. 1. loc. verb. Caer enfermo.
caer en la ~. 1. loc. verb. Acostarse rendido y con mucho sueño. **2.** loc. verb. desus. **caer en cama.**
caerse alguien **de la ~. 1.** loc. verb. coloq. *Cuba.* Ser demasiado crédulo.
estar en ~, o **guardar ~. 1.** locs. verbs. Permanecer en ella algún tiempo por motivos de salud.
guardar la ~. 1. loc. verb. desus. Permanecer acostado en la **cama.**
hacer ~. 1. loc. verb. desus. **estar en cama.**
hacer la ~. 1. loc. verb. Prepararla para acostarse en ella.
hacerle a alguien **la ~. 1.** loc. verb. Trabajar en secreto para perjudicarlo.
irse a la ~. 1. loc. verb. **acostarse** (‖ para dormir y descansar).
saltar de la ~. 1. loc. verb. coloq. Levantarse de ella con rapidez.
tender la ~.1. loc. verb. coloq. *Arg., Col., C. Rica, Méx., Perú, Ur.* y *Ven.* **hacer la cama.**
□ V.**casa de camas**, **coche cama**, **colgadura de cama**, **cosido de la cama**, **salto de cama**, **sofá cama**

Figure 22. Entry for ‘cama’ [bed] in the DRAE

This comparison shows that the treatment of collocational information in these Spanish dictionaries is far from systematic. There seems to be no special reason for offering a collocation as an example, in the definition, in the section reserved for phrases or in cross-references. Particularly, in the case of the DUE, the presentation of

collocations seems to be based on the arbitrary decision of the author (Corpas 1990: 338).

In contrast, English monolingual dictionaries (e.g. the *Collins Cobuild Dictionary of English*) include more collocational information than their Spanish counterparts. COBUILD stands for *Collins Birmingham University International Language Database* and refers to a project directed by John Sinclair in the 1980s at the University of Birmingham. Its objective was to investigate and describe the English language. The novelty of the COBUILD project was that it compiled the first dictionary to be based on a systematic analysis of a large-scale corpus: *The Collins Cobuild Dictionary of English* was first published in 1987, as an English learners' dictionary. It has gone through many editions (1987, 1995, 2001, 2003, 2006), and the last version can be freely accessed online[26]. To date, the COBUILD group has compiled a large number of dictionaries and various lexicographic resources in various languages[27]. The corpus of the COBUILD dictionaries is called *The Bank of English*. *The Bank of English* is also a result of the COBUILD project, and has more than 650 million words of mainly UK English, but also includes American and Australian variants. It is mainly composed of written texts, although it also has a large repository of oral discourse.

Figure 23 shows the entry for 'bed' in COBUILD. In the same way as its Spanish counterparts, phraseological information is provided in COBUILD as follows:

(i) examples in italics, after the definition (e.g. sense 3, after the definition 'sleep' or 'rest', the construction 'time for bed' is found);

(ii) the definition. For example, in senses 19 and 21, 'bed of roses' and 'bed of nails' are included, followed by their definitions. The same occurs in the case of verbal constructions (e.g. 'go to bed', sense 23; 'put to bed', sense 24; 'take to one's bed', sense 25);

(iii) phrases, such as 'get out of bed on the wrong side' (sense 22).

[26] <http://www.collinsdictionary.com/dictionary> [02/06/11].
[27] For more information concerning Collins products, see:
<http://www.collinslanguage.com/Default.aspx> [02/06/11].

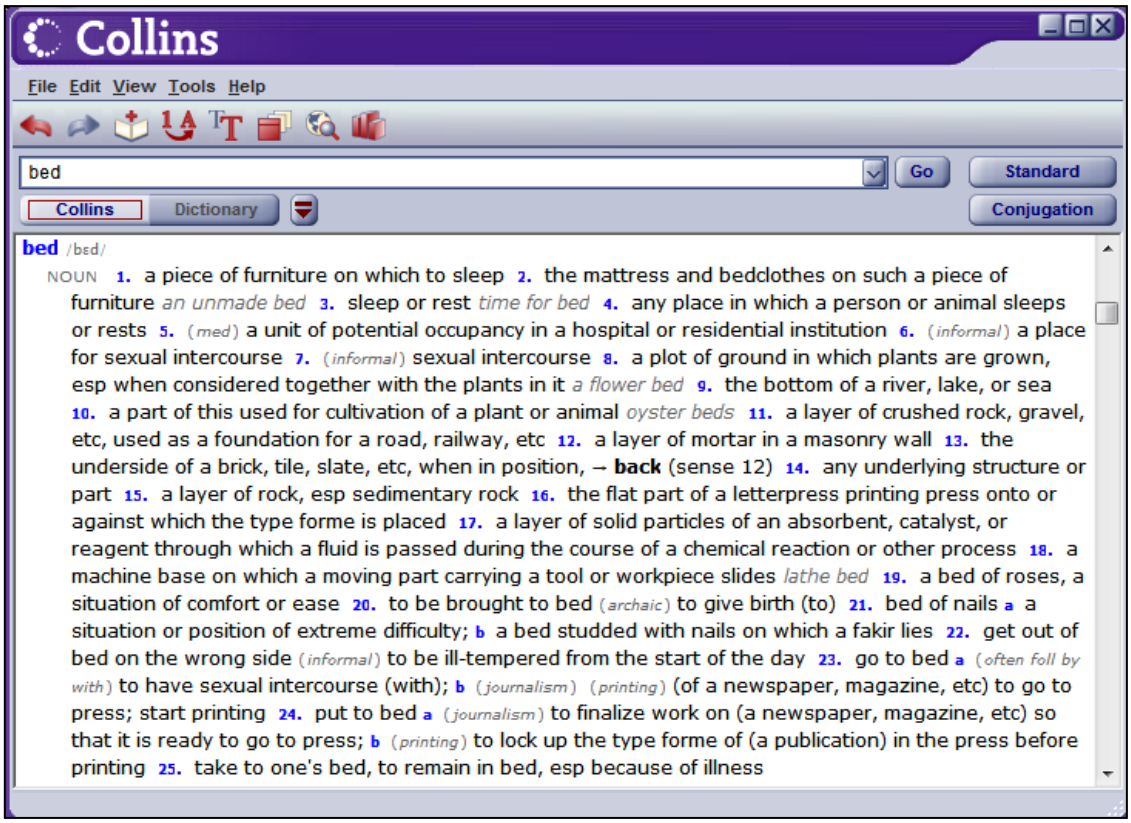


Figure 23. Entry for 'bed' in COBUILD

This description shows that the treatment of phraseological information is also not very systematic in English dictionaries either. In addition, there are also certain inconsistencies in terms of encoding. For instance, the entry for 'bed' lacks basic collocations such as 'single' or 'double'. However, the entry for 'single' includes 'bed' as one of its collocations (i.e. sense 4 of 'single'*: designed for one user* ⇒ 'a single room', 'a single bed'). Surprisingly, 'double bed' is not included in the entry for 'double', which only includes the example of 'a double room'. Curiously enough, this dictionary includes a separate entry for 'twin bed'. This is only one example of the lack of consistency in the treatment of phraseological information in monolingual English resources.

There is no doubt that the COBUILD group revolutionized lexicography in the 1980s, and led to a new generation of corpus-driven dictionaries and various lexicographic reference works for English language learners. Other dictionaries based on corpus are the *Cambridge International Dictionary of English* (CIDE) (subsequently published under the title *Cambridge Advanced Learner's Dictionary*[28]), the *Oxford Advanced Learner's Dictionary of Current English* (OALDCE) (which afterwards

[28]<http://dictionary.cambridge.org/dictionary/learner-english/> [25/03/2012].

became the current *Oxford Advanced Learner's Dictionary*[29]*)*, and the *Longman Dictionary of Contemporary English*[30] (LDOCE). Since all of these learners' dictionaries deal with phraseological information in a very similar way, only the description for COBUILD is provided as an example. For a comparison of the four dictionaries mentioned in this section, see Bogaards (1996). Walker (2009) also compares the COBUILD, OALDCE and LDOCE.

Despite the lack of systematicity in the presentation and description of collocations in dictionaries, the situation has improved over the years, largely thanks to Internet. There are now new ways of collecting and organizing data as well. These processes are enhanced by the use of corpus analysis tools that enable lexicographers to extract more information, However, the systematicity and coherence of any lexicographic or terminographic resource requires the previous selection of a set of criteria regarding the following: (i) the types of collocation to be encoded; (ii) the kind of collocational information to be offered; (iii) the location in the micro or macrostructure of the dictionary where collocations should be included.

2.2.2 Phraseology in bilingual general dictionaries

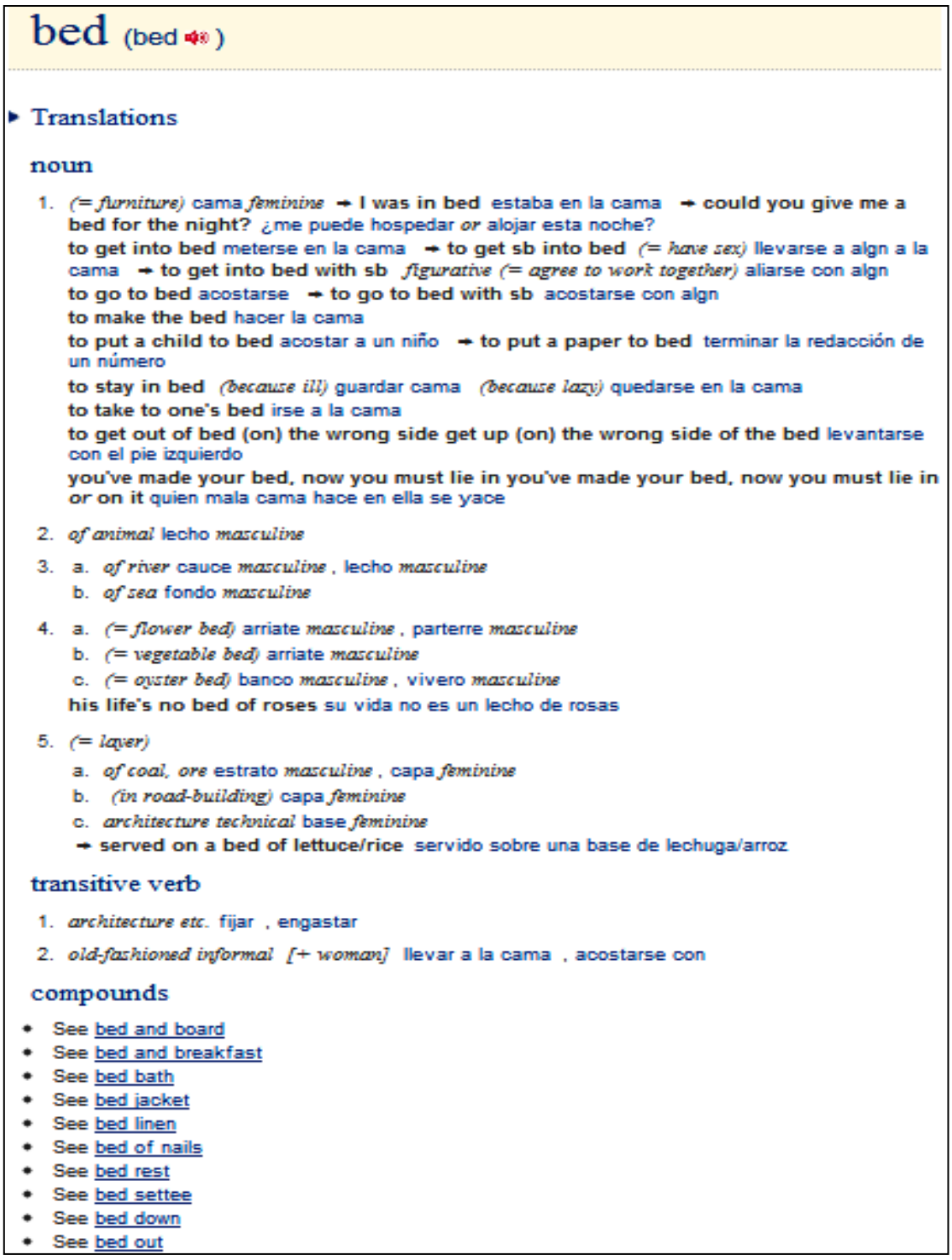
The aim of bilingual general language dictionaries is obviously to help users retrieve correspondences of a word or phraseme in another language. This makes them useful for translators, especially in the production of the target language text. Since 'literal' correspondences are often not the best choice, such dictionaries generally include a significant number of collocations in order to show the differences between the two languages and to help users communicate with native-like fluency.

The entries in these dictionaries frequently do not include definitions since it is assumed that users already know the meaning of the word or have previously looked it up in a monolingual dictionary. Generally speaking, these dictionaries are bidirectional, i.e. they allow searches from the source language to the target language (e.g. English-Spanish) and from the target language to the source language (e.g. Spanish-English).

The bilingual dictionaries published by Collins COBUILD[31] and Oxford[32] are two of the most representative English-Spanish bilingual dictionaries. Even though

[29] <http://oald8.oxfordlearnersdictionaries.com/> [25/03/2012].
[30] <http://www.ldoceonline.com/> [25/03/2012].
[31] <http://www.collinsdictionary.com/> [06/04/2011].
[32] < http://www.diccionarioinglesespanoloxford.es/bed/en> [06/04/2011].

Larousse[33] and Cambridge[34] also have published bilingual dictionaries, the information in them is not as complete. Therefore, the examples in Figures 24 and 25 are taken from the COBUILD and Oxford dictionaries.



bed (bed)

▸ **Translations**

noun

1. *(= furniture)* cama *feminine* ➞ **I was in bed** estaba en la cama ➞ **could you give me a bed for the night?** ¿me puede hospedar *or* alojar esta noche?
   to get into bed meterse en la cama ➞ **to get sb into bed** *(= have sex)* llevarse a algn a la cama ➞ **to get into bed with sb** *figurative (= agree to work together)* aliarse con algn
   to go to bed acostarse ➞ **to go to bed with sb** acostarse con algn
   to make the bed hacer la cama
   to put a child to bed acostar a un niño ➞ **to put a paper to bed** terminar la redacción de un número
   to stay in bed *(because ill)* guardar cama *(because lazy)* quedarse en la cama
   to take to one's bed irse a la cama
   to get out of bed (on) the wrong side get up (on) the wrong side of the bed levantarse con el pie izquierdo
   you've made your bed, now you must lie in you've made your bed, now you must lie in or on it quien mala cama hace en ella se yace
2. *of animal* lecho *masculine*
3. a. *of river* cauce *masculine* , lecho *masculine*
   b. *of sea* fondo *masculine*
4. a. *(= flower bed)* arriate *masculine* , parterre *masculine*
   b. *(= vegetable bed)* arriate *masculine*
   c. *(= oyster bed)* banco *masculine* , vivero *masculine*
   his life's no bed of roses su vida no es un lecho de rosas
5. *(= layer)*
   a. *of coal, ore* estrato *masculine* , capa *feminine*
   b. *(in road-building)* capa *feminine*
   c. *architecture technical* base *feminine*
   ➞ **served on a bed of lettuce/rice** servido sobre una base de lechuga/arroz

transitive verb

1. *architecture etc.* fijar , engastar
2. *old-fashioned informal [+ woman]* llevar a la cama , acostarse con

compounds

- See bed and board
- See bed and breakfast
- See bed bath
- See bed jacket
- See bed linen
- See bed of nails
- See bed rest
- See bed settee
- See bed down
- See bed out

Figure 24. Entry for 'bed' in the Collins Cobuild English-Spanish Dictionary

[33] <http://www.larousse.com/en/dictionaries/english-spanish> [06/04/2011].
[34] <http://dictionary.cambridge.org/dictionary/english-spanish/> [06/04/2011].

bed /*sustantivo*,*verbo*
sustantivo

1 cama

to go to bed irse a la cama

to be in bed estar en la cama

a single/double bed una cama individual/de matrimonio
to make the bed hacer la cama
I've put the children to bed. He acostado a los niños.
It's time for bed. Es hora de irse a la cama.

NOTA Nótese que sólo cuando **bed** se refiere al mueble se usa el artículo. Compárense los siguientes ejemplos
I came to pay for the bed. He venido a pagar la cama.
Did I get you out of bed? ¿Te he sacado de la cama?
Ver más vocabulario relacionado con la palabra **bed**

2 colchón
3 (*flores*) macizo
4 lecho (*marino o de un río*)
5 base
grilled chicken, served on a bed of rice pollo a la parilla, servido sobre una capa de arroz
6 (*Geología*) estrato
LOC
bed and board pensión completa
get out of bed on the wrong side ; (USA)
get up on the wrong side of the bed levantarse con el pie izquierdo
go to bed with sb (coloquial) irse a la cama con alguien
(not) be a bed of roses (no) ser un lecho de rosas
take to your bed meterse en la cama enfermo
you've made your bed, so you must lie on it (refrán) tú te lo has buscado*Ver tb* **wet**

verbo transitivo (**-dd-**)
1 fijar
The bullet bedded itself in the wall. La bala se incrustó en la pared.
2 (anticuado) acostarse con
PHR V
bed down pasar la noche , dormir , acostarse (*en un lugar improvisado*)
air bed
bed and breakfast
bed-wetting
camp bed
double bed
flower bed
river bed

Figure 25. Entry for 'bed' in the Oxford bilingual English Spanish dictionary

In both dictionaries, the entry can be for a single word (e.g. 'bed') or a compound (e.g. 'bed and breakfast'). The printed version of the dictionaries includes the phonetic description of the lemma, whereas the online version allows users to listen to the pronunciation of the word. As shown, the grammatical category of the headword is specified as well as its different senses. For each sense, the translation is provided, followed by the various collocations in which the headword participates.

The source-language collocations in the COBUILD dictionary are in boldface type and the target-language correspondences are in dark blue. In contrast, the source-language correspondences in the Oxford dictionary appear in red italics and boldface type, and the translations in normal font.

In both dictionaries, collocations are either offered explicitly (e.g. in sense 1 in both dictionaries, 'to make the bed' – 'hacer la cama'*)* or inserted in examples (e.g. sense 1 in Oxford, 'I've put the children to bed'), as a totally arbitrary decision. For the sake of coherence with the other examples in sense 1 (i.e. 'to go to bed', 'to be in bed', 'to make the bed'), it would have been preferable to explicitly include the collocation (i.e. 'to put sb to bed') instead of giving it in an example sentence. This entry could lead users to believe that 'to put children to bed' is a fixed expression. This is evidently not the case since it is possible to say 'I put my grandmother/dolls/ brother to bed'.

The same inconsistency in the treatment of collocations also occurs in COBUILD. In the first sense of 'bed' (as a noun), the collocation is sometimes given (e.g. 'to get sb into bed'), whereas at other times, it has to be deduced from the examples (e.g. 'I was in bed'). Needless to say, it would have been preferable to include 'to be in bed' as an explicit collocation and then give additional examples if needed.

Another example can be found in 'bed' in its sense of *layer*, the fifth sense in both dictionaries: COBUILD illustrates the collocation with the example 'served on a bed of lettuce/rice', and Oxford with 'grilled chicken served on a bed of rice'. Once again, the user has no means of knowing whether 'bed' in this sense can only be used with 'rice' and 'lettuce' or on the contrary, can be followed by other nouns such as 'bed of sand'. There is thus no way of differentiating a fixed collocation from a trivial example.

The same inconsistency is also found in the treatment of compounds. For example, Oxford includes 'double bed' both as a collocation (see sense 1) and as a compound with its own lemma. In contrast, 'single bed' only appears as a collocation and not as a compound. Surprisingly, COBUILD does not include either 'double bed' or 'single bed' in the entry for 'bed'. However, it does include 'double bed' as a compound

though it does not provide a direct link to this compound from the entry for 'bed'. As shown in Figure 24, COBUILD only offers cross references to compounds which start with 'bed', without mentioning other ones (e.g. 'river bed' is not included).

The problem lies in the fact that collocations in bilingual dictionaries are not well differentiated from typical examples provided for a particular sense. Users must thus decide whether the examples are representative or if they merely illustrate how the word is used in context. This decision is somewhat difficult for non-native speakers. The presentation and configuration of the information within an entry is essential to provide quick retrieval of data as well the effective acquisition of knowledge. In this regard, the layout of the Oxford dictionary is more user-friendly than that of COBUILD. Accordingly, in the proposal for encoding and describing verb phrases in this thesis, attention is paid not only to content, but also to the way that the content is displayed.

2.2.3 Phraseology in collocations or combinatorial dictionaries

The growing importance of phraseological information can also be seen in the recent publication of combinatorial or collocational dictionaries. In this section, the most representative English and Spanish collocation dictionaries for general language are described[35].

The Explanatory Combinatorial Dictionary (ECD) (Mel'čuk, Clas, and Polguère 1995; Mel'čuk 1984-1999) (§2.1.3.1) deserves special mention in that it has been and still is the most influential combinatorial dictionary for any language. In fact, various dictionaries described in this section are based on the theoretical and methodological premises of the ECD. A full description of the ECD along with its theoretical background, the Meaning Text Theory is given in 2.1.3.1 since it has served as a model for many lexicographical and terminographical resources.

To date, there are three general collocational dictionaries on the market for English[36], namely, the *BBI Combinatory Dictionary of English* (BBI) (1986, 1997, 2009), *Oxford Collocations Dictionary for Students of English* (OCD) (2002, 2009), and *Macmillan Collocations Dictionary* (MCD) (2010). Spanish also has three important

[35] Terminographic resources containing phraseological information are displayed in 2.3.

[36] Two other English collocation dictionaries, although not being described in this section, deserve to be mentioned: (i) *A Dictionary of English Collocations* (Kjellmer 1994); (ii) *LTP Dictionary of Selected Collocations* (Hill and Lewis 1997). In the same line as McGee (2012: 327), the reason of not including Kjellmer's is that it was not written for learners of English and it focuses on adjacent collocations. LTP is not included either since it provides less coverage than the other three dictionaries described in this section for English. In fact, the headword 'bed' is not included in LTP.

collocations dictionaries: *Redes: Diccionario combinatorio del español contemporáneo* (REDES) (Bosque 2004), *Diccionario combinatorio práctico del español contemporáneo* (PRÁCTICO) (Bosque 2006), and the *Diccionario de colocaciones del Español* (DICE) (Vincze, Mosqueira, and Alonso 2011). In this section, an overview of how collocational information is treated in these collocations dictionaries based on the headwords 'bed' and 'cama', for English and Spanish resources, respectively. Since the DICE is restricted to the domain of emotion, the entry for 'indignación' [indignation] is chosen. This section analyzes their principal advantages and limitations of the design of our entry template for verb phrase collocations. Other analyses of collocational dictionaries can be found in Cowie (1998a), Appleby (2000), Klotz (2003), Walker (2009), Coffey (2011), McGee (2012), Ferrando (2012), *inter alia*.

However, the way that each resource encodes, classifies, and displays collocations varies substantially. In the words of Nuccorini (2003: 367):

> […] the delimitation and description of contents, the theoretical principles adopted for the inclusion, selection, classification and presentation of headwords, the sources and the layout of phraseological dictionaries vary considerably both linguistically and lexicographically.

There is a great range in the number of headwords in the three English dictionaries (McGee 2012: 335). The BBI includes approximately double the number of entries as the OCD, and the OCD contains twice as many entries as the MCD. Therefore, collocations for a less common word are more likely to be found in the BBI. One thing in which collocational dictionaries coincide is that they are meant for production (Nuccorini 2003: 367):

> English monolingual collocational dictionaries agree on one point: they are meant for encoding purposes and are consistently addressed to advanced learners and translators.

We agree with Nuccorini in that collocational dictionaries are especially used for productive purposes, but as shall be seen, some of the most recent dictionaries on the market can also be useful for decoding purposes (e.g. Bosque 2004). In our opinion, any collocational resource should be helpful both for encoding and decoding. As such, the information in the verb templates in this thesis helps translators to better understand the source text and to produce a suitable target text.

The BBI Dictionary of English Word Combinations

Of the six collocational dictionaries in this section, the *BBI Dictionary of English Word Combinations* (Benson, Benson, and Ilson 1986, 1997, 2009) was the first to be compiled. Its third edition is evidence of its success and great acceptance. The 1997 edition included 18,000 entries and roughly 90,000 collocations. The most recent 2009 edition is an extensive revision with new collocations in the field of computing and Internet (although the dictionary does not specify the exact number of new combinations). It also includes new example sentences and more detailed descriptive usage notes (Benson, Benson, and Ilson 2009: xi).

Although it was conceived in ignorance of the works of Hausmann (Benson 1989), both the content and form of the BBI reflects the principles proposed by Hausmann (1989). The first edition of the BBI is described as a learner's dictionary of English to help users to express themselves in English as naturally as possible. In this regard, it follows the long tradition of learner's dictionaries of English which began with the OALDCE. Benson, Benson and Ilson, authors of the BBI, mention the principal drawbacks of learner's dictionaries of that time. Such limitations include their treatment of grammatical information; the presentation of collocations within examples and definitions; and the large number of free combinations, which are perfectly predictable from the meaning of their components (§2.1.4.1). Fortunately, the situation of learner's dictionaries has dramatically changed for the better over the last twenty years.

Unlike the OCD and MCD, the BBI is the only dictionary that is not corpus-based, but rather 'corpus-refined' (McGee 2012: 330), which means that it is based on the authors' intuition. In the words of their Benson, Benson, and Ilson (2009: viii):

> Nowadays, our task is eased not only by the availability of corpuses of contemporary English (such as the British National Corpus) but also by the amazing resource of the Internet itself, which enables us to search in it for a word and find superb examples of that word in context. Nor should it be forgotten that an important source of new information in BBI 3 is, paradoxically, BBI 2, now that the computer allows material from an entry in BBI 2 to be added to other entries in BBI 3 when such material is appropriate.

The BBI is the only dictionary of the three English dictionaries that makes use of the new approach to corpus referred to as *Web as Corpus* (WaC). Consequently, it uses the web directly as a corpus to extract information (see §4.2.2.1 and §4.2.2.2 for a

detailed description of the *Web as Corpus* and the *Web for Corpus* approaches, respectively).

Another striking difference of the BBI in comparison to the OCD and MCD is that apart from including lexical collocations, the BBI is also a kind of syntactic dictionary that supplies extensive information about the complementation patterns of verbs, nouns, and adjectives. The dictionary is thus referred to as a *combinatory dictionary,* rather than a *collocational dictionary* since it includes not only phraseology, but also valency (Benson, Benson, and Ilson 2009: i):

> Traditionally, the combination of words into grammatical patterns has been called colligation or complementation or construction (though in BBI it is called collocation, too) and its result has been called valency. A dictionary that provides both phraseology and valency is a dictionary of word combinations; or, in the terminology of Igor Mel'čuk, whose work has inspired us, a combinatory or combinatorial dictionary. BBI is a combinatory dictionary.

In this sense, the BBI includes two types of collocations as explained in Benson (§2.1.3.1): (i) grammatical collocations; and (ii) lexical collocations. Grammatical collocations, also referred to as *colligation*[37] by other authors such as Hoey (2005), are always listed under the dominant word. Eight major types of grammatical collocations are designated: (G1) noun + preposition; (G2) noun + *to* + infinitive; (G3) noun + *that*-clause; (G4) preposition + noun; (G5) adjective + preposition; (G6) predicative adjectives + *to* + infinitive; (G7) adjectives + *that*-clause; (G8) nineteen English verb patterns (designated by the capital letters A to S). An extract of the verb lemma 'begin' is offered as an example (Table 10):

begin
1. (D; intr.; tr.) to ~ as (to ~ a new career as a teacher)
5. (E) she began to work
6. (G) she began working

Table 10. Extract of the lemma 'begin' in the BBI

[37] See Stubbs (2002) for a complete description regarding the differences between collocation and colligation.

As shown, some of the letters displayed for 'begin' are D, G and E. As described in the introduction to the dictionary, letter D means that the verb forms a collocation with a certain preposition; E is reserved for verbs that are followed by *to* + infinitive; and G is used for verbs ending in *-ing*. Grammatical collocations are always listed under the dominant word. As shall be seen, in the description of our verbs, the constructions in which verbs participate are also taken into account. They can be implicitly deduced from the usage examples displayed for each verb.

Lexical collocations are divided into seven types, labelled L1, L2, L3, and so on. L1 collocations are mainly transitive verbs denoting Creation or Activation + a noun/pronoun (e.g. 'come to an agreement'). These combinations are referred to as CA collocations (Benson, Benson, and Ilson 2009: xxxi). L2 collocations include verbs meaning Eradication and or Nullification (e.g. 'reject an appeal'), which are referred to as EN collocations (Benson, Benson, and Ilson 2009: xxxii). However, the BBI insists on the arbitrary character of EN and CA combinations, making it difficult for foreigners to produce them spontaneously. L3 collocations have the pattern of adjective + noun (e.g. 'strong tea'). L4 collocations are noun + verb combinations (e.g. 'alarms go off'). L5 collocations are noun + of + noun combinations (e.g. 'a bouquet of flowers'). L6 collocations are adverb + adjective combinations (e.g. 'deeply absorbed'); and finally, L7 collocations are verb + adverb combinations (e.g. 'affect deeply'). In this thesis, only verb collocations are described.

The listing of lexical collocations in the BBI also follows the principles of Hausmann for the configuration of collocations in learner's dictionaries. Thus, when there is a noun in the collocation, the collocation is placed under the noun. If there are two nouns, it appears under the second noun; if there is no noun, then it appears under the adjective; when there is no noun or adjective, it is placed under the verb.

Generally speaking, the BBI does not generally include free combinations, which are defined as "elements that are joined in accordance with the general rules of English syntax and freely allow substitution" (Benson, Benson, and Ilson 2009: xix). Table 11 shows an example of an entry of the BBI corresponding to 'bed'. Compounds are also listed as headwords in the dictionary.

Bed *n.* [“article of furniture for sleeping”] 1. to make, make up a ~ (I’ll make you up a ~ = I’ll make a ~ up for you) USAGE NOTE: *to make up a bed* is to put bedclothes on it. *to make a bed* is to re-arrange the bedclothes on it neatly after someone has slept in it. 2. to strip; unmake a ~ 3. to go to ~ with (she went to ~ with a heating pad)
USAGE NOTE: *to go to bed with smb.* means “to have sexual intercourse with smb.” 4. to be, lie, stay in ~; to lie; sit on a ~; to wet the ~ (“to urinate in it accidentally”) 5. to get out of ~ 6. to put smb. to ~ (to put the children to ~) 7. to take to one’s ~ (“to remain in bed because of illness”) 8. a double; king-size; queen-size; single; twin ~ 9. a bunk; camp (BE); folding; four-poster; hospital; rollaway; sofa; truckle (BE), trundle (AE) ~ 10. a water ~ 11.an unmade ~ 12. a feather ~ (“a feather mattress”) (see also *bed and breakfast* at **breakfast**) [“ground at the bottom of a body of water”] 13. a river ~ [“plot of ground”] 14. a flower ~

Table 11. Entry for ‘bed’ in the BBI

As shown in Table 11, after the headword and its grammatical category, a definition is provided in square brackets and double quotation marks. After the definition, lexical collocations are listed. In the BBI, lexical collocations always precede grammatical collocations. Since the example does not contain grammatical collocations, all the combinations displayed correspond to lexical ones. For certain collocations, an additional explanation is given in double quotation marks and sample squares, e.g. 7 to take to one’s ~ (“to remain in bed because of illness”). Other collocation senses also offer a usage example in squares, e.g. 1. (I’ll make you up a ~ = I’ll make a ~ up for you). For some articles, usage notes are included (see sense 1), and these provide additional information concerning the headword. As shown in Table 12, varieties of English are marked as AE, for American English (e.g. ‘trundle bed’), and BE (for British English) (e.g. ‘camp bed’).

The order of presentation of lexical collocations inside entries are verb + noun (CA collocations) (e.g. ‘make up a bed’), verb + noun (EN collocations) (e.g. ‘unmake a bed’), adjective + noun (e.g. ‘a double bed’), noun + verb (none in this entry), noun + noun (e.g. ‘a flower bed’). The BBI thus has a more systematic approach since all collocations are assigned a labelled pattern (L1, L2, L3, G1, etc.), and then patterns are consistently arranged in each dictionary entry in the order established in the introduction to the dictionary. For example, the order for nouns is L1, L2, L3, L4, L5, G1, G2, G3, G4. Despite this systematization, the BBI includes many fixed phrases that do not fit into any of the types of grammatical and lexical collocations described above

(Benson, Benson, and Ilson 2009: xxxv). Such phrases are normally listed under the label pattern *misc* (miscellaneous).

One of the positive aspects of the dictionary is that it does not exclusively rely on lexical collocations. As previously mentioned, it is also a syntactic dictionary which supplies information about the complementation patterns of verbs, nouns and adjectives. Although Cowie (1998c: 225) argues that complementation of verbs and nouns should be treated in a valency dictionary and not in a collocational one, in our opinion, constructional information for verbs should be included in a dictionary of collocations. In the words of Siepmann (2005: 416):

> [...] collocation and verb complementation are intimately related, since many noun-verb collocations require a specific distribution of semantic roles.

The main drawback of the BBI is that it does not provide semantic characterisation of collocations. McGee (2012), however, states that there is a kind of semantic classification since semantic classes of collocates are grouped together, though not explicitly labeled with a subheading. Nevertheless, this classification only describes collocations that follow the pattern of verb + noun in CA (creation/activation) and EN (eradication/nullification). Other semantic areas are not explored.

Moreover, despite the fact the BBI focuses on restricted collocations, a number of free combinations are also encountered. In his analysis of the first edition, Cowie (1998c: 226–227) highlights that in the lemma for ‘book’, ‘write’ is found, which has a literal sense and can occur in many contexts. In contrast, in the entry for ‘letter’, the free combination ‘receive a letter’ is displayed. Collocations such as ‘write a book’ or ‘write a letter’ are also included in the 2009 edition (McGee 2012: 332).

Despite these minor criticisms, the BBI is an enormous contribution to the study of phraseology. It includes a great deal of collocational information and tries to formalize these phenomena. In addition, its definition of collocations makes a distinction between collocation and free lexical combinations.

Oxford Collocations Dictionary

The Oxford Collocations Dictionary was first published in 2002, and a second edition was published in 2009 (McIntosh, Francis, and Poole 2009). Like the MCD, it is an English dictionary of collocations especially aimed at upper intermediate to advanced

students of English for text production purposes. In this regard, it can be used for the preparation of official exams, such as the Cambridge First Certificate in English (FCE), Advanced Certificate in English (CAE), Proficiency Certificate in English (PCE) and the International English Language Testing System (IELTS).

Since the term *collocation* is not used in the same way by all linguists (§2.1.4), it becomes necessary for any dictionary of collocations to first define it since the information in each entry depends on this. According to Klotz (2003: 58), the OCD's notion of collocation is more inclusive than Hausmann's. For example, collocations such as 'public announcement', 'honest answer' or 'feel anxiety' are included although they are semantically quite predictable. However, at the same time, the OCD notion of collocation is more restrictive than Sinclair's since 'yesterday's announcement', 'only answer', and 'some anxiety' are not present.

The introduction to the last edition (McIntosh, Francis, and Poole 2009: v) highlights that the OCD does not follow any particular theory, but rather is more oriented to learner's needs. Collocations are thus selected on the basis of the answers to the following questions (McIntosh, Francis, and Poole 2009: v): (i) Is it a typical use of language? (ii) Might students of English want to express this idea? (iii) Would users look up this entry to find out how to do this? The result is a dictionary whose most recent edition contains 250,000 word combinations and about 9,000 headwords for nouns, verbs, and adjectives, along with 75,000 examples showing how collocations are used.

Like the BBI, the OCD does not provide noun collocates for verb and adjective entries. For example for the verb 'study', the OCD provides adverb collocates (e.g. 'carefully', 'closely', 'in depth'); phrases ('be easily'/ 'well studied'); and prepositions (e.g. 'for'), but no information is given regarding noun collocates.

The corpus used to extract the most salient combinations was the *British National Corpus*, in the first edition, and the *Oxford English Corpus* of 2 billion words in the most recent edition. When necessary, this edition also provides variants for British and American English, and includes a CD containing the dictionary plus exercises. There are also 25 usage notes throughout the dictionary on a variety of subjects, (e.g. religions) along with ten special thematic pages (e.g. sports). Table 12 shows the entry in the OCD for 'bed'.

bed *noun*
1 piece of furniture for sleeping on

- ADJ. **double, king-size, single, twin | bunk, camp, feather, folding, four-poster, hospital, sofa | warm | unmade** *a messy room, with an unmade bed and clothes on the floor* | **marriage**
- VERB + BED **do** (*informal*), **make, make up | strip** *Please strip the beds and put the sheets in the washing machine.* | **climb into, crawl into, get into, go to, tumble into** *She crawled into bed exhausted.* | **climb out of, get out of, leap out of | lie (down) on, lie in, sit on** *He lay in bed, reading his book. ◊ Elisabeth was sitting on her bed writing a letter.* | **put sb to, tuck sb up in** *It's your turn to put the children to bed.* | **wet** *Don't punish a child who wets the bed.*
- BED + NOUN **clothes, linen**
- PREP. **in ~** *I like to be in bed before 11 o'clock.* **out of ~** *Are you out of bed yet?*
- PHRASES **bed and breakfast, the edge/side of the bed, the foot/head of the bead, get sb into bed** (=have sex with sb), **go to bed with sb** (=have sex with sb), **take to your bed** (=go to bed because you are ill), **time for bed** *Come on, children, it's time for bed.*

2. piece of ground for growing flowers, vegetables, etc.

- ADJ. **flower, rose, strawberry | ornamental | raised**
- PREP. ~**of** *ornamental beds of roses*

Table 12. Entry for 'bed' in OCD

In the OCD, when the headword is polysemous, as is the case for 'bed', a definition of each sense is provided. However, when the headword only has one sense, no definition is given. As shown in Table 12, immediately after each definition, the associated collocations are displayed in bold in order to be easily identifiable from the rest of the text. Collocations are first organized in regards to their grammatical structure. (e.g. Adj; Verb + *bed*; *bed* + Verb; Prep, Phrases).

The OCD is the only one of three dictionaries described in this section that devotes a special section for phrases. Authors of the MCD justify not including phrases by saying that the "OCD has a 'phrases' category where items are often included on grounds of non-transparency of meaning" (Kilgarriff et al. 2010: 373).

It is true that the OCD does not clearly define the concept of *phrase*. For example, many phrases that were regarded as relevant in monolingual and bilingual dictionaries

(e.g. 'get out of bed on the wrong side', 'bed of rice/vegetables') (§2.2.1 and §2.2.2) are not included in the phrase section of the OCD. In addition, for some strange reason, it includes the combination 'bed and breakfast', but not 'bed and board'.

Nevertheless, despite these minor inconsistencies, the OCD approach to phrase description is very useful. Phrases are frequent in general language and are difficult for non-native speakers to master. The idea of putting phrases in a separate section in a dictionary makes them easier to understand for the user. Phrases are beyond the scope of our research, but will be included in subsequent studies, based on the methodology in this thesis.

Differences in meaning within the same grammatical relation are expressed in OCD by means of a vertical dash "|". As previously mentioned, McGee (2012:333) states that collocations within OCD entries are "semantically organized, but this organization is not actually labeled". However, in our opinion, the semantic organization in this dictionary is rather tenuous since the semantic relationship between the headword and the collocations is not stated and thus must be deduced by users. For instance, regarding the collocates for the first sense in Table 12, 'double', 'king-size', 'single', and 'twin' are separated from 'bunk', 'camp', 'feather', 'folding', 'four-poster', 'hospital' and 'sofa'. However, it is difficult for users to discern the relationship between 'feather bed' and 'sofa bed'.

In addition, the OCD provides explanations in brackets when needed, both regarding meaning and usage. Finally, it offers usage examples in italics to illustrate how each collocation should be used in context.

MacMillan Collocations Dictionary

The Macmillan Collocations Wordlist (Rundell 2010) is the most recent dictionary of collocations in English though its compilation process started in the 1990s. Like the OCD, its objective is to help upper intermediate to advanced English students improve their writing skills in order to pass the International English Language Testing System (IELTS). In the same way as the OCD and in contrast to the BBI, it is corpus-based. The corpus used is a two-billion word corpus, known as the *World English Corpus*[38]. Its novelty lies in its methodology since it is the first dictionary to extract collocations automatically in the form of *word sketches* thanks to a new software tool, incorporated

[38]For a complete description of the corpus, see <http://www.macmillandictionary.com/corpus.html> [12/04/2012].for a complete description of the corpus.

into the Sketch Engine corpus query system[39] (Kilgarriff et al. 2004). As shall be seen in 4.2.2.2, a *word sketch* is an automatic corpus-derived summary of a word's grammatical and collocational behavior. In fact, the OCD is a model for what was intended to be automatically produced by word sketches (Kilgarriff et al. 2010: 373):

> Our goal for what word sketches aim to do is provide a grammatically-organized list of collocates which would form a suitable entry for a collocations dictionary such as OCD.

As evidenced by Kilgarriff et al. (2010), word sketches offer reliable and valuable information These authors formally evaluate word sketches for four languages, namely, Dutch, English, Japanese, and Slovene. To that end, forty-two headwords were selected and twenty collocates for each headword were analyzed in the four languages by asking users whether the collocation is suitable for inclusion in a published collocation dictionary. The evaluation showed that more than two thirds of the collocations were of publishable quality.

In line with the OCD, the MCD also has a very broad classification of collocation for MCD, as can be seen in the introduction of the dictionary[40]:

> Collocations are 'semi-preconstructed phrases' which allow language user to express their ideas with maximum clarity and economy. Not only that, there is a strong correlation between frequency in a corpus and typicality, which means that the use of common collocations contributes to the naturalness of a text.

The MCD contains about 4,500 entries. Its editors justify this relatively small number compared to other collocational resources such as the OCD by saying that they "prefer to give full coverage to a smaller number of words"[41]. Even though the MCD contains fewer entries than the OCD, both dictionaries are similar in length (McGee 2012: 333):

> Although [the MCD] contains around half as many entries as the OCD – with around 4,500 keyword entries – the two dictionaries are of comparable length (MCD 911 pages, versus OCD 963 pages).

[39] < http://www.sketchengine.co.uk/> [12/04/2012].
[40]<http://www.macmillandictionaries.com/features/how-dictionaries-are-written/macmillan-collocations-dictionary/#1> [12/04/2012].
[41]<http://www.macmillandictionaries.com/features/how-dictionaries-are-written/macmillan-collocations-dictionary/#1> [12/04/2012].

Table 13 shows the entry for 'bed' as described in the MCD. For each headword, first the definition(s) is provided. In contrast to OCD, in the MCD, the definition of each headword is always given, even when the headword only has one meaning. This means that the MCD can be used both for encoding as well as decoding. This is extremely useful, and thus, in our verb template, definitions of each verb frame are always provided. The idea is to help the translator find correspondences in the production phase as well as in the early reception stages.

bed N
a piece of furniture for sleeping on

adj +**N** types of bed **bunk, divan, double, folding/foldaway/fold-up/four-poster/4-poster, king-size/king-sized, single, sofa, twin** *If you sleep in a double bed up against a wall, ensure you sleep by the wall side.*
► ***describing*** *a bed* ***comfortable*****, comfy** informal, **unmade** *The beds were comfortable and the kitchen had everything we could ever need.*

v+into+**N climb, collapse, crawl, fall, get, roll, sink, tumble** *We dumped our bags and just crawled into bed.*

v+*to*+*N come, crawl, get, go, put sb, retire, send* sb *He was glad to get to bed. Send your child to bed at a regular and reasonable time.*

Table 13. Entry for 'bed' in MCD

Collocations are subsequently provided after the definition of the headword, which is first categorized in terms of grammatical category. The MCD gives the grammatical relation between the headword and collocate by means of part of speech patterns (e.g. adj + N; v + *into* + N; v + *to* +N), but not explicitly like the OCD does (i.e. adj+ *bed*). According to Fuertes-Olivera (2011: 59), it would have been better to have explicitly specified the grammatical labels for the sake of language students:

> My main contention is that the structural labels, i.e. the grammatical codes, are not explained, which hinders its usability in some teaching/learning situations, e.g. Spanish universities, in which students are unfamiliar with grammar information.

In our opinion, it is improbable that the absence of grammatical labels is a problem for advanced students of English. However, their explicit presentation makes lexicographic entries more transparent. As previously mentioned, unlike the OCD, the

MCD does not put prepositions in a different section, but rather includes them in the constructions and examples.

The MCD expresses different ideas within the same grammatical category through semantic groupings with a definition headed by the symbol ►. As can be seen for the grammatical category adj + n, two semantic groupings are provided: (1) *types of bed*, which include collocations such as 'bunk', 'divan', 'double', 'folding/foldaway/fold-up/four-poster/4-poster', 'king-size/king-sized', 'single', 'sofa', 'twin'; and (2) *describing a bed*, with collocates such as 'comfortable', 'comfy' or 'unmade'. Its authors seem to be very proud of the fact that they classify collocations according to meaning: "it is the only fully corpus-based collocations dictionary which incorporates semantic groupings"[42].

The underlying idea is very good though, as envisaged in the dictionary, it seems to be more viable from a grammatical or syntactic point of view rather than from a semantic perspective. This is evidenced, for instance, by the use of constructions described in terms of *v + into* or *v + to*, and the exclusion of basic collocations such as 'make' which could pose problems for students of English who must decide between 'do the bed*' or 'make the bed'. Regarding the second meaning (describing a bed), only the collocations 'comfortable', 'comfy' and 'unmade' are given, even though a bed can also be 'hard' or 'uncomfortable'.

The importance of classifying collocations according to meaning is of paramount importance. This is the reason why in this thesis, collocations are specified in lexical domains and semantic frames in order to better group together lexical units sharing the same properties. This classification, along with the rest of parameters proposed is meant to help predict verb meaning, and thus can be applied to translation (Sánchez and Buendía 2012: 558).

The MCD also includes a section devoted to usage notes that provides the following information: (i) colligation, i.e. the tendency of a word to appear in a particular form (Hoey 2005) (e.g. a verb that is mainly used in its passive form; or a noun used primarily in its plural form); (ii) alternatives to collocations, which are common phrases that can be used instead of the collocation. Colligation information is displayed against a pink background, and possible alternatives to collocations against a grey one. The systematicity achieved by clearly stating what is included in each type of

[42] <http://www.macmillanenglish.com/courses/macmillan-collocations-dictionary/> [21/01/2012].

usage note and the color distinction gives users easy access to this information. Finally, like the OCD and the BBI, the MCD includes usage examples in italics to illustrate how each collocation should be used in context.

However, the most striking difference of the MCD is the type of collocational information provided. In other words, it is the first dictionary to include noun collocates for adjectives and verb entries. Neither the BBI nor the OCD provides nouns for adjectives and verbs (McGee 2012: 334):

> The standard practice in collocation dictionary production until the publication of the MCD was to place the 'independent' or 'autonomous' base of the collocation (usually the noun) as an entry word, where one can find its 'dependent' collocates (verbs or adjectives).

This is indeed one of the most valuable and distinctive aspects of the MCD compared to the other English collocational dictionaries. This is something that we have also done in our research as shall be seen in chapter 5.

Redes. Diccionario combinatorio del español contemporáneo

Redes. Diccionario combinatorio del español contemporáneo, henceforth REDES (Bosque 2004), is the first combinatorial dictionary published for the Spanish language. It was developed by Ignacio Bosque and his team at the Complutense University of Madrid. The dictionary was elaborated from a corpus of 250 million words composed of texts from 68 Spanish and Latin American newspapers from 1993 to 2003. It is for linguists studying lexical restrictions in Spanish and, in general, for anyone interested in the use of the Spanish language. In 2006, another more practice-oriented dictionary appeared, which is derived from the data contained in REDES: the *Diccionario combinatorio práctico del español contemporáneo*, henceforth PRÁCTICO (Bosque 2006).

Apart from being the first work to deal with word combinations for Spanish, the novelty of REDES lies in the fact that predicates are the nucleus of the dictionary. This is in vivid contrast to the general tendency of traditional collocational dictionaries in other languages, where the noun (which normally corresponds to the base) is the primary focus of attention (Barrios 2007: 1):

> […] *Redes*, el primer diccionario combinatorio del español, obra innovadora en su enfoque pues en ella los predicados se convertían en núcleo y principal objeto de estudio.

Bosque prefers to refer to his dictionaries as *diccionarios combinatorios* [combinatorial dictionaries] in the same way as the BBI or the ECD. The reasons are the following (Barrios 2007: 2): (i) Bosque's dictionaries include something more than collocations, i.e. "fenómenos de combinatoria en el sentido amplio" [combinatory phenomena in the broadest sense]; (ii) the notion of dictionary of collocations lends itself easily to multiple interpretations (e.g. the entry for 'book' in the OCD includes 'good' and 'great', whereas in PRÁCTICO, neither 'buen libro' [good book] nor 'libro genial' [great book] is included). This is evidence that certain resources, commonly referred to as *dictionaries of collocations*, also include frequent combinations with the subsequent risk of draining the notion of *collocation* of its substance.

REDES contains 7,115 lemmas composed of nouns, adjectives, adverbs, verbs as well as noun, adjective, adverbial, and verbal phrases. It has two types of entry: (i) *entradas analíticas* or *entradas largas* [analytical or long entries]; (ii) *entradas abreviadas* or *entradas cortas* [abbreviated or short entries]. Generally speaking, long entries are for *selecting* base words, whereas short entries are for *selected* words or *collocates*, as stated by Bosque (2004: xxxviii):

> De manera muy simplificada, puede decirse que las palabras que aparecen en las entradas analíticas son, en la mayor parte de los casos, PALABRAS SELECCIONADORAS, mientras que las voces que aparecen las referencias cruzadas son PALABRAS SELECCIONADAS.

Long entries can be various pages. Table 14 shows an extract from a long entry in REDES for the phrase 'a bombo y platillo' which, in reference to a piece of news or an event, means to spread far and wide[43]. As shown in Table 14, after specifying the grammatical category (adverbial phrase) [loc.adv.], the *entradilla* is offered with an explanation of the combinatorial potential of the lemma.

[43] 'Cama' [bed] is not displayed for this example, since 'cama' is not a long entry.

a bombo y platillo *loc.adv.* ▉ Admite la variante, menos usada, *a bombo y platillos*. Se combina con…

A VERBOS QUE DENOTAN DIFUSIÓN O TRANSMISIÓN DE INFORMACIÓN: **1 anunciar++: …**un congreso en el que se anunciaría a *bombo y platillo* que la familia nacio-nalista se iba a vivir a una misma casa. CAN080101 **2 proclamar ++:** …algo que la Comisión proclamó *a bombo y platillo* hace diez días pero que ahora no resulta tan claro. EME190895. **3 difundir ++**: Esta noticia fue difundida *a bombo y platillo* en una rueda de prensa…EME 200694 **4 propagar +:** …la prensa falangista lo habría propagado *a bombo y platillo*. LVE190395. **5 pregonar +:** …ha sido pregonado *a bombo y platillo* su rotundo éxito…LVE161095. **6 presentar:** …fue presentado *a bombo y platillo* a los medios de comunicación… ABC190595 **7 transmitir:** …produce rubor el recuerdo de la infame entrevista (…) transmitida *a bombo y platillo* por los telediarios…EPE131099 **8 publicar:** … para que fuesen publicadas *a bombo y platillo*. LVE020395 **9 airear:** …aireadas *a bombo y platillo* por asociaciones estadounidenses…EME240796 **10 promulgar:** …unos carteles enormes promulgan *a bombo y platillo* el «gran» esfuerzo del Ayuntamiento…EPE261201

B VERBOS QUE DESIGNAN OTRAS ACCIONES VERBALES, MÁS FRECUENTEMENTE SI EXPRESAN LA DE PONER ALGO DE MANIFIESTO: 11 afirmar +: Se afirma *«a bombo y platillo»* que el precio del dinero está en los niveles más bajos de su historia…EME120194

Table 14. Extract from a long entry in REDES: 'a bombo y platillo'

The most important aspect of long entries is that they are semantically divided into lexical classes, identified with capital letters (A, B, C, D, etc.). The descriptor of the lexical class is the text that defines or characterizes it. This description is in capital letters so that users can easily identify the sense that they are looking for. In this regard, in Table 14 two lexical classes are displayed: (i) A. *VERBOS QUE DENOTAN DIFUSIÓN O TRANSMISIÓN DE INFORMACIÓN* [verbs that denote diffusion or transmission of information]; (ii) B. *VERBOS QUE DESIGNAN OTRAS ACCIONES VERBALES*, MÁS FRECUENTEMENTE SI EXPRESAN LA DE PONER ALGO DE MANIFIESTO [verbs that designate other verbal actions, more specifically if they express the idea of making something evident]. The last lexical class descriptor is normally referred to as *POSIBLES USOS ESTILÍSTICOS* and includes infrequent uses.

Some lemmas also include a lexical class in the end referred to as *POSIBLES USOS CRUZADOS*, which provides possible combinations that are options to the one given.

A meaning-based classification and description of combinations is one of the most positive features of REDES. A semantic classification for collocations is necessary since as stated by Siepmann (2005: 424): "[…] dependencies exist not merely between lexical units, but also between semantic features". In our research, verbs are first classified in lexical domains according to the LGM (§3.1.2.1) on the premise that verbs belonging to the same lexical domain or subdomain have similar paradigmatic and syntagmatic information. More specifically, verbs are further classified into semantic frames and the arguments activated by each predicate are grouped in semantic categories (§5.4.2).

Subsequently, the combinations activated within each lexical class are given in bold typeface headed by a number. However, the numbering does not reinitiate each time a new lexical class is encountered since these numbers are subsequently used to automatically create short entries. Then, the frequency of each combination is represented by means of the following symbols: "++" is a very frequent combination; "+" signifies a fairly frequent one; no symbol means that it is an acceptable combination for native speakers; "-" means that the combination is not very frequent though possible.

Documented examples showing the combination in context are provided, along with the reference that identifies the source date. An explanation of the labels used for documenting the examples is given in the introduction to the dictionary. For example, as shown in Table 14, the example for the verb 'presentar', is *…fue presentado a bombo y platillo a los medios de comunicación…*, which is identified as ABC190595. This means that the example comes from the Spanish newspaper *ABC*, dated 19 May 1995. However, as stated by Bosque (2004: xlix), sometimes it is very difficult to retrieve a certain combination in a referenced source because it is not very frequent, even though it sounds natural to a native speaker. Should this happen, the authors create the examples themselves, labeling them as *undocumented* (INDOC) (Bosque 2004: xlix):

> Por muy amplio que sea el corpus con el que se trabaje, no es posible encontrar en él todas las combinaciones que correspondan a una clase léxica determinada.

Therefore, REDES includes both natural and frequent combinations in a corpus. Generally speaking, frequency in a corpus coincides with naturalness (acceptable for a native speaker). However, there are times when infrequent expressions seem to be very natural. In a similar way, the examples of the verb templates outlined in this thesis are retrieved from the corpus compiled for the research (i.e. following a conventional *web for corpus approach*, see §4.2.2.1). However, when certain word combinations are not encountered, they are directly searched for in the web, following a *web as corpus approach*, i.e. considering the web as a corpus (§4.2.2.2).

At the end of long entries, two more specifications are provided: (i) a grey square with the phrase *se combina con* [combines with], which offers word combinations that do not fit in the other lexical classes; (ii) cross-references to other entries by means of *véase también* [see also].

In short entries, the lexical classes of long entries disappear since these short entries were not written one by one, but rather automatically obtained from long entries. In other words, the computer extracted nouns, phrases, verbs, adjectives, and adverbs from the body of analytical entries, and converted them into lemmas. As shown in Tables 15, 16, 17, 18, 19, neither grammatical information nor definitions (even when the lemma has various senses) are provided for short entries. The various senses within a lemma are distinguished by means of this symbol: ◆. In REDES, five types of short entries can be distinguished:

1. *Referencias cruzadas a las voces* [cross-references to terms] (Table 15), which are lemmas for selected words, i.e. collocates or words that can combine with many other words. As shown in the extract in Table 15, a large number of the combining words included in this type of lemmas contain a superscript which comes from the number of the corresponding analytical entry. This is the reason why numbers in analytical entries do not start from zero each time a new lexical class is described. When the combining words do not contain a number, this means that they were extracted from the *entradilla* or that they have been explicitly added by the author to provide the user with more combinatorial information. In the words of Bosque (2004: liii), these entries are a kind of extended index resulting from the rearrangement of the information contained in analytical entries with some supplementary information added:

En general, las referencias cruzadas a las voces están concebidas en Redes como índices ampliados. Constituyen por tanto el resultado de reordenar la información contenida en las entradas analíticas y añadir algunas informaciones complementarias (Bosque 2004: liii).

problema ◆ abrumador[45],abstruso[1], acuciante[1], a cuestas[3], alambicado, álgido[14], apreciable[20], apremiante[7], arduo[39], banal, candente[7], capital, clásico, complejo, congénito15, controvertido[34], coyunturarl[1], crucial[58], decisivo[82], de consideración, delicado, descomunal, desencadenante, difícil, endémico, endemoniado, endiablado[15], enmarañado, enrevesado[12], espinoso, eterno, fácil, galopante[22], gordo, grave, grueso[5], hondo[37], imprevisible[51], imprevisto, inextricable[8], ingente[70], insalvable[13], insignificante, insoluble[1], insoslayable[25], integral[60], intrincado[9], irresoluble[1], irreversible[12], latente, leve, ligero, mamayúsculo[24], nimio[17], pasajero[24], peliagudo, perentorio[52], profundo[138], sencillo, serio[1], severo[70], simple, soterrado[34], tangencial[39], trivial, vasto[18], vigente[34] ◆ a la medida (de)[16], a la vista (de)[33]

Table 15. Extract of the short entry for 'problema' in REDES

2. *Referencias cruzadas a los conceptos* [cross-references to concepts], which do not refer to concepts rather than words. The lemma comes in small caps so that it can be easily distinguished from cross-references to terms, which are displayed in bold. The combining words included in these short entries contain letters as superscripts instead of numbers. The letters come from the lexical class of the correspondent analytical entry.

PERCEPCIÓN
◆ (SUSTANTIVOS) Véase: abigarrado[B], abismal[B], acusado[I], adulterar[E], afilado[C], afinar[C], a flor de piel[B], agridulce[A], aguzar[A], amortiguar[C], analíti-, co[O], arduo[I], avieso[A], beatífico[B], beligerante[C], borroso[B], bosquejar[O], centrípeto[C], cerrar los ojos (ante)[A], cobrar fuerza[B], corroborar[C], cristalino[A].

Table 16. Extract from the short entry for 'percepción' in REDES

3. *Entradas del índice conceptual* [entries of the conceptual index] that also designate concepts. In contrast to cross-references to concepts, which are indexes of the lexical classes of the corresponding analytical entries, these entries are

indexes of the lemmas or words that have an entry in REDES. They are identifiable since they are displayed in small caps and italics (see Table 17).

INFLUENCIA, EFECTO Y CONSECUENCIA Véase: ◆impactante, mortífero, viciado ◆en (mil) pedazos, en consecuencia, mortalmente, sin efecto ◆a resultas (de) ◆atracción, atractivo, cicatriz, consecuencia, dependencia, eco, efecto, encanto, escombro, estigma, huella, impacto, impronta, influencia, influjo, magisterio, magnetismo, rastro, repercusión, rescoldo, resonancia, resto, restos, revés, revuelo, roce, ruina, secuela (de), víctima ◆afectar, incidir, influir, repercutir, resonar □Véase también: *ATINGENCIA Y CORRESPONDENCIA*; *DAÑO Y PERJUICIO*.

Table 17. Extract of the short entry for 'influencia', 'efecto' *y* 'consecuencia' in REDES

4. *Series abreviadas* [abbreviated series], to which 'cama' belongs. These entries do not contain any superscript. An analytical entry is not proposed for these entries since, in general, some of the most frequent combinations for these entries do not fit into the analytical entry. As shown in the example for the lemma 'cama' (Table 18), only verbal collocations are listed. This entry does not include any kind of semantic information, examples or more nouns or adjective phrases collocating with the headword. The collocational information provided for lemmas, however, is greatly enhanced in the dictionary PRÁCTICO.

cama ◆ acostar(se) (en), deshacer, guardar, hacer, levantar(se) (de), meter(se) (en), postrar(se) (en), quedar(se) (en), recostar(se) (en), sacar (de), salir (de), tender(se) (en)

Table 18. Entry for 'cama' in REDES

5. *Remisiones* [cross-references], which refer to both suggestions (*véase también*) [see also] or cross-references (*véase*) [see] (Table 19):

a grito limpio Véase: **a gritos** □ Véase también: **grito**
en vivo Véase: **en directo**
EFECTO Véase*: INFLUENCIA, EFECTO Y CONSECUENCIA*
ABANDONO Véase*:* DEJACIÓN; SALIDA

Table 19. Short entry in REDES for 'a grito limpio' and 'en vivo'

Diccionario combinatorio práctico del español contemporáneo

The *Diccionario combinatorio práctico del español contemporáneo* (Bosque 2006), henceforth PRÁCTICO, is a combinatory dictionary for Spanish derived from REDES. It is the practice-oriented version of REDES, conceived for text production (Barrios 2007: 1–2):

> Si *Redes* era un intento de reflexión acerca de la restricción léxica, *Práctico* pretende ayudar a encontrar la palabra adecuada, sumándose a la lista de trabajos onomasiológicos orientados a la producción y no a la comprensión […].

It thus targets teachers and students of Spanish as a foreign language, translators, and generally anyone that aspires to speak Spanish with native fluency. Although it is shorter than REDES (1,305 pages in contrast to 1,839), it contains almost twice the number of entries (14,000 in contrast to the nearly 8,000 of REDES), and about twice the number of word combinations (400,000 in contrast to 200,000). Conceptual entries are not included in PRÁCTICO because it is a less conceptual and more practical dictionary. This is indeed the reason why it has more word combinations and less descriptive information. PRÁCTICO has three types of entry:

(i) *simple entries*, which merge analytical and short entries in REDES (see Table 20). The lemmas are displayed in bold roman characters;

(ii) *generic entries*, which are not present in REDES. They group the words with the same combinatorial potential in the same semantic field in one entry so as to avoid repetition. For instance, since all the months of the year combine with a similar set of words, there is a generic entry *'MES'* [month] that includes all of the months of the year. The lemmas of generic entries come in capital letters.

(iii) *cross-references entries*, which are enclosed in square brackets and are supplementary aids designed to facilitate information retrieval. For example, the

lemma *'julio'* [July] corresponds to a cross reference entry that refers the user to the entry for *'MES'* [month], as previously above.

The content of the simple entries in PRÁCTICO is very different from that in the brief entries of REDES. This is the case of 'cama' [bed] as shown in Table 20, which corresponds to the entry for 'cama' [bed] in PRÁCTICO. When compared to the same lemma in REDES, most of the information in PRÁCTICO is new. The microstructure of an entry in PRÁCTICO is similar to that of the OCD (see Table 20).

cama s.f.

- CON ADJS. **mullida · acogedora · confortable** *Leía tumbado sobre la confortable cama del hotel* **· placentera · cómoda · blanda · dura · incómoda ‖ destartalada · desvencijada ‖ funcional · utilitaria ‖ de matrimonio · individual · separada** *Están casados, pero duermen en camas separadas* **‖ plegable · adicional · supletoria · abatible ‖ de agua · turca · elástica**
- CON SUSTS. **cabecero (de) · pie (de)** *Te he dejado los zapatos a los pies de la cama* **‖ ropa (de) ‖ escena (de) ‖ compañero,ra (de) ‖ nido** *Las dos hermanas duermen en una cama nido*
- CON VBOS. **hacer** *Mientras yo hago las camas, tú prepara el desayuno* **· deshacer ‖ compartir ‖ meter(se) (en) · ir(se) (a)** *Buenas noches, me voy a la cama* **· quedar(se) (en) · acostar(se) (en) · tender(se) (en) · recostar(se) (en) · postrar(se) (en) · permanecer (en) ‖ levantar(se) (de) · salir (de) · sacar (de) · incorporar(se) (de) ‖ dar vueltas (en) ‖ servir (de) ‖ llevar (a)**
- CON PREPS. **debajo (de) · bajo · sobre · en**

□EXPRESIONES **caer en cama** [ponerse enfermo] **‖ cama redonda** [acción de mantener relaciones sexuales más de dos personas a la vez] **‖ {estar en/guardar} cama** [estar en la cama por enfermedad] *Su médico le ha aconsejado guardar cama* **‖ hacer la cama** (a alguien) [actuar secretamente para perjudicarle] *col.*

Table 20. Entry for 'cama' in PRÁCTICO

In contrast to the analytical entries in REDES, entries in PRÁCTICO do not contain an *entradilla*. Like the OCD, in PRÁCTICO, when the headword has more than one sense, a brief definition of each sense is first provided in square brackets headed by

a black square "▌". For 'cama' (Table 20), no description is provided since it only has one meaning.

As can be seen in Table 20, collocations in PRÁCTICO are first organized according to grammatical category. In this regard, PRÁCTICO distinguishes between combinations *CON ADJS.* [with adjectives], *CON SUSTS.* [with nouns], *CON VBOS.* [with verbs], *CON PREPS.* [with prepositions], and *EXPRESIONES* [phrases]. Like the OCD, it also devotes a special section to phrases, which is a completely new feature, especially considering that REDES does not provide this kind of information. However, this dictionary does not include a section for non-verb combinations. Users must thus deduce this type of information, which might be a problem for non-native speakers of Spanish (Model 2008: 196):

> […] el diccionario carece de indicaciones explícitas, por lo cual el usuario tiene que detectar la función de un sustantivo a partir de otras indicaciones menos obvias. Pese a que se ha elaborado un sistema para revelar si un sustantivo es sujeto u objeto del verbo indicado, este sistema resulta algo enrevesado y no se aplica en todos los casos.

Like the OCD, collocations are grouped in terms of semantic proximity. The resulting groups are separated by "‖". As shown in Table 20, 'meter(se) (en)', 'ir(se) (a)', 'quedar(se) (en)', 'acostar(se) (en)', 'tender(se) (en)', 'recostar(se) (en)', 'postrar(se) (en)', and 'permanecer (en)' are similar in meaning. However, only a native speaker of Spanish would realize that 'acostarse' is somewhat different from 'quedarse en'.

In our opinion, again, this type of classification is often confusing since the semantic relation between members is not explicitly labeled and is often difficult to infer. This is indeed one of its main drawbacks when compared to REDES. In PRÁCTICO, the semantic classes of REDES are not included. There is only an attempt to implicitly group collocates by semantic proximity as stressed by Barrios (2007: 7):

> Si el lector de Práctico es un lingüista y busca paradigmas de clases léxicas, no encontrará un análisis de los rasgos que permiten definir los subgrupos pero aunque los rasgos semánticos no se hagan explícitos permanecen implícitos en la forma de agrupar.

In addition, PRÁCTICO provides explanations in square brackets, when needed, concerning meaning and usage. Finally, it provides usage examples in italics to illustrate how the collocations should be used in context. It is important to emphasize here that the documented examples in REDES have been replaced by examples created by the authors themselves in order to help students of Spanish and also not to create an excessively large dictionary.

Diccionario de Colocaciones del Español

The *Diccionario de Colocaciones del Español*[44], henceforth DICE (Vincze, Mosqueira, and Alonso 2011; Alonso, Nishikawa, and Vincze 2010) is an online collocations dictionary of Spanish, currently being developed at the University of A Coruña (Spain). To date its contents are restricted to the domain of 'feeling' and it only contains 211 lemmas, which are each associated with one or various lexical units (LUs). Therefore, the number of LUs described in the dictionary is larger. The website provides two access points: (1) to the dictionary itself; (2) to an advanced search option consultation. An additional didactic model is currently being built.

As in most of the collocations dictionaries in this section, the DICE focuses on encoding and is especially conceived for learners of Spanish (Alonso 2010). As shall be seen, like the BBI, REDES and PRÁCTICO, it is based on the premises of Explanatory Combinatorial Lexicology (§2.1.3.1). However, it differs from them in that it is the only one that describes the semantic content of collocations in terms of lexical functions.

Accordingly, the DICE's notion of *collocation* is in line with Mel'čuk's, which in turn is in consonance with that of Hausmann (§2.1.3.1), and differs from the frequency-oriented approach of British contextualism (§2.1.3.2). In other words, the DICE assumes a particular meaning relationship between the constituents of a collocation. As such, it has a broad conception of collocations, and therefore, includes phrases as transparent as 'muy cansado' [very tired] or more opaque ones, such as 'armado hasta los dientes' [fully armed]. The feature shared by both phrases is their selection restriction. As underlined by Alonso (2010: 57), the adverb 'muy' has a wider combinatorial potential than 'hasta los dientes', but it is not a free combination since when 'cansado' is combined with 'muy' the resulting phrase has the meaning of *being very tired*:

[44] <http://www.dicesp.com> [16/05/2012].

> Desde nuestra perspectiva, no hay problema en incluir bajo el mismo término de *colocación* sintagmas completamente transparentes como *muy cansado* hasta sintagmas completamente opacos como *armado hasta los dientes*. En ambos sintagmas un elemento está seleccionado restringidamente. Es cierto que el adverbio *muy* tiene una combinatoria amplísima, que la combinación es transparente. Sin embargo, la combinación *muy cansado* no es libre en el sentido de que *cansado* se combina con *muy*, no tanto porque *muy* significa 'muy', sino porque es uno de las maneras de expresar ese sentido, predicado de *cansado*.

As previously mentioned, the DICE only focuses nouns of feeling, and thus does not include 'cama', the word used as a comparison the other resources. Therefore, the headword 'indignación' [indignation] is given as an example to describe the microstructure of an entry in the DICE (see Figure 26).

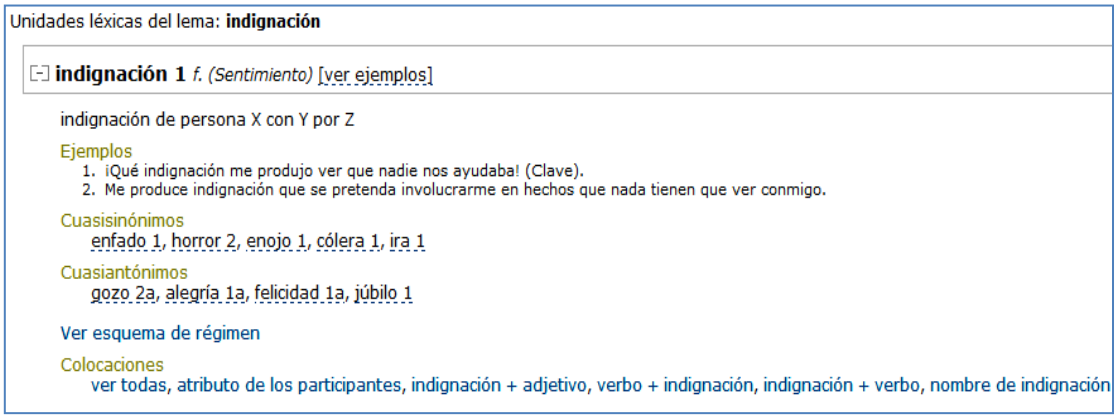


Figure 26. Entry for the LU 'indignación1' in the DICE

As shown in Figure 26, the information provided for each LU in the DICE is the following: (1) a semantic tag, which is the general meaning of the LU (in this case, the semantic tag is *sentimiento* [feeling]); (2) its actantial structure, which lists the participants in the situation designated by the LU (i.e. *indignación de persona X con Y por Z* [indignation of person X with Y because of Z])[45]; (3) usage examples, mainly derived from the *Corpus de Referencia del Español Actual* (CREA)[46] as well as from the corpus *LexEsp* (Sebastián et al. 2000), the *Corpus del Español*[47], directly from the web, and from other lexicographical works; (4) quasi-synonyms and quasi-antonyms of the LU; (5) the syntactic schema (*esquema de régimen*), which is displayed in another window (see Figure 27) with syntactic information regarding the linguistic realizations

[45] As shall be seen in 2.3, specialized language dictionaries, such as the DAFA, the DiCoInfo and the DiCoEnviro also provide the actantial structure of participants in the same way as the DICE.
[46] <http://corpus.rae.es/creanet.html> [16/05/2012].
[47] <http://www.corpusdelespanol.org> [16/05/2012].

of actants (e.g. the participant Y with 'indignación' can be linguistically headed by the prepositions 'con', 'contra', 'ante' and 'hacia' (e.g. 'La indignación con/contra/ante/hacia el gobierno'); (6) collocations.

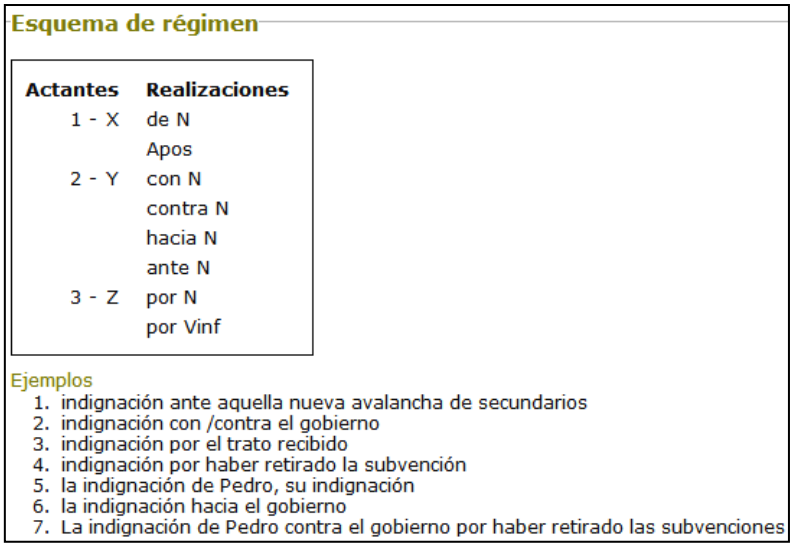


Figure 27. Syntactic schema of the LU 'indignación1' in the DICE

As shown in Figure 27, the collocations for each LU are classified into five groups: (i) *atributo de los participantes* [attributes of the participants], (e.g. 'en un momento de indignación'); (ii) *indignación* + *adjetivo* [LU + adjective] (e.g. 'indignación popular'); (iii) *verbo* + *indignación* [verb + LU] (e.g. 'aumentar', 'acrecentar', 'alimentar'); (iv) *indignación* + *verbo* [LU + verb] (e.g. 'aplacarse'); (v) *nombre de indignación* [noun *de* + LU], which lists all collocations headed by the combination of a noun + *de* (e.g. 'sentimiento de indignación'). By clicking on each collocation group, the system displays a list of glosses or brief linguistic descriptions that are characteristic of a specific meaning (Alonso, Nishikawa, and Vincze 2010: 370-371) (see Table 21). In other words, they are linguistic adaptations of LFs. Table 21 displays the glosses for the construction *verb* + *indignación*.

As can be seen, the LF may also appear. In Table 21, LFs are included in the screenshot, but the system permits them to optionally be displayed with the glosses. In addition, the number of collocates that are part of the meaning of each gloss is specified. For example, as shown in Table 21, for the gloss 'causar que la ~ sea mayor' [to cause the indignation to become greater], the associated LF is *Caus Pred Plus*. Also given are three collocates in consonance with this semantic description.

sentir ~ Oper1 (3 valores en total)
sentir una ~ intensa Magn + Oper1 (3 valores en total)
continuar sintiendo ~ Cont Oper1 (1 valor en total)
hacer a algo objeto de ~ Labor12 (1 valor en total)
causar que la ~ sea intensa Caus Pred Magn (1 valor en total)
causar que la ~ sea mayor Caus Pred Plus (3 valores en total)
causar que la ~ sea menor Caus Pred Minus (5 valores en total)
causar ~ en alguien Caus Func1 (6 valores en total)
causar que alguien sienta ~ Caus Oper1 (2 valores en total)
causar que alguien sienta un ~ intenso Magn + Caus Oper1 (1 valor en total)
hacer que la ~ no se note non Perm1 Fact (4 valores en total)
no permitir que la ~ se manifieste non Perm1 Manif (2 valores en total)
la ~ se manifiesta en algo Conv21 Manif (1 valor en total)
manifestar corporalmente la ~ V Sympt (3 valores en total)
verbo cuasisinónimo V (1 valor en total)
sentir ~ por simpatía con otro que lo siente V no estándar (1 valor en total)
manifestar ~ Caus1 Manif (4 valores en total)
causar que el ~ desaparezca Liqu1 Func (1 valor en total)

Table 21. Glosses for the group of collocations *verb + indignación* in the DICE

By clicking on the plus symbol at the left of each gloss, the system displays the collocates that transmit the meaning of the gloss. Accordingly, as shown in Figure 28, for 'causar que la ~ sea mayor', the verbs 'aumentar', 'acrecentar' and 'alimentar' are encountered. This does not mean, however, that the verbs are synonyms, but rather that they can express a similar meaning.

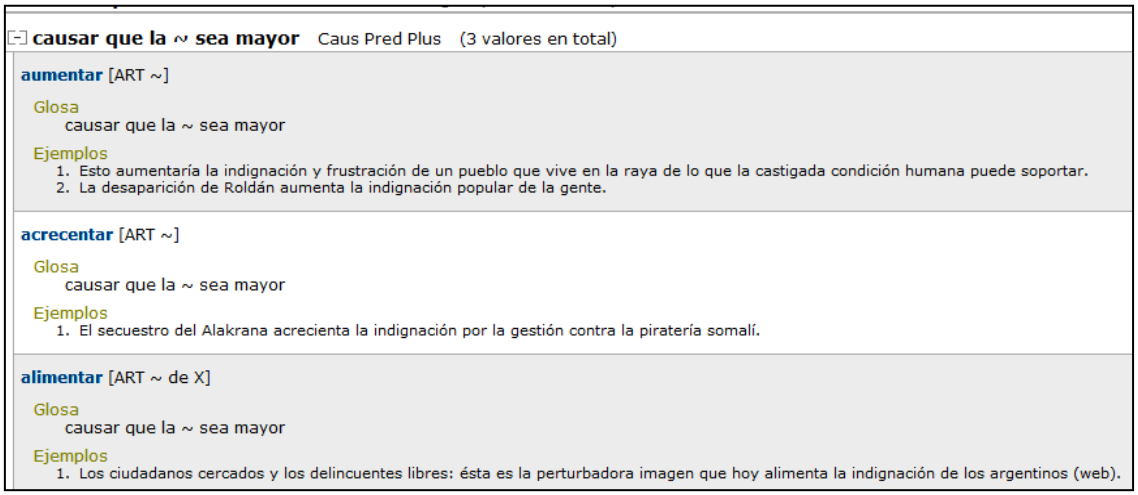


Figure 28. Collocations of the gloss 'causar que la indignación sea mayor' for the LU 'indignación' in the DICE

As can be observed, the DICE includes the following information for each collocate: (i) the syntactic schema of the collocation in square brackets (e.g. for the verb 'alimentar' the schema is [ART ~ *de X*], which means that 'alimentar' comes in a construction headed by an article and followed by the preposition 'de' and an actant X); (ii) the gloss; (iii) usage examples extracted from the corpus (e.g. 'Los ciudadanos cercados y los delincuentes libres: ésta es la perturbadora imagen que hoy alimenta la indignación de los argentinos'). At times, certain entries also include expressions that despite not being collocations, express the same meaning as the collocation.

As previously mentioned, the DICE also has an advanced search component that permits users to find answers to specific questions. The system allows three types of searches: (1) *directas* [direct searches]; (2) *inversas* [indirect searches]; (3) *ayuda a la redacción* [help in writing]. Direct searches permit users to find the collocates of a base by means of an LF. Nevertheless, the system only allows users to search for one LF at a time. Additionally, they can specify the lemma and the LU.

For example, as shown in Figure 29, a direct search could provide an answer to the question of which verb can combine with 'indignación' to express the meaning of the lexical function *Caus.* For this query, the system displays 30 collocations organized in terms of the LFs underlying their meaning. 'Indignación' is found within the LF *Caus Pred Plus.*

In our opinion, the idea of providing a query system is very helpful for final users. This is the reason why, along with the verb entry template designed in this thesis, a query system to extract information from these templates is also proposed. Nonetheless, the problem of direct searches in the DICE lies in the fact that meaning can only be expressed in terms of LFs. This rather difficult metalanguage makes it unsuitable for the majority of users, except for linguists. Paradoxically, however, the intended users of the dictionary are learners of Spanish: "tomamos como punto de referencia al usuario de diccionario, especialmente al aprendiz de español" (Alonso 2010: 55).

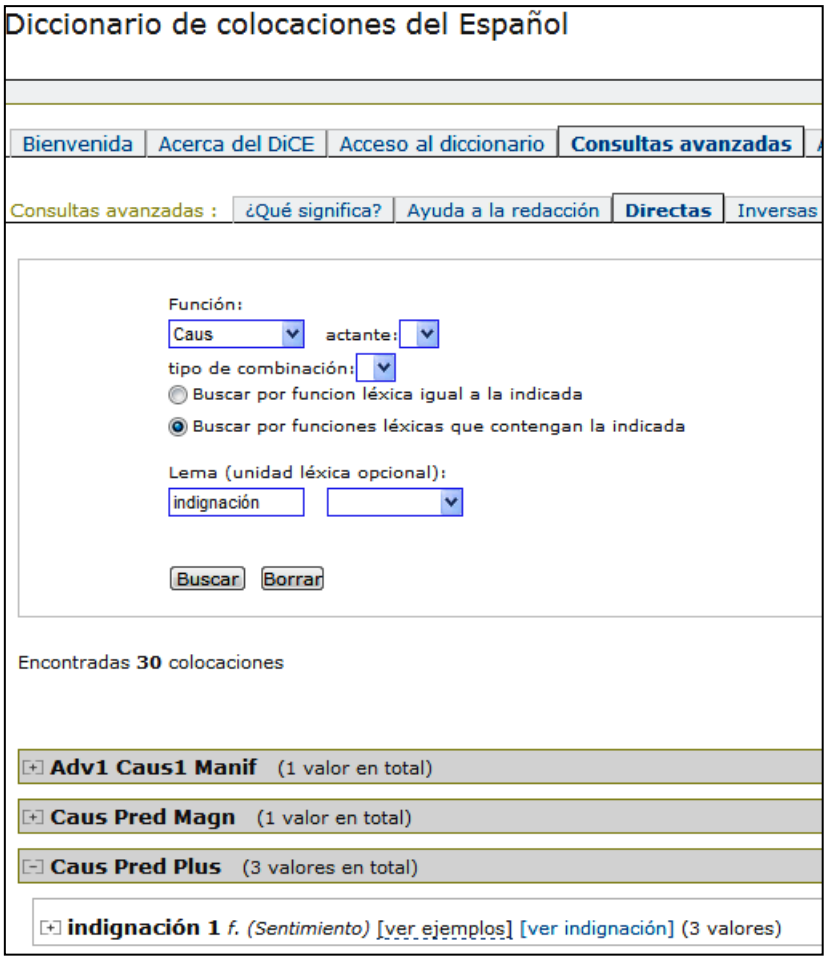


Figure 29. Direct search option in the DICE

The indirect search option (Figure 30), allows for two types of queries. The first one permits users to find the base of a collocation from the collocate (referred to as *valor* [value] in the DICE). Additionally, the LF can also be specified. For example, as shown in Figure 30, it can be used to find out which lemma is selected by the collocate 'alimentar' to express the meaning of the LF *Caus.* Once again, the problem is the metalanguage of the LFs. It is true that users can launch a query without having to specify the LF, but in that case, an excessively large number of records are displayed.

Figure 30. Indirect search option in the DICE

The second query that can be launched with the indirect search option permits users to find the bases associated with a specific LF. This option is "oriented towards comprehension" (Alonso, Nishikawa, and Vincze 2010: 372), and thus is for decoding purposes. Once again, the queries in the DICE do not seem to respond to the initial linguistic production aims of the authors of the dictionary, which according to Alonso (2010: 65), are oriented towards linguistic production.

In our opinion, this kind of search option is not very useful. Evidently, if users have doubts about the meaning of a word, they would consult a monolingual dictionary, which is easier, and less time consuming than obtaining the meaning of a word through its LF.

Finally, the *help in writing* option verifies whether a specific combination is correct. It permits the following two kinds of consultations: (i) users can ascertain whether a specific base combines with a specific collocate; (ii) users can obtain the collocates for a given base when wishing to express a specific meaning, by means of the gloss. For example, as shown in Figure 31, it permits users to launch the query and

obtain all the verbs that can combine with 'indignación', when it is used as a direct object with the meaning 'causar que la indignación sea mayor'. For this consultation, the system displays the three verbs: 'acrecentar', 'alimentar', and 'aumentar'.

Figure 31. *Ayudas a la redacción* search option in the DICE.

Of the three types of search in the DICE, the option of *ayudas a la redacción* seems to be the only one that is accessible to all users, and which is of great help for encoding. As shall be seen in 5.5, the query system proposed in our research is in consonance with this type of query and allows users to find the answer to their doubts about text production.

Succinctly put, the DICE is the only one of the six dictionaries with free online access. The fact that it is available online means that collocations can be easily accessed by users. It also allows for various types of queries to access collocations, which is another of its major advantages. However, as already mentioned, the metalanguage used makes it unsuitable for a layperson since too much linguistic knowledge is assumed. Another of its problems is it limited number of lemmas in comparison to the other resources.

2.3 Phraseology in terminographic resources

Despite the fact that most terminographers agree that phraseological information in terminographic resources is extremely important, few specialized resources actually

contain word combinations (L'Homme and Leroyer 2009: 260). It goes without saying that those resources that do include them are frequently not consistent in their treatment of collocational information (Montero and Buendía 2012). Nevertheless, the situation is gradually improving and the representation of collocations in specialized dictionaries is becoming more and more frequent.

As previously mentioned, specialized dictionaries that include collocations differ considerably in their method of listing them and representing them in entries. As such, there is still no general agreement as to which word combinations should be included in dictionaries or how they should be described and classified. Apparently, characteristics in which all of them coincide are the following (L'Homme 2009: 239):

(i) collocations are listed under a headword that has already been defined as a term in a specialised subject field;
(ii) the keyword of the collocations is usually a noun or a noun phrase;
(iii) lexical units that typically combine with terms can be verbs, nouns, or adjectives.

In this section a brief description of the most representative terminographic resources that include phraseological information in English, Spanish and French are described. The reason that French resources are included here is that they were the first to include phraseological information in specialized domains. The resources analysed are grouped by domains: (i) domain of accounting/economy/business/law; (ii) domain of computing and the Internet; (iii) domain of the environment; (iv) multiple domains. More specifically, the whole set of dictionaries analysed are the following:

(1) *Lexique de cooccurrents—Bourse et conjuncture économique* (Cohen 1986);
(2) *Vocabulaire et cooccurrents de la comptabilité* (Caignon 2001), a French monolingual dictionary on accounting with English equivalences;
(3) *Dictionnaire d'apprentissage du français des affaires* (DAFA) (Binon et al. 2001), a monolingual French business language dictionary available both in paper and electronic form with equivalences in English, Spanish, German, Dutch, and Italian;
(4) *Dictionnaire contextuel du français économique* (DICOFE) (Verlinde et al. 1993-2003), whose online version allows searches for Dutch-French collocational information for business terms;

(5) *The Accounting Dictionaries* (ADs), a group of five electronic dictionaries with accounting terms;

(6) *Diccionario de términos económicos, financieros y comerciales (inglés-español, Spanish-English),* a paper bilingual English-Spanish dictionary on economics, finance and commerce;

(7) *Internet. Répertoire bilingue de combinaisons lexicales spécialisées français anglais* (Meynard 2000), English-French dictionary for Internet terms;

(8) *Dictionnaire fondamental de l'informatique et de l'Internet* (DiCoInfo), online dictionary in French, English, and Spanish of computing and Internet terms;

(9) *Dictionnaire fondamental de l'environnement* (DiCoEnviro), incipient online dictionary in French, English, and Spanish on the environment, and more precisely on the field of the climate change;

(10) Termium Plus®, a large terminological and linguistic database that covers almost every single field in English, French, Spanish and Portuguese. It is a representative example of a resource with multiple specialized domains.

Most of the specialized resources listed here correspond to the field of business and accounting since as Leroyer (2006: 183) aptly points out:

> The language of written business communication is characterised by the extensive use of phraseology, not only in terms of collocations and idiomatic expressions, but also of standard phrases in prototypical genres.

Apart from describing the main features of the micro and macrostructure of each resource, the focus of our analysis is on how each resource deals with the following: (i) access to collocations; (ii) classification of collocations; (ii) description of collocations. In addition, the positive and negative aspects of each resource are underlined with a view to designing a terminological entry that integrates the good points of these resources and avoids their drawbacks. Because of the various domains covered, it was impossible compare these resources on the basis of entries for the same lemma.

2.3.1 Domain of accounting/economy/business/law

Lexique de cooccurrents—Bourse et conjuncture économique

The *Lexique de cooccurrents—Bourse et conjuncture économique* (Cohen 1986) is a specialized dictionary of collocations in economy (i.e. the stock exchange). It adapts Mel'čuk's methodology to specialized language, and thus uses paraphrases of certain lexical functions to denote the phases of stock market processes (see first column on the left of Table 22).

As shown in Table 22 for the headword 'chomâge' [unemployment], each entry appears in a table with its definition. Collocations are then classified horizontally according to grammatical categories (nouns; verbs for which the headword is the subject; verbs for which the headword is the direct object; adjectives). They are also classified vertically, according to their meaning (the phases of the economic cycle). This includes the paraphrases of lexical functions, namely, *début* [start], *croissance* [growth], *decline* [decline] and *fin* [end]. Apart from these four categories, the dictionary uses two more: (i) *indéterminés* [undetermined] for unspecified temporal movement; (ii) *autres cooccurrents* [other co-ocurrences] for co-occurences that do not fit in the other categories.

For example, as shown in Table 22, the category of *autres cooccurrents* includes units that are semantically quite different, such as 'accidental', 'caché', 'chronique', 'conjoncturel', 'déguisé' or 'frictionnel'. In addition, collocations within each section are listed in alphabetical order and regrouped in consonance with the general sense or the situation that they express. This way of displaying information in tables allows access to collocations at two different levels: (i) base term → syntactic schema of the collocate → semantic description of the collocation → collocation; (ii) base term → semantic description of the collocation → syntactic behaviour of the collocate → collocation (Montero 2002: 210). Collocations are thus classified both according to grammatical category and general meaning.

CHOMÂGE : Manque d'emploi ou diminution involontaire du temps d'emploi pour la main-d'oeuvre.				
	NOMS	**VERBES (SUJET)**	**VERBES (OBJET)**	**ADJECTIFS**
DÉBUT			engendrer provoquer susciter	
CROISSANCE	accroissement aggravation augmentation extension montée poussée	s'accroître s'aggraver augmenter s'étendre monter	accroître aggraver augmenter	considérable élevé
INDÉTERMINÉS	évolution fluctuation stabilisation niveau taux	évoluer fluctuer	se stabiliser	fluctuant variable stable
DÉCLIN	affaiblissement baisse diminution réduction	baisser diminuer	diminuer freiner limiter lutter (contre) réduire restreindre	faible
FIN			maîtriser remédier (à) résorber	
AUTRES COOCCURRENTS			être (au) mettre (au)	accidentel caché chronique conjoncturel déguisé frictionnel partiel récurrent saisonnier structurel technique technologique total

Table 22. Entry of 'chomâge' [unemployment] in the *Lexique de cooccurrents—Bourse et conjuncture économique*[48]

The *Lexique de cooccurrents—Bourse et conjuncture économique* was the first attempt to include phraseological information in a specialized domain. The fact that it classifies collocations by meaning is one of its virtues, although the section *others* is not very informative because it includes any collocation that cannot be included in the other categories. Its main limitations are that it does not have an electronic version and can

[48] <http://www.linguatechediteur.com/lexique-de-coocurrents-exemple-darticle.html> [31/03/2011].

only be consulted in book form. This considerably limits access to information since searches are only permitted from the base term.

Caignon's accounting dictionaries

Caignon is the author of two dictionaries of collocations in the domain of accounting, namely, the *Essential Lexicon in Accounting* (Caignon 2000) and the *Vocabulaire et cooccurrents de la comptabilité* [Vocabulary and co-occurences in accounting] (Caignon 2001). The first is an English accounting lexicon with French correspondences for lemmas, whereas the second is a French accounting lexicon with correspondences in English. Definitions and the rest of the phraseological information are only given in the main language. Since both resources follow the same methodology and present the information in the same way, the *Vocabulaire et cooccurrents de la comptabilité* (2001) is described as an example.

The *Vocabulaire et cooccurrents de la comptabilité* is a specialized French accounting dictionary with English correspondences. Targeted user groups are specialists, students, translators, and accounting teachers. It is basically a monolingual vocabulary since apart from the English correspondences for the headword, the rest of the information contained, including the definition and phraseological information, is only given in French. The dictionary contains a English-French index at the end, which can be of great value for translators.

For example, the entry for 'obligation[49]' is composed of a short definition, synonyms, abbreviations and the English correspondences (it specifies *G.B.* for British English, and *E.U,* for American English, when there is geographical variation). It also provides linguistic and encyclopaedic information (see Table 23).

Regarding collocations, the dictionary classifies them according to grammatical category: (i) nouns; (ii) verbs with the headword as the subject; (iii) verbs with the headword as the direct object; (iv) adjectives. There is sometimes a category for adverbs and another called *autres* [others] for those collocations that do not fit in the rest of categories. It also includes collocations that can be found in texts, but which are apparently incorrect. These are identified with an asterisk "*".

[49]<http://www.linguatechediteur.com/vocabulaire-et-coocurrents-de-la-comptabilite-exemple-darticle.html> [06/04/11].

OBLIGATION (n.f.) Titre d'emprunt collectif qu'une collectivité publique ou qu'une entreprise émet lorsqu'elle fait une demande d'emprunt à long terme, représentant un montant remboursable à terme et procurant un revenu fixe à son porteur.
SYNONYME : bon[1]
TERMES ANGLAIS : bill[2] (É.-U.) bond[2] debenture (G.-B.) debenture stock (G.-B.) note[2] (É.-U.)
REMARQUES : 1. Ce synonyme n'est pas perçu comme un anglicisme, mais plutôt comme un terme dont l'usage se trouve surtout dans le secteur public, par exemple *bon du trésor*. 2. Selon le *Dictionnaire de la comptabilité et de la gestion financière* (1994, p. 84), aux États-Unis, le terme *note* renvoie de nos jours aux obligations émises par l'État fédéral et par les entreprises – principalement aux obligations à moyen terme. Ce terme s'oppose ainsi à *bill*, qui représente des obligations à court terme, et à *bond*, qui désigne des obligations à long terme.
COOCCURRENTS : **noms:** achat d'~s, accroissement des ~s, appréciation d'une ~, augmentation des ~s, baisse d'une ~, certificat d'~, chute d'une ~, *coupon d'une ~, émetteur d'~s, émission d'~s, fléchissement d'une ~, montant d'une ~, montée d'une ~, porteur d'~s, prix d'une ~, progrès d'une ~, taux des ~s, total des ~s, variation d'une ~, vente d'~s **verbes** (sujet)**:** ~ augmenter, ~ baisser, ~ chuter, ~ comprendre, ~ coter, ~ coûter, ~ donner, ~ être rémunérée, ~ fléchir, ~ monter, ~ progresser, ~ s'apprécier, ~ s'échanger **verbes** (complément)**:** acheter des ~s, acquérir une ~, arbitrer une ~, céder une ~, constater une ~ par un titre, détenir une ~, échanger une ~ [contre], émettre des ~s, liquider une ~, négocier des ~s, posséder des ~s, remettre une ~, se composer d'~s, souscrire une ~, traiter une ~, vendre des ~s **adjectifs:** *~ à court terme, *~ à long terme, ~ cautionnée, ~ chère, ~ convertible, ~ échangeable, ~ émise, *~s étrangères, ~ garantie

Table 23. Entry for 'obligation' in the *Vocabulaire et cooccurrents de la comptabilité*

As shown, Caignon's dictionary includes extensive phraseological information, which is very valuable. However, although it is supposed to be for translators, only the correspondences of the headword are given in the other language. No correspondences for the collocational information are included. In our opinion, a resource for translators should contain comparable information in the two languages. In addition, the fact that collocations are only classified in terms of part of speech (POS) and not according to meaning can make it difficult for users, especially for non-native speakers of French, to choose the correct collocate (e.g. the difference between 'obligation baisse' and 'obligation chute'). In this regard, usage examples that show the word in context would have been very useful.

Dictionnaire contextuel du français économique (DICOFE)

The *Dictionnaire contextuel du français économique* (Verlinde et al. 1993-2003)**,** henceforth DICOFE, is a combinatorial dictionary which focuses on French business language developed by the research group on Pedagogical Lexicography *(Groupe de Recherche en Lexicographie Pédagogique, GRELEP)* of the Catholic University of Leuven. It is composed of four volumes covering the areas of *l'entreprise* [company], *le commerce* [commerce], *les finances* [finance], and *l'emploi* [employment]. The structure of the dictionary is not alphabetical. On the contrary, it is conceived as an introductory text to the world of economics in which, as users acquire conceptual knowledge, they also acquire lexical knowledge about the specialized units at the same time. To that end, in the paper version of the DICOFE, even and odd pages have different functions.

In Figure 32, an extract of an odd page for everyday financial operations ('les operations financières courantes') is shown as an example[50]. The various operations are explained with sentences, collocations, and more complex combinations of words, typical of this particular field:

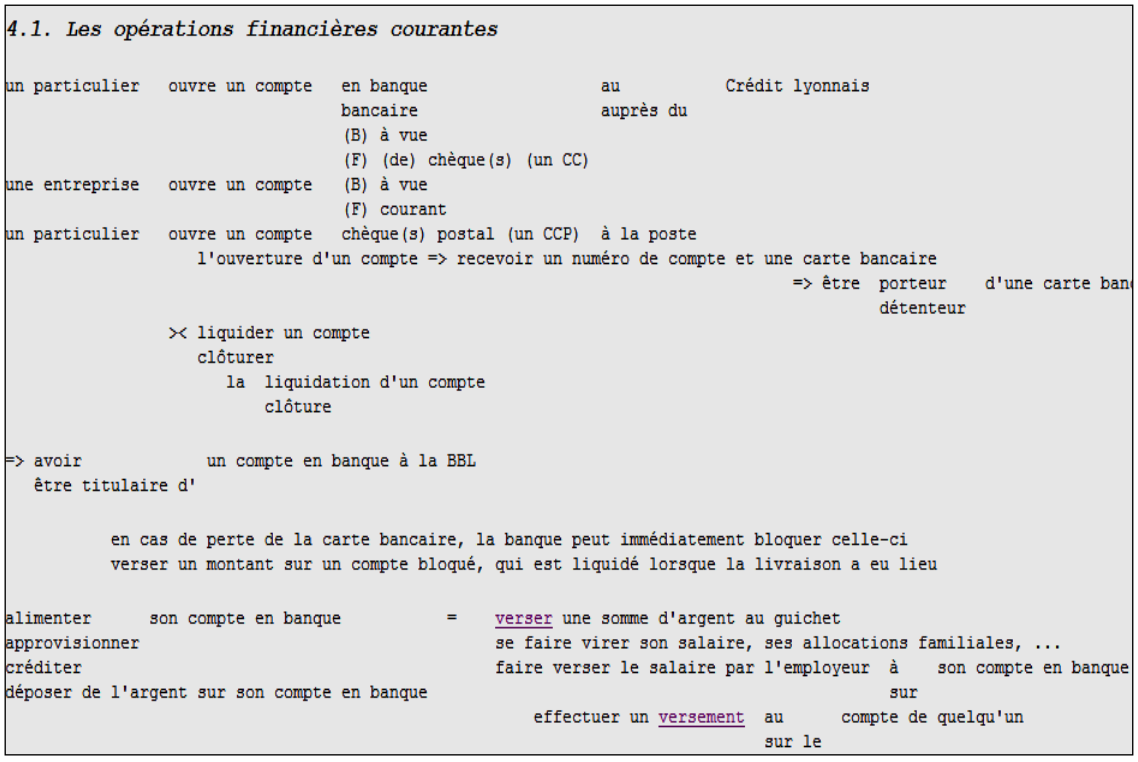


Figure 32. Extract of an odd-numbered page from the DICOFE

[50] All the screenshots displayed in the section were extracted from the website of the research group GRELEP: <http://ilt.kuleuven.be/Grelep/projets/dicofein.htm> [20/03/2011].

As can be seen, the text is divided into sections in order to identify the various elements in each sentence (either a collocation or a complex combination) and to include the variants for each particular case. For example, in the collocation ‘effectuer un versement’, prepositional phrases such as ‘au compte de quelqu’un’ or ‘sur le compte de quelqu’un’ can be used. In order to describe the semantic relations between the different units, the dictionary uses symbols, such as “=” for synonymy; “=>” for cause-effect; “X” for antonyms, etc. In addition, when necessary, geographical variants from Belgium and France are identified with a (B) and (F), respectively.

Pages even display the terminological entries for some of the specialized units in the examples of the odd pages. These units are grouped by meaning, and each of their components appears with a definition, headed by a number (1, 2, 3); an example (*ex*); and collocations and other combinations of words that were not listed on the corresponding odd page (1.1, 1.2, 1.3). This microstructure is shown in Table 24, which displays the terminological entry for the terms ‘verser’ [to pay] and ‘versement’ [payment]:

verser -versement		
substantif/nom qui	quoi	Verbe
un versement		verser
un versemement 1. Action d’apporter de l’argent à une personne, à une caisse afin de payer (à ne pas confondre avec un *virement*) ex. Veuillex effectuer un versement de 1.000 euros en faveur de notre association 1.1. un versement anticipé (V. L’impôt sur le revenu) 1.2. un bulletin de versement *verser* 1.1. verser une somme d’argent à/sur un compte		

Table 24. Extract of an even-numbered page from the DICOFE

As shown in Table 24, the deverbal noun ‘versement’ is derived from ‘verser’. Each element appears in a different sub-entry. The entry for ‘un versement’ includes its definition, a usage example, and two noun phrases, ‘un versement anticipé’ and ‘un bulletin de versement’, which does not appear in the entry on the odd page (see Figure 32). The sub-entry ‘verser’, includes a phraseme that was not previously displayed, ‘verser une somme d’argent à/sur un compte’.

In order to facilitate the consultation of the dictionary, an exhaustive index of all the terms that appear in the introductory texts is included. Table 25 shows an extract from the index for ‘banque’ [bank]. As can be observed, each term is accompanied by

information concerning its frequency (asterisks), grammatical category, a list of all of its collocations and other noun phrases in which the headword appears, and finally, the page(s) of the dictionary where more information can be found:

*** banque (nf) 60; 7; 17; ...
une — centrale 7; 30; ...
une — commerciale 7; 30; 55
la Banque mondiale 59
une — mutualiste 61
...
un billet de — 7; 55; ...
un compte auprès d'une — 63; 64
un compte en — 7; 18; 21; ...
...
émettre des billets de — 7, 55
encaisser un chèque auprès d'une — 65
mettre des billets de — en circulation 7
...
la mise en circulation de billets de — 7
**** banquier, -quière (n) 60; 30
** barème (nm) 49

Table 25. Extract corresponding to the alphabetical index of the DICOFE

An electronic version of the dictionary can be accessed at the webpage <http://www.kuleuven.be/grelep/dicofe/>, but it is somewhat different from the paper version of the dictionary. It is a kind of bilingual glossary that permits searches from Dutch to French, but not the other way round. Figure 33 displays the interface of the online version of the DICOFE:

Figure 33. Interface of the online version of the DICOFE

For example, when looking up the term 'financiering' in Dutch [financial operation] (see Table 26), the system displays 35 occurrences of the search word. As

can be observed in the middle column of Table 26, the occurrences in boldface are lemmas and the ones in normal typeface are multiword expressions. The column on the left of the occurrences displays the volume in the written version of the dictionary where the lemma or multiword expression is found. The column on the right gives the French translation of each lemma or multiword expression. In addition, this column shows the grammatical category and the page/s in the paper dictionary where the lemma or multiword expression is found. For multiword expressions, the lemma associated with the particular expression is provided between hash symbols (#).

Tome		***entrée***; (catégorie grammaticale); renvoi aux pages *combinaison de mots* (#entrée#); renvoi au x pages
A	**financiering** (v)	**financement** (nm) 30
C	de financiering van een aankoop	le financement d'un - 23 #achat#
C	**zelffinanciering** (v)	**autofinancement** (nm) 8; 33; 38; 42
C	een financieringsbehoefte	un - de financement 29; 71 #besoin#
C	het brutofinancieringstekort	le solde - à financer 52 #brut, brute#
C	de financiering van het begrotingstekort	le financement du déficit - 53 #budgétaire#
C	de financiering van het begrotingstekort	le financement du - budgétaire 53 #déficit#
C	**financiering** (v)	**financement** (nm) 8; 23; 29; 31; 33; 44; 53; 55; 71
C	de financiering van een aankoop	le - d'un achat 23 #financement#
C	een kortlopende financiering	un - à court terme 8 #financement#
C	de financiering van het begrotingstekort	le - du déficit budgétaire 53 #financement#
C	de financiering van een investering	le - d'un investissement 33 #financement#
C	een langetermijnfinanciering	un - à long terme 8 #financement#
C	een middellange (termijn)financiering	un - à moyen terme 8 #financement#
C	een financieringsbehoefte	un besoin de - 29; 71 #financement#
C	een financieringscapaciteit	une capacité de - 71 #financement#
C	financieringsmiddelen	un moyen de - 55 #financement#
C	een financieringsbron	une source de - 29 #financement#
C	het nettofinancieringstekort aanvullen; bijpassen	combler le solde net à - 53 #financer#
C	een nettofinancieringstekort vertonen	enregistrer un solde net à - 53 #financer#
C	het brutofinancieringstekort	le solde brut à - 52 #financer#
C	het nettofinancieringstekort	le solde net à - 52; 53 #financer#
C	de financiering van een investering	le financement d'un - 33 #investissement#
C	het nettofinancieringstekort	le solde - à financer 52; 53 #net, nette#
C	het nettofinancieringstekort aanvullen; bijpassen	combler le solde - à financer 53 #net, nette#
C	een nettofinancieringstekort vertonen	enregistrer un solde - à financer 53 #net, nette#
C	het brutofinancieringstekort	le - brut à financer 52 #solde#
C	het nettofinancieringstekort	le - net à financer 52; 53 #solde#
C	het nettofinancieringstekort aanvullen; bijpassen	combler le - net à financer 53 #solde#
C	een nettofinancieringstekort vertonen	enregistrer un - net à financer 53 #solde#
C	een kortlopende financiering	un financement à court - 8 #terme#
C	een langetermijnfinanciering	un financement à long - 8 #terme#
C	een middellange (termijn)financiering	un financement à moyen - 8 #terme#
D	**autofinanciering** (v); zelffinanciering (v)	**autofinancement** (nm) 36
D	**financiering** (v)	**financement** (nm) 36

Table 26. Results for the search ‘financiering’ in the DICOFE

Accordingly, for the expression 'een kortlopende financiering' [short term financing], the DICOFE specifies that the translation into French is 'un financement à court' and that more information can be found in page 8 of volume C of the lemma *terme*.

The paper version of the DICOFE is a very valuable repository for knowledge and lexical acquisition since users can access collocations either by means of the index or based on subject matter. In other words, it is of tremendous help for decoding purposes. In contrast, the online version especially focuses on encoding, and includes a long list of phraseological combinations for each headword. However, as pointed out by L'Homme and Leroyer (2009: 265), "If lists are long, users might experience problems accessing the specific combination for which they are looking". In addition, the fact that searches can only be launched from Dutch to French makes term access for French native speakers and learners of French even more difficult.

Dictionnaire d'apprentissage du français des affaires (DAFA)

The *Dictionnaire d'apprentissage du français des affaires* (Binon et al. 2001), henceforth DAFA, is a reference dictionary of the French business language also compiled by the GRELEP research group of the Catholic University of Leuven. The target user groups are French native speakers who wish to improve their knowledge and language skills in the world of business, intermediate/advanced learners of French business administration, business language teachers, translators and interpreters, and anyone interested in French business language.

Words in the DAFA are organized in categories so that each article describes a semantic field. All the words in the alphabetical index are listed under 135 lemmas. The dictionary includes entries for more than 3,200 words and 11,000 word combinations. It also has over 3,000 examples of sentences extracted from a specialized corpus of 25 million words. It also includes usage notes and the main geographical variants for Belgium, France, Switzerland, and Canadian French.

In order to facilitate term access for non-native French speakers, the word list appears in five languages, namely, German, English, Spanish, Italian and Dutch. Thus, it is also helpful for natives of French who wish to translate a text into one of these five languages. Searches can be performed either by domain or by exact matching in one of the languages. However, collocations and other linguistic information are only given for French terms.

Table 27 shows an extract of the entry for 'commerce[51]' [trade]. The information included concerns the first sense and first meaning, headed by 1.1, (two other meanings are described). As can be seen, the entry first offers cross-references to other words that are relevant for comparison and additional knowledge ('vente', 'achat', 'marché', 'marketing'). As previously mentioned, this resource also includes the words derived from the lemma *(le commerce, la commercialisation, le commercial, un commerçant, une commerçante, un commercial, une commerciale, commerçant, -ante, commercial, -iale; -iaux, -iales, commercialisable, commercialement, commercer, commercialiser).* After giving the term's phonetic representation and grammatical category, the dictionary offers the definition, a usage example, and its collocations.

Collocations are listed according to the part of speech of the collocates: *+adjectif* (for adjectives), *+nom* (for nouns) and *+verbe* (for verbs). As shall be seen in Table 28, the verbs for each lemma are displayed in a separate table. All collocates in a lemma are further classified according to their meaning, which is based on lexical functions. Although greatly simplified, the lexical description is in line with the linguistic description developed by Mel'čuk within the framework of the Meaning Text Theory (§2.1.3.1).

For example, as shown in Table 27, the adjective collocates include a subsection *LOCALISATION DU COMMERCE*. This shows the various phraseological units for 'commerce', when it combines with adjectives that specify geographic location. Examples include 'le commerce mondial/international' [world/international trade], 'le commerce extérieur' [foreign/external trade], and 'le commerce intérieur, national, domestique' [domestic trade]. In addition, for certain collocations (e.g. 'le commerce électronique'), a more detailed explanation is given (*commerce à l'aide de la télématique —combinaison de l'informatique et des moyens de communication—*), as well as a usage example ('Lorsque l'on parle de commerce électronique, on pense généralement au World Wide Web, mais le courrier électronique peut également devenir un média privilégié pour commercer sur le Net') (see Table 27):

[51] <http://www.kuleuven.be/grelep/projets/COMMERCE.PDF> [21/03/2011].

COMMERCE

→ **vente - achat - marché**
→ **marketing**

1 le commerce	3 un commerçant,	4 commerçant, -ante	6 commercer
2 la commercialisation	une commerçante	5 commercial, -iale;	6 commercialiser
6 le commercial	6 un commercial,	-iaux, -iales	
	une commerciale	6 commercialisable	
		6 *commercialement*	

1 le COMMERCE - [k*m£Rs(!)] - (n.m.)
1.1. Activité (de service) d'un agent économique (un commerçant, une entreprise, parfois un État - X) qui
consiste soit à acheter des marchandises ou des valeurs (Y) pour les (re)vendre ou les louer à un client (Z) sans y apporter de transformation matérielle, soit à proposer des services (Y).
♦ Syn.: (**+** Pour en savoir plus, Commerce (sens 1.1.) et synonymes).
Le commerce naît lorsque les produits bruts ou transformés font l'objet d'échanges: produits contre produits,produits contre valeurs ou valeurs contre valeurs (Gaeng).
+ adjective

• TYPE DE COMMERCE (sens 1.1.)
Le commerce électronique: commerce à l'aide de la télématique (combinaison de l'informatique et des moyens de communication). *Lorsque l'on parle de commerce électronique, on pense généralement au World Wide Web, mais le courrier électronique peut également devenir un média privilégié pour commercer sur le Net*.
Le commerce sédentaire: se caractérise par un point de vente fixe.
>< **Le commerce non sédentaire**, comme p. ex. **le**
commerce ambulant: commerçants présents sur les marchés ou les foires.
+ nom

• CARACTÉRISATION DU COMMERCE (sens 1.1.)
Un commerce lucratif: qui rapporte beaucoup.

• LOCALISATION DU COMMERCE (sens 1.1.)
Le commerce mondial.
Le commerce extérieur, **international**: achat (l'importation) et vente (l'exportation) de produits international). (**+** Pour en savoir plus, Les termes commerciaux internationaux).
>< **Le commerce intérieur**, **national**, **domestique**: achat et vente de produits à l'intérieur d'un pays.

(sens 1.1.)

• **Un représentant de commerce**. (V. 6 autres dérivés ou composés).

• **Une maison de commerce**: entreprise commerciale traditionnelle, souvent familiale. *La maison de commerce japonaise Sumitomo a pris une participation de 20 % dans le capital de Hamilton Standard Space Systems International, un groupe américain spécialisé dans les technologies spatiales*.

• **Le code de commerce**: ensemble des lois et des
règlements qui s'appliquent au commerce. *Le code*
de commerce exige que, dès le début de son activité commerciale, tout commerçant demande son immatriculation au Registre du Commerce et des Sociétés auprès du greffe du Tribunal de Commerce compétent (Gaeng).
(B) **Le Registre de Commerce** (**le RC**), (F) **le Registre du Commerce et des Sociétés** (**le RCS**),
(S) **Le Registre du Commerce** (**le RC**): registre où sont centralisées certaines informations sur les entreprises et les sociétés commerciales (Wagner).

commercial et industriel et qui représente leurs intérêts.

• **Le ministère du Commerce extérieur**; **le ministre du Commerce extérieur**.
L'Organisation mondiale du commerce (**l'OMC**): organisation internationale créée par l'Uruguay Round destinée à se substituer au GATT pour gérer l'accord général sur les tarifs douaniers et le commerce (Silem). (**+** Pour en savoir plus, Le commerce international).

• **Une école** (**supérieure**) **de commerce**.

• TYPE DE COMMERCE (sens 1.1.)
Le commerce de + nom qui désigne une marchandise ou une valeur. Le commerce du bois, du charbon.
Le commerce (**mondial**) **des marchandises**. (V. marchandise, 1).
Le commerce des services.
Le commerce de gros: commerce qui consiste à acheter des marchandises par quantités importantes et à les vendre à des revendeurs, détaillants ou grossistes. *La flotte d'agents commerciaux qui représentent le commerce de gros sont les gros utilisateurs des nouvelles technologies de communication*. {**un**, **une grossiste**}.
>< **Le commerce de détail**: commerce qui

Cette semaine, quinze nouvelles sociétés ont demandé leur immatriculation au Registre de Commerce.
Le tribunal de commerce: traite les contestations
entre commerçants.

- **Un effet de commerce**: tout document par lequel un tireur (un créancier) donne l'ordre à un tiré (un débiteur) de payer le bénéficiaire à l'échéance de la dette (p. ex. une lettre de change, une traite (**tirer une traite sur qqn**); un billet à ordre). (V. paiement, 1). **Un bénéficiaire d'un effet de commerce**. (V. bénéfice, 2). **Un effet de complaisance**, **de cavalerie**. (V. effet, 1).
Le protêt est l'acte authentique par lequel le porteur d'un effet de commerce fait constater que cet effet n'a pas été accepté par le tiré ou qu'il n'a pas été payé à l'échéance (RQ).
- **Les livres** (**de commerce**). (Syn.: **les documents comptables**). (V. comptabilité, 2).
- **Une chambre de commerce et d'industrie**: établissement public autonome, géré par les représentants élus des entreprises du secteur

consiste à acheter des marchandises pour les revendre au consommateur ou à l'utilisateur final, en général par petites quantités (Gaeng). (Syn.: **la distribution**). *Les difficultés de circulation dans les villes et les difficultés de stationnement limitent le développement du commerce de détail*.
{**un détaillant**, **une détaillante**, **détailler**}.
Le commerce (**de gros**) **en libre-service**: commerce (où la vente se pratique dans un entrepôt de gros et) où le client se sert lui-même.
Le commerce de dépôt-vente: commerce où les marchandises sont déposées par des particuliers et vendues par le commerçant, qui perçoit une commission sur le prix de vente.

- MESURE DU COMMERCE (sens 1.1.)

Le volume du commerce. *L'essor économique rapide des pays asiatiques a beaucoup contribué à la croissance du volume du commerce mondial des produits manufacturés*.

Table 27. Noun and adjective collocations for the sense 1 of 'commerce' in the DAFA

Verbs are displayed in a separate table (cf. Table 28). They are organized according to their actantial structure and their general meaning, which is also based on Mel'čuk's lexical functions. As is well-known, the MTT refers to predicate argument structure as *actantial structure* and to arguments as *actants*[52].

As previously mentioned (see Table 27), predicate argument structure is specified in the definition which states that 'commerce' has three arguments, represented by the variables X, Y and Z. These arguments are defined as generic terms that are regarded as typical arguments (e.g. the first argument X is an economic agent, namely a trader, enterprise, and sometimes, a country, etc.).

As shown in the list of verb collocations in Table 28, all of the arguments are appear in the very first column on the left of the table. The middle column contains the verbs and constructions that collocate with these arguments. Finally, in the column on the right, nominalisations of verb forms, if they exist, are given. For example, in 'faire du ~ avec Z' *(Z: souvent un État)*, the subject (X) is generally 'une entreprise', 'un État' [an enterprise or a country]. Finally, an example is provided at the bottom of the table (e.g. 'Pour des raisons politiques, certains pays font moins de commerce avec la Chine').

[52] However, in this thesis, we have used the term arguments to refer to the participating entities in a verb's argument structure.

+verbe : qui fait quoi ? (sens 1.1.)				
X (une entreprise)	**faire** (le) ~ **de** Y (**avec** Z)		-	1
X (une entreprise, un État)	**faire du** ~ **avec** Z (Z : souvent un État)		-	2
une mesure politique un accord	**stimuler** le ~ (**de** X) **avec** Z/**entre** X **et** Z **encourager** le ~ **de** X **avec** ... >< **entraver** le ~ **de** X **avec**... (X, Z : un État)		la stimulation du ~ (de X) avez Z/entre X et Z - -	3
une mesure politique un accord	**libéraliser** le ~ (de Y) >< **réglementer** le ~ (de Y)		la libéralisation du ~ (de Y) la réglementation du ~ (de Y)	4
le commerce (de Y) le commerce (de Y)	+ ++	**se développer** **connaître une croissance** **être en pleine expansion**	le développement du ~ (de Y) la croissance du ~ (de Y) une expansion du ~ (de Y)	5
le commerce (de Y)	+=	**stagner**	la stagnation du ~ (de Y)	6
le commerce (de Y) le commerce (de Y)	, , ,	**régresser** **s'effondrer**	la régression du ~ (de Y) un effondrement du ~ (de Y)	7
1 *Un négociant doit posséder une carte de négociant-manipulant pour obtenir le droit de faire le commerce de champagne*. - 2 *Pour des raisons politiques, certains pays font moins de commerce avec la Chine*. - 3 *L'instauration d'un code de concurrence loyale permettrait de stimuler le commerce entre les pays industrialisés et certains pays en développement*. - 4 *L'Organisation mondiale du commerce doit mettre fin au protectionnisme économique et veiller à la libéralisation du commerce mondial*. - 5 *Les importations massives pour éviter la sous-alimentation sont un des facteurs déterminants de la vigoureuse expansion du commerce international de denrées alimentaires*. - 6 *Le commerce international a stagné une première fois de façon significative au début des années 80*. - 7 *Certaines informations alarmantes concernant la qualité de la viande ont causé un effondrement du commerce de la viande.*				

Table 28. Verbal collocations for sense 1.1 of 'commerce'

The DAFA also has an electronic version (http://www.projetdafa.net) (see Figure 34). Like the printed version of the dictionary, searches can be performed by domain or by an exact matching of the word in French, English, Spanish, Italian, German or Dutch. Although the DAFA permits users to find correspondences in all of these languages (and is thus useful for translators), it is basically a monolingual dictionary in the sense that detailed information is only provided for French terms.

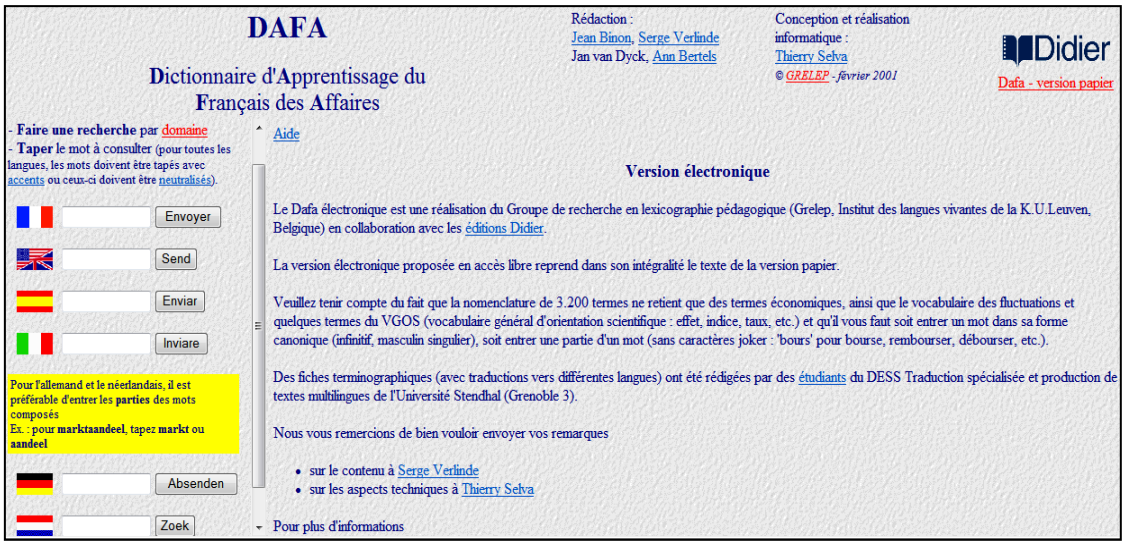


Figure 34. Search interface of the DAFA online version

Collocations are not displayed all at once (Figure 35), but are rather organized in terms of grammatical category: *+adjective* (for adjective collocates); *+nom* (for noun collocates) *+verb* (for verb collocates). As in the printed version, each category is subdivided into meaning areas for each sense. For example, the entry for 'commerce' is shown in Figure 35. By clicking on the various hyperlinks, users can visualize the corresponding list of collocations[53]. In the same way as in the printed version, verbs are displayed on a separate table.

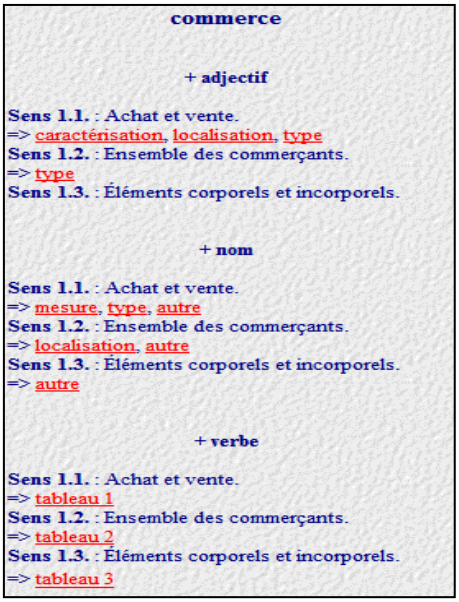


Figure 35. Classification of the collocations for 'commerce' in the DAFA

[53] The whole list of collocations for each subdivision is not included since it corresponds with collocations offered in Tables 27 and 28 for the printed version of the dictionary.

The way that collocations are described in the DAFA is an example for other resources to follow. Of the dictionaries listed in this section, the DAFA is one of the resources that provides the best collocational information in each entry. It classifies collocations in relation to their meaning, and includes a definition as well as a usage example. In our opinion, these three parameters are obligatory for each collocation within a lemma. The DAFA also gives preferential treatment to verb collocates, which in itself is proof of the importance of this type of information. Nevertheless, the metalanguage used by the DAFA, especially for verb description, though comprehensible for linguists, is not very user-friendly. In the words of Leroyer (2006: 187):

> The DAFA is both an impressive and at the same time disappointing dictionary: impressive because of the extensiveness of the phraseological information, disappointing because of the very limited interlingual assistance and the linguistic access structures. Thus, the main problem does not lie in the analytic quality of the phraseological information itself, but in the adaptation of the phraseological data types to the planned dictionary functions. […] For the non-native users, the general problem remains the easy access to the information needed because the concept presupposes a very high degree of proficiency in the foreign language.

The Accounting Dictionaries: Diccionario inglés-español de contabilidad

The Accounting Dictionaries, henceforth ADs, are a set of electronic dictionaries in accounting developed by the Centre for Lexicography (University of Aarhus, Denmark) designed to assist both native and foreign language learners with accounting terminology. According to Fuertes-Olivera and Niño-Amo (2011), these ADs are both descriptive and normative: descriptive in that they include various spellings (even misspelt variants that potential users can come across in accounting texts); and normative, since they advise their potential user of correctness, and make recommendations regarding spellings, and preferred variants.

Currently, there are five dictionaries in the network available online[54], namely, (1) *Den Danske Regnskabsordbog* [The Danish Dictionary of Accounting]; (2) *Den Dansk-Engelske Regnskabsordbog* [The Danish-English Dictionary of Accounting]; (3) *Den Engelske Regnskabsordbog* [The English Dictionary of Accounting]; (4) *Den Engelsk-*

[54]<http://www.asb.dk/en/research/researchcentresandteams/researchcentres/centreforlexicography/centreprofile/> [02/04/2011]

Danske Regnskabsordbog [The English-Danish Dictionary of Accounting]; and (5) *The English-Spanish Dictionary of Accounting*. Currently in elaboration is the *Diccionario Español-Inglés de Contabilidad* [The Spanish-English Dictionary of Accounting] and the *Diccionario Español de Contabilidad* [The Spanish Dictionary of Accounting]. Since all of these dictionaries follow the same methodology and display collocations in the same way, the *Diccionario inglés-español de contabilidad* is described as an example.

The *Diccionario inglés-español de contabilidad*[55] [The English-Spanish Dictionary of Accounting] has been elaborated at the University of Valladolid (Spain), thanks to the cooperation of two research groups: one from the Centre for Lexicography, mainly composed of Bergenholtz, Nielsen and Almid, and another one from the University of Valladolid, including researchers such as Fuertes-Olivera, Gordo Gómez, Niño-Amo, De los Ríos Rodicio, Sastre Ruano, Tarp, and Velasco Sacristán. The main objective of the dictionary is to help users to read, understand, and produce accounting texts, as well as to translate English accounting texts into Spanish and acquire knowledge about accounting (Fuertes-Olivera and Niño-Amo 2011). Consequently, it is conceived for both encoding and decoding purposes.

Special attention is given to information access. Consequently, four ways of accessing the information are provided. As shown in Figure 36, once users enter the keyword in the search box, they can choose various options from the menu on the left. Such options include the following: (i) "is" retrieves the records in which the word appears as a lemma; (ii) "begins with", "contains" and "ends" displays all of the entries whose lemma starts, contains, or ends with the word entered in the search box. For example, when users look up 'market' and choose the *begins with* option, the dictionary displays 31 instances, such as 'market', 'market cap', 'market capitalisation', 'market capitalization', 'market communications', etc. However, unlike the DiCoInfo or DiCoEnviro (see below), it does not allow direct searches by collocation. In other words, if users are seeking a phraseological unit (e.g. 'volatile market'), they must first access one of the words in the combination (i.e. 'volatile' or 'market') and then scroll down the list of collocations provided for the lemma in order to retrieve the desired combination.

[55]< http://www.accountingdictionary.dk/regn/gbsp/regngbsp_index.php> [02/04/2011]

Figure 36. Search interface of the *Diccionario Inglés-Español de Contabilidad*

Regarding the microstructure of 'market' with a special focus on collocations (see Table 29), the dictionary offers grammatical information (e.g. it specifies whether the headword has a plural form, should be used with a definite/indefinite article, etc). Subsequently, it gives a brief description in English of the concept, and its translation into Spanish ('mercado') and synonyms, if there are any. The next section is for collocations. In this regard, each lemma lists the most representative collocations for the headword along with their equivalences in Spanish.

Collocations are provided in the form of typical phrases containing the headword. They can either be in the form of a short phrase (e.g. 'an active market' > 'un mercado activo'), or a long one (e.g. 'affected by the thinness of the market' > 'afectado por la inestabilidad del mercado'). Therefore, the dictionary includes complex units as headwords. In other words, units such as 'active market', which appear as a headword, can also be listed as a collocation in entries such as 'market' (see Table 29).

market <noun a, the, -s>
definition A market is a place, whether physical or electronic, where transactions take place between sellers and buyers, i.e. where goods and services are bought and sold. A market is usually governed by supply and demand which means that the prices of goods and services are determined by the available supply and the number of buyers.
Mercado
collocations • a difficult market un mercado difícil • a fluctuating market un mercado fluctuante • a geographical market un mercado geográfico • a homogenous market un mercado homogéneo • a volatile market un mercado volatil • affected by the thinness of the market afectado por la inestabilidad del mercado • an active market un mercado activo • an economically comparable market un mercado económicamente comparable • consolidate one's position on the market consolidar su posición en el mercado • demonstrate the existence of a market demostrar la existencia de un mercado • develop a market desarrollar un mercado • gain a foothold on a market asentarse en un mercado • introduce to the market introducir en el mercado • leave a market dejar un mercado • the items traded within the market los artículos comercializados dentro del mercado
example The enterprise should be able to demonstrate the existence of a market for the output of its intangible asset. La empresa debería poder demostrar la existencia de un mercado para la salida de su activo intangible.

Table 29. Entry of 'market' in the *Diccionario de contabilidad inglés-español*

In addition, usage examples appear at the end of the entry with their Spanish translation. When needed, usage notes are included, which explain differences between English and Spanish. In certain cases, the entry may also include notes regarding the nature of the Spanish accounting system or information concerning recent regulations and legislation. When necessary, the dictionary also provides the variants for the International Accounting System (as stipulated by the International Accounting Standards Board) or the British or American variants for terms by means of the tags *IAS/IFRS*, *UK*, and *US*, respectively. It also includes the label *not recommended, use instead* for terms that can be found in accounting texts, but which have more suitable alternatives. It also adds cross-references to other lemmas by means of the tag *see also*. Finally, some dictionary records incorporate a hyperlink (*source)* that links the article to texts where more information can be retrieved.

Among the most positive aspects of the dictionary are the extensive phraseological information it offers for each lemma and the fact that it provides translations for every phraseological unit. This dictionary thus affords valuable information for users who seek a corresponding expression in Spanish. However, when browsing through articles (especially when lists of collocations are very long), it is hard to perceive how collocations are organized or sorted. This means that users may encounter problems when trying to retrieve the combination that they are looking for. In fact, no type of classification is provided and collocations are merely listed alphabetically within lemmas. Evidently, since translators are the targeted user group, more usage examples should have been provided.

As previously mentioned, collocations cannot be extracted directly. The *contains* search option could help in the retrieval of collocations since it displays all the articles where the search term appears. However, this only seems to work when searching for the base. For examples, when users looking for ‘a difficult market’ enter ‘market’ in the search box with the *contains* option, they obtain ‘a difficult market’, among other information. However, when they introduce ‘difficult’ in the search box (again with the *contains* options), no results are obtained.

Diccionario de Términos económicos, financieros y comerciales (inglés-español, Spanish-English)

The *Diccionario de términos económicos, financieros y comerciales (inglés-español, Spanish-English)* (Alcaraz-Varó and Hughes 2008)[56], henceforth DTEFC, is an English-Spanish bilingual dictionary of economic, financial, and trade terms, which continues and complements the *Diccionario de términos jurídicos (inglés-español, Spanish-English)* (Alcaraz-Varó and Hughes 2007)[57] developed by Alcaraz-Varó during the 1990s. Other bilingual dictionaries compiled by this research team are Alcaraz-Varó (2006ab); Alcaraz-Varó and Castro-Calvín (2007); Campos Pardillos and Alcaraz-Varó (2002); Castro-Calvín and Alcaraz-Varó (2003); and Mateo-Martínez and Alcaraz-Varó (2003).

The DTEFC targets translators, students, and professionals of the various branches of economics, business, and law, along with journalists and entrepreneurs. It contains about 25,000 entries, including both American and British English variants. For its compilation, newspapers, such as *The Financial Times, The Wall Street Journal, the International Herald Tribune, The Economist* (for English), and *Cinco Días, Expansión, Gaceta de los Negocios, El País de los Negocios, ABC, El Mundo, Diario 16, La actualidad económica, Ranking* (for Spanish) were used. In addition, a large number of specialized texts from numerous resources were consulted for searching and validating examples and definitions. As is highlighted in the introduction of the dictionary, law terms were excluded from the DTEFC, since these terms were already in the *Diccionario de Términos Jurídicos*.

The DTEFC has two sections: (i) English-Spanish; (ii) Spanish-English. Although in theory, the two parts should be similar in length, this is not case. The reason is that some of the terms or phrases that are simple or polylexical units either in English or Spanish are translated as a paraphrase in the other language. Most of the lemmas in the dictionary are simple or compound lexical units, although there are some syntactic or periphrastic units, especially in the Spanish-English part, e.g. ‘opción extinguible de acuerdo con las fluctuaciones máximas o mínimas del precio del activo subyacente’.

The DTEFC gives a detailed description of each lemma by including a wide variety of combinations associated with the lemma. For example, ‘market’ has a total of 116 different combinations (see Table 31). This is the reason why Table 30 only

[56] It corresponds to its 5th edition.
[57] It corresponds to its 10th edition. The first issue was published in 1992.

displays an extract from the headword 'market' as described in the DTEFC. As shown in Table 30, after specifying the grammatical category of the headword, the various translations of the headword are provided, headed by the subdomain to which they belong (e.g. COMER, stands for commerce; FINAN, for finances; and BOLSA for the stock exchange). The translations are separated by commas when they are regarded as synonymous, and by a semicolon when they refer to different senses. Subsequently, usage examples are provided after the symbol ◊ (e.g. 'A market can be a network of buyers and sellers who deal with each other over a computer screen'). Entries in the DTEFC also include cross-references to other lemmas headed by "V", which stands for *véase* [see]. Finally, the various combinations with the headword are listed.

Word combinations are highlighted in bold typeface. The information displayed for each combination has the same structure as the information provided for the headword to which the construction belongs (i.e. domain and translation of the phraseological unit, usage examples, cross-references to other word combinations). Most of the constructions for each headword in the DTEFC are noun compounds. In other words, they are polylexical terms, such as 'market access', 'market allocation', 'Market and Trading Information System'. Therefore, they are not constructions as envisioned in EcoLexicon. More interesting for our research are constructions in the DTEFC, such as *be on the market* (last example), which focus on verbs.

The most positive aspect of the DFTEC is indeed the great quantity of combinations provided. It is thus a veritable treasure house of information for a translator. The accuracy and reliability of the information in the DTEFC are evidenced by its number of editions, the result of many years of work and revision. However, the fact that it mainly focuses on terms in the form of compound nouns and considerably less on collocations themselves, signifies that users may find it difficult to retrieve the verb or adjective that best fits in a sentence when they are trying to produce a text. In other words, if users consult the dictionary to find out what actions can be performed within the context of the stock market (e.g. 'develop', 'monopolize', 'hold', 'supply', 'play') or to find the right Spanish terms for these actions, the DTEFC is not the resource to consult.

market *n/v:* COMER, FINAN, BOLSA mercado, bolsa, plaza; valor de mercado; comercializar, vender o vender en el mercado, explotar comercialmente, lanzar al mercado, introducir en el mercado; ponerse en venta; ◊ *A market can be a network of buyers and sellers who deal with each other over a computer screen;* V. *be priced out of the market, black market, bull market, Common Market, flea market, free Enterprise/market economy, forward markets, over-the-counter market; ready market, market maker; dampen the market, clobber the market.* [Exp: **market abuse** (MERC FINAN/PROD/DINER utilización abusirva del mercado **market access** (COMER, FINAN acceso a los mercados), **market access commitments** (COMER, FINAN compromiso [en materia] de acceso a los mercados), **market access concessions** (COMER, FINAN concesiones sobre accesos a los mercados), **market access opportunity** (COMER, FINAN oportunidades de acceso al mercado), **market allocation** (COMER, FINAN reparto/distribución de mercados), **market amplitude** (FINAN amplitud del mercado; volumen negociado), **Market and Trading Information System, MANTIS** (FINAN sistema de información de la Bolsa de Londres que permite la ejecución automatic de transacciones; **market area** (COMER zona commercial), **market, at** *US* (BOLSA orden de comprar o vender títulos tan pronto sea recibida al precio del Mercado o el mejor posible también llamada *at market order;* orden ilimitada; V. *at best; in the market; at the close order; day order, good until cancelled, limited order; no-limit order; on the market*), **market basket** (ECO, GRAL bolsa de la compra), **market, be on the** (GRAL estar en venta; V. *come on to the market*) […]

Table 30. Extract from the headword 'market' in the DTEFC

2.3.2 Domain of computing and the Internet

Internet. Répertoire bilingue de combinaisons lexicales spécialisées français anglais

The *Internet. Répertoire bilingue de combinaisons lexicales spécialisées français anglais; Bilingual Inventory of Lexical Combinations English-French* (Meynard 2000) is a dictionary of word combinations in the field of the Internet which provides users with English and French translations of typical phraseological units. This bilingual repository provides a non-exhaustive description of English and French usages structured around a set of keywords or *base nouns*. The dictionary is divided into two parts: English >French and French>English. Two indexes, one in English and one in French, make it possible to find each combination. Table 31 shows the entry for the term 'link' (L'Homme 2009: 241):

LINK definition Hypertext object that allows a word, an image or a data element to connect with another.	
French base noun: lien	
Collocate noun + Base noun • Attributes of a link • Behaviour of a link • Colour of a link • Creation of a link • Deletion of a link • […]	 Attributs d'un lien Comportement d'un lien Couleur d'un lien Création d'un lien Suppression d'un lien
Collocate verb + Base noun • To activate a link • To click on a link • To create on a link • To delete a link • To display a link • […]	 Activer un lien Cliquer sur un lien Créer un lien Supprimer un lien Afficher un lien
Base noun + Collocate verb • Link displays • Link initiates • Link points to • Link specifies • Link works	 Lien affiche Lien lance Lien évoque Lien spécifie Lien fonctionne
Collocate adjective + Base noun • Basic link • Broken link • Direct link • External link • Functional link • […]	 Lien élémentaire Lien rompu Lien direct Lien externe Lien fonctionnel

Table 31. Entry corresponding to 'link' in the *Internet. Répertoire bilingue de combinaisons lexicales spécialisées français anglais*

As can be seen, for each lemma or base noun, a definition is provided as well as the correspondence for the base noun in the other language (in this case the translation into French, 'lien'). The classification of collocations is initially based on their syntactic structures. There are four categories: (i) collocate noun + base noun (e.g. 'attributes of a link'-'attributs d'un lien'); (ii) collocate verb + base noun (e.g. 'to activate a link'-'activer un lien'); (iii) base noun + collocate verb (e.g. 'link displays'-'lien affiche'); (iv) collocate adjective + base noun (e.g. 'basic link'-'lien élémentaire').

This dictionary contains corresponding collocations in the other language, which are invaluable for translators. For this reason, the verb template in this thesis is bilingual English-Spanish. In other words, for each lemma, a comparable description is offered for both languages. Although our aim is not to provide exact correspondences for each collocation within a lemma, the fact that collocations are classified according to meaning makes it possible to establish equivalents in the other language.

Some of the possible drawbacks of Meynard (2000) are the limited number of articles included and the fact that it is only available in paper format. This considerably restricts accessibility and means that information can only be accessed by means of the base term (generally a noun), which means that verbs can only play a very secondary role.

Dictionnaire fondamental de l'informatique et de l'Internet (DiCoInfo)

The *Dictionnaire fondamental de l'informatique et de l'Internet*[58], henceforth *DiCoInfo*, is an online specialized dictionary in French, English, and Spanish that contains more than 1,000 entries in the fields of computing and the Internet. It was created by the research group *ÉCLECTIK (Équipe de recherché en combinatoire lexicale, terminologie et informatique)* of the Meaning-Text Linguistics Observatory *(Observatoire de Linguistique Sens-Texte, OLST)* of the University of Montreal. The proposal mainly draws on the theoretical principles of the Meaning-Text Theory and the Explanatory Combinatorial Dictionary (§2.1.3).

The dictionary contains basic terms related to computer science and the Internet. Therefore, lexical items unrelated to computer science, even if they are frequently used in the field and found in specialized corpora (L'Homme 2009: 244), and those used only in narrow sub-fields of computer science are not included in the resource. The lemmas can be nouns, verbs, adjectives, or adverbs. They can describe objects (e.g. 'file'), human actants (e.g. 'user'), representations (e.g. 'bit'), activities (e.g. 'delete'), properties (e.g. 'compatible'), or units of measurement[59]. The dictionary also contains proper names and acronyms.

The DiCoInfo can be searched alphabetically by using an index, or a search engine. The search engine (see Figure 37) allows the user to perform searches based on the following criteria: (i) *mode*, i.e. by term, word, lexical relation[60] or expression; (ii) *language* (English, French, Spanish or the three languages together); (iii) *precision*, which permits searches by "exact matching", "starting from" or "containing" the word introduced in the search box:

[58] <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi> [25/03/2011].

[59]Information extracted from the DiCoInfo user guide:
<http://olst.ling.umontreal.ca/dicoinfo/DiCoInfo%20manual_Eng.pdf> [23/02/2011].

[60]As shall be seen, collocational information is included in the section for lexical relation. Therefore, in contrast to the other resources described before, DiCoInfo allows direct searches of collocations.

Figure 37. Interface of the DiCoInfo

Table 32 shows the entry corresponding to 'file$_1$'. Each lemma in the dictionary is associated with one or various lexical units for each specific sense. Therefore, each entry in the dictionary corresponds to a specific sense. As can be seen, entries in the DiCoInfo are divided into the following default categories: (1) *headword* (multi-word expressions can also appear as headwords if their meaning is non-compositional); (2) *grammatical information*; (3) *degree of completion* of the entry ("0" means the entry is completed, "1" that the sections are in an advanced stage of editing, and "2" that the entry is still being developed); (4) *actantial structure* which describes the typical actants activated by a verb and their semantic role; (5) *definition*, only for terms whose state is "0"; (6) *synonyms*; (7) *linguistic realizations of actants*, which lists all linguistic realizations of the arguments; (8) *context(s)* that are brief extracts from the corpus texts; (9) *lexical relations*[61], which describe the relations between the headword and the semantically-related lexical units, and is thus the section where collocational information is given; (10) *correspondences* in the other languages; and (11) *administrative information*, such as the last update of the entry and the person responsible for compiling it.

[61] We list lexical relations in a separate table since it is the most complex information category (L'Homme 2009). It is also the most interesting for our purposes because it contains phraseological information.

file$_1$, n Status: 2
Actantial structure: a file: ~ created by Agent{user 1} to act on Patient{data 1}
Linguistic realizations of actants
Context(s)
Lexical relations
español: archivo1 *français:* fichier1
Written by: LPD MEP MCLH Last update:31/01/2009

Table 32. Entry for 'file$_1$' in the DiCoInfo

Table 32 shows that the LU 'file$_1$' is not complete since its status is "2"[62], which is also the reason why the definition is not as yet offered. The actantial structure provides access to the actants, which are labelled with actantial roles. Approximately 15 roles have been defined, of which the most frequent are AGENT, DESTINATION, INSTRUMENT, and PATIENT (L'Homme 2010: 148–149). In the example in Table 32, the actantial structure is "a *file* is created by an AGENT and acts on a PATIENT". The translation into French ('fichier') and into Spanish ('archivo') is also provided, as well as who compiled the entry and when it was last updated (i.e. January 2009). By clicking on *linguistic realizations of actants*, a new window is displayed (see Table 33). As can be observed, the role of AGENT is realized by the *user* and that of PATIENT by *command*, *data*, and *text*. Contexts of use are also provided.

Linguistic realizations of actants
Agents
user$_1$
Patient
command$_1$, data$_1$, text$_1$
Contexts
With all your important files saved on a removable disc such as a floppy or CD, you'll be able to continue your work on another PC.(Source: CALL THE PC DOCTOR) *Here is a quick view of some the files and folders that should be created on our Web server.* (Source: BLOGGING AWAY TO GLORY) *By contrast, dictation software recognizes your speech as incoming words to be inserted into a file as text.*(Source: ABOUT UNIFIED MESSAGES)

Table 33. Linguistic realization of actants within the entry for 'file$_1$' in DiCoInfo

[62] The vast majority of entries in the dictionary have status 2. The reason why we decided to include this entry, despite its degree of incompletion is that, it offers more collocational information.

In addition, by clicking on the section for *lexical relations*, the user can access the paradigmatic and syntagmatic relations of the headword and semantically related lexical units. It is thus the section for collocations. The information is displayed in two columns (see Table 34). In the right column, the semantically-related terms are listed. The left column explains the semantic relations between the headword and the terms on the right. As previously mentioned, this description is based on a simplified version of Mel'čuk's lexical functions (§2.1.3.1). It thus uses everyday-language versions of definitions in the *Dictionnaire de combinatoire* (DiCo) (Polguère 2000) and in the *Lexique actif du français* (Mel'čuk and Polguère 2007).

Relations are classified as follows: (i) *related meaning*, which includes recurrent paradigmatic relations (e.g. 'document', a quasi-synonym of 'file'); (ii) *type_of* relations (e.g. 'pdf file'); (iii) *combinations*, which account for verbal collocations (e.g. the user creates a f. -> 'create a file'); (iv) *others*, which lists less standard paradigmatic relations (e.g. 'file format').

In regards to verbal collocations, such as 'open a file', the dictionary describes three linguistic properties of collocations in accordance with the ECD (§2.1.3.1) (L'Homme 2009: 249-250):

- the syntactic relationship between the base and the collocate, evidenced by the phrase containing the collocate and the headword ('file' is the object complement);
- the actantial structure of the base, shown by the explanation of the meaning of the collocation. Table 33 shows that 'file' has two actants: an AGENT instantiated by terms denoting a *user*, and a PATIENT realized by terms denoting *data*. In the explanation of the combination 'open a file', *the user starts using a file*, the first actant (user) is involved;
- the sense of the collocate in simple language ('use', 'create', 'cause', 'start', etc.). In this case, 'open' means *start using.*

Lexical relations	
	Related term
Related Meanings	
≈	document
Types of	
That does not have a content	empty ~
That has a specific format	HTML ~
That has a specific format	binary ~
That has a specific format	PDF ~
That has a specific format	XML ~
That has specific data	data $_{1}$~
That contains specific data	text ~
That is added to an email	attachment $_{1}$
That is used to carry out a specific task	configuration ~
That is used to carry out a specific task	installation $_{2}$~
That is used to be executed	executable $_{1}$
That is used to obtain help	help $_{1}$~
That contains a series of instructions	batch $_{1}$~
Combinations	
The user creates a f.	create $_{1a}$ ~
The user destroys a f.	delete $_{1a}$ ~
The user reduces the size of a f.	compress $_{1a}$ ~
The user restores the original size of a f.	decompress $_{1a}$ ~
The user creates a f. according to predefined parameters	generate $_{1a}$ ~
The user places a f. on a storage device	backup $_{1a}$ ~
The user places a f. on a storage device	copy $_{1a}$ ~ to ...
The user places a f. on a storage device	save$_{1a}$ ~ to ...
The user prepares a f. to allow the f. to operate	install $_{1a}$ ~
	download a ~
The user places a f. on a computer from another computer	load $_{1ba}$ ~
The user starts using a f.	open $_{1a}$ ~
The user starts using a f.	edit $_{1a}$ ~
The user uses a f.	share $_{1a}$ ~
Tne user uses a f. with other users	parse $_{1a}$ ~
Someone or something uses a f.	process $_{1a}$ ~
Someone or something uses a f.	damage a ~
Someone or something causes that a f. stops functioning properly	
Others	
Format	~ format $_{1}$

Table 34. Lexical relations for ‘file’ in the DiCoInfo

DiCoInfo gives a vast array of collocational information and a detailed description of each word combination. This includes its meaning and syntactic schema. Another advantage of DiCoInfo in comparison to the other resources described is that it permits users to access collocational information in various ways, not only by the headword, but also directly by specific combinations. In addition, the fact that the DiCoInfo is theory-based is also one of its virtues.

Nevertheless, it could be more user-friendly since most people do not know what an actant is. Furthermore, the dictionary does not seem to target any specific user group since parts of it assume too much linguistic knowledge for it to be useful to anyone except a linguist. L'Homme (2010: 142) states that they are currently working on converting the formal encoding into a more accessible description for users who are not familiar with the theoretical principles of the ECD (§2.1.3.1). Another limitation is that the dictionary does not deal with the problem of representing relationships between collocations across languages (L'Homme 2009: 254). In other words, it does not give the correspondence for a given collocation in the rest of languages in the database. This would have been of enormous help to translators.

2.3.3 Domain of the environment

Dictionnaire fondamental de l'environnement (DiCoEnviro)

The *Dictionnaire fondamental de l'environnement*[63], henceforth *DiCoEnviro*, is an incipient online dictionary in French, English, and Spanish related to the field of the environment, and more precisely to the field of the climate change. It has also been developed by ÉCLECTIK, the same research group who compiled the DiCoInfo, and follows exactly the same methodology used in the DiCoInfo. Currently, it only contains 150 entries for the three languages concerned and most of the entries are in state 2, which means they are still incomplete. Table 35 shows the entry for 'erosion', but there is no need to describe it further since both the interface of the dictionary and the information included in terminological entries follow exactly the same structure and methodology developed for the DiCoInfo.

[63] <http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi> [21/03/11].

erosion$_1$,n

Status: 2

Actantial structure: erosion: ~ of Patient{shoreline} by Natural_force {water}

Linguistic realizations of actants

patient
coastal, shoreline, soil
natural_force
water, wind

Context(s)

Increased erosion of Arctic shorelines is expected from a combination of rising sea level, permafrost thaw and increased wave action as a result of increased open water.
(Source:CHANG 7IPCCREGION)
For example, mangroves in low-island coastal regions where sedimentation loads are high and erosion processes are low may be better able to respond to sea-level rise because deposited sediments will create new habitat for mangrove colonization.
(Source:CHANG 2IPCCBIODIVERSITE)
The severity of this drought was such that Lake Karun, in the middle of the Fayyum Depression some 30 kilometres from the Nile, dried up and its sediments were partially carried away by wind erosion.
(Source:CHANG JACLET)

Lexical relations

Actantial roles

Explanation -Typical term	Related term
Other Parts of Speech and Derivatives	
Verb	erode

español: erosión$_1$

français:

érosion$_1$

Written by: ALS MG MCLH
Last update: 14/03/2011

Table 35. Entry of the DiCoEnviro correspondent to the term ‘erosion’

Regarding collocational information, the entry only lists ‘erode’ as a related term. Thus, although the dictionary focuses on phraseological information, based on the limited number of records, collocates only appear in a few entries, and when listed, this information is incomplete.

2.3.4 Multiple domains

Termium Plus®

Termium Plus®[64] is a terminological and linguistic database created by the Government of Canada. It is the result of over 30 years of research and development in Terminology. It contains almost four million English and French terms, more than 200,000 terms in Spanish, and has recently incorporated 18,000 terms[65] in Portuguese. Queries can be formulated in any of the four languages. It is a work in progress, i.e. more than 4,000 monthly updates are performed, which include record creation, deletion of obsolete data, and expansion of existing records.

According to its website, Termium Plus® covers "almost every field of human endeavour […] from a simple tool or a complex machine, to a disease or plant, association or committee". Two terminographic repositories in this resource provide extensive information on terminological combinations, namely, the *Vocabulaire combinatoire de la CFAO mécanique* [Combinatory Vocabulary of CAD/CAM in Mechanical Engineering] (Lainé 1993) and the *Vocabulaire des systèmes dynamiques et de l'imagerie fractale*[66] [Combinatory Vocabulary of Fractal Imagery] (Pavel and Boileau 2003). Both resources, published by the Translation Bureau of the Government of Canada, have been integrated into Termium Plus® (Orliac 2004: 95). Termium® is a tool created by the Canadian government to facilitate standardization. As such, it provides access to 16 electronic resources and provides writing assistance.

Figure 38 displays the search interface of Termium®. As shown, searches are performed by exact matching. The application also permits users to perform either simple or advanced searches by means of the button *search keys*. In the simple search, it is possible to choose terms in English, French, Spanish, Portuguese, or all of them. This simple search limits the search to one language and to records that contain the term being queried in the main entry, synonyms, abbreviations, spelling variants, and key terms. However, the system does not search for the term in the definition, contexts, observations or examples.

Thus, for example, if users are looking up 'data recorder' and they type 'data' in the search box, the system will not give the term 'data recorder' because this term is not

[64] <http://www.termiumplus.gc.ca/site/termium.php?lang=eng&cont=001> [21/03/2011].
[65] This information corresponds to March 2011.
[66] Also available online: < http://www.btb.gc.ca/publications/documents/fractal.pdf>.

an exact match for the search term. If no satisfactory response is achieved, the advance search option can always be used[67].

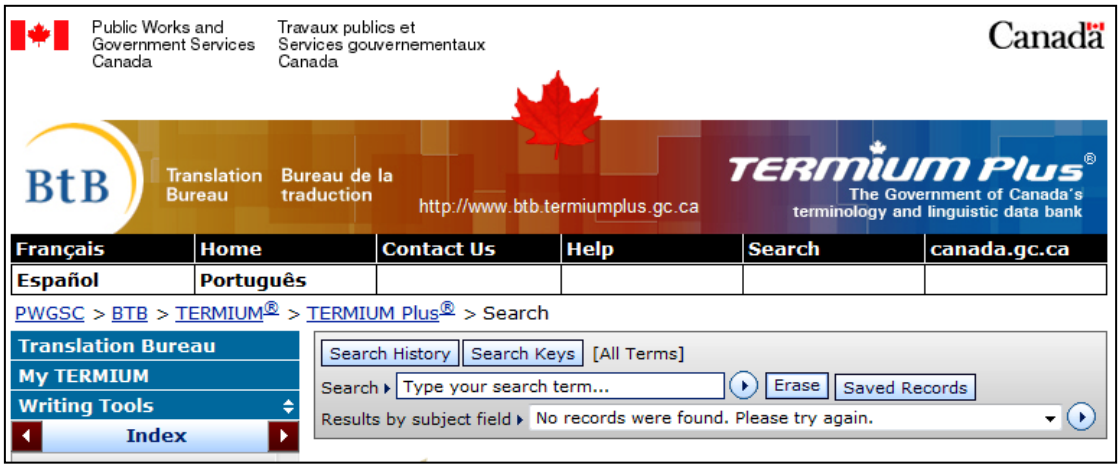


Figure 38. Search interface of Termium Plus®

Table 36 reproduces the results after entering the term, 'data'. The search was restricted to *English terms* in the simple search option since within the option of *all terms*, 200 records were offered. For the English search, five fields are displayed (Communication and Information Management; General Conduct of Military Operations; General Scientific and Technical Vocabulary; ISO/IEC JTC Information Technology Vocabulary; and Information Theory). These fields correspond to that of *Information Theory*. The reason for this choice of term was the fact that the phraseological information contained is one of the most complete in the database.

[67] For more information regarding advance search options, see the section entitled "Understanding search key":<http://www.termiumplus.gc.ca/tpv2alpha/alphaeng.html?lang=eng&srchtxt=&i=1&index=alt&page=srch-clky-eng> [31/03/2011].

Subject Field(s) Information Theory	Domaine(s) Théorie de l'information	Campo(s) temático(s) Teoría de la información
data	données	dato
DEF – Reinterpretable [representations] of information in a formalized manner suitable for communication, interpretation, or processing.	**DEF –** [Représentations réinterprétables] d'une information sous une forme conventionnelle convenant à la communication, à l'interprétation ou au traitement.	**DEF –** Contenido informático que puede expresarse mediante la unidad elemental de un sistema de transmisión de información.
OBS – Data can be processed by humans or by automatic means.	**OBS –** Les données peuvent être traitées par des moyens humains ou automatiques.	**PHR –** Adquisición de datos; mantenimiento de datos; administración de datos; gestión de datos; transferencia de datos; preparación de datos.
OBS – The singular term "datum" is rarely used in the field of information technology. The plural form "data", which is commonly used in this field, is often perceived as a collective noun, which takes a singular verb. However, "data" can take a plural verb if attention is being focused on each individual piece of data (e.g. The data were transferred.).	**OBS –** Bien qu'en anglais le terme singulier «datum» soit rarement utilisé dans le domaine des technologies de l'information, ce n'est pas le cas en français. Certains auteurs, dont l'ISO/CEI, emploient le singulier (donnée) comme équivalent de «data».	**OBS –** dato: Término utilizado generalmente en plural.
OBS – data: term and definition in the singular standardized by ISO/IEC and CSA International.	**OBS –** donnée : terme et définition au singulier normalisés par l'ISO/CEI.	
PHR – To analyze, archive, assemble, collect, compile, convert, edit, process, save, transfer, translate data.	**OBS –** données : terme uniformisé par le Comité de terminologie française du Conseil de doctrine et de tactique de l'Armée de terre.	
PHR – Alphanumeric, analog, binary, digital, dynamic, input, output, static data.	**PHR –** Analyser, archiver, assembler, collecter, compiler, convertir, éditer, enregistrer, mémoriser, modéliser, stocker, traduire, traiter, transférer les données.	
PHR – Data analysis, archiving, collection, compilation, conversion, creation, processing, transmission.	**PHR –** Donnée alphanumérique, analogique, binaire, dynamique, fonctionnelle, informatisée, numérique, paramétrique, séquentielle, statique.	
	PHR – Données de conception, d'entrée, de fabrication, d'impulsion, de sortie.	
	PHR – Accès aux données, analyse, archivage, collecte, compilations, conversion, création, échantillonnage, enregistrement, extraction, gestion, manipulation, mise en forme, stockage, traitement, transmission des données.	
	KEY TERM(S) – donnée	

Table 36. Entry of ‘data’ in Termium® corresponding to the field of Information Theory

Table 36 shows that the subject field is first specified in each entry. It is followed by the headword (data) in the three languages with their respective definitions, headed by DEF. Subsequently, observations (OBS), i.e. more information related to the term entry or contexts of use are then offered. Lastly, collocational information (PHR) is given. Collocations are classified in terms of part of speech though it is not made explicit and the user must infer it. In this regard, there are three fields: one for verbal collocates; one for adjectives; and one for nouns.

For example, verbal collocations in English for 'data' are to 'analyze', 'archive', 'assemble', 'collect', 'compile', 'convert', 'edit', 'process', 'save', 'transfer', 'translate'; adjective collocations include 'alphanumeric', 'analog', 'binary', 'digital', 'dynamic', 'input', 'output', 'static'; and noun collocations include 'data analysis', 'archiving', 'collection', 'compilation', 'conversion', 'creation', 'processing', 'transmission'. It is worth mentioning that the phraseological information displayed for the three languages are not translation correspondences.

Termium® is a huge database, which is a gold mine of information. However, it has some limitations in relation to collocations. First of all, in terms of access, the search options in Termium® only permit users to search for exact matches. Therefore, if users only enter the first element of a compound, there will be no hits. A second problem is the excessive number of hits for certain headwords. In addition, although Termium® claims to include phraseological information, collocates only appear in a limited number of term records, and when they are listed, they are incomplete.

For example, when looking up common terms pertaining to our EXTREME EVENT frame, such as 'hurricane' or 'tropical cyclone', no phraseological information is offered. Nor is the grammatical category of the collocate made explicit. In the words of L'Homme (2009: 240), "this method assumes that a user already knows the part of speech of the collocate he or she is looking for and will access the right part of the record", which unfortunately is not always the case.

2.4 Summary

Thanks to the availability of large corpora and lexical analysis tools, it is increasingly frequent for general language dictionaries of all types to include phraseological information in their entries. Nevertheless, at the same time, there is still no systematic treatment of such information. The decision to include a collocation as an example in

the section for phrases or by cross-references to compounds all too often seems to be entirely random. The same inconsistency can be observed in bilingual dictionaries, where phraseological information is given in the form of examples, either in a section for idioms or in cross-references to compounds. The problem in bilingual dictionaries lies in the fact that it is often not possible to differentiate between a fixed collocation and a free example.

Notwithstanding, there is growing evidence of the importance of phraseological information in lexicographic resources. This is evident in the publication in recent years of a wide variety of combinatory or collocation dictionaries. More specifically, the BBI, REDES, PRÁCTICO, and DICE are based on the underlying theoretical principles of the MTT. In contrast, the OCD follows a learner's dictionary approach rather than the premises of any linguistic theory. The same can also be said of the MCD. All of these dictionaries, except the BBI, are corpus-based. In our opinion, any lexicographic or terminographic resource should be theoretically based since this is a guarantee of internal coherence and systematicity.

As shall be seen, our proposal incorporates insights from the theoretical principles of the Lexical Grammar Model (§3.1.2.1), Frame Semantics (§3.1.3.2), and Role and Reference Grammar (§3.1.2.2). Although examples can be created by lexicographers, the use of corpora also enhances credibility.

With the exception of the DICE and the MCD, the dictionaries analyzed only provide one way of accessing collocations. This is done either by the base or the collocate itself. In contrast, the DICE allows users to access collocations in various ways. The search engine allows users to enter one of the following: (i) base term; (ii), collocate; (iii) lexical function. A very positive aspect of the MCD is that it is the only dictionary that includes noun and adjectives in verb entries. Consequently, users can access collocations either by the base term or the collocate.

Regarding the classification of collocations within an entry, only the short entries within REDES, do not offer any type of classification, whereas the other resources do. All of the dictionaries, with the exception of REDES and PRÁCTICO, classify collocations by syntactic schema though the organization and formulation of this information is far from homogeneous. The BBI systematically classifies collocations within the microstructure of each entry, but does not explicitly specify this organization. However, the order followed is described in the introduction of the dictionary: verb +

noun (CA collocations); verb + noun (EN collocations); adj + noun; noun + verb; and noun + noun.

In contrast, the OCD explicitly specifies the grammatical category and order of the two components (i.e., for the lemma *bed*, adj, verb + *bed*; *bed* + verb, prep, phrases). The MCD also provides this information but without indicating the headword. This means that instead of verb + *bed*, it says verb + noun. The order followed is adj + n; v + *into* + n; v + *to* + n. The DICE first includes attribute combinations, followed by n + adjective; verb + noun, noun + verb, and noun *of* + noun.

PRÁCTICO also classifies collocations by part of speech. However, the syntactic schema between the noun and verb is not expressed. The only resource that gives a meaning-based classification of collocation is REDES in its long entries, which are classified in lexical classes. It is true that the OCD and MCD claim to differentiate meaning within the same grammatical category. However, their attempt is somewhat less than successful since the relationship between the various components in the group is not expressed, and there are inconsistencies in the members of each group.

Concerning the description of collocations, the vast majority of resources provide explanations of the meaning of collocations. When necessary, usage notes and usage examples are often provided. However, the BBI is the only resource that differentiates American English from British English, whereas REDES is the only one that indicates the frequency of each combination. Finally the DICE has the advantage of offering a gloss for each combination

The inclusion of collocations in specialized resources is also becoming more and more frequent. However, specialized dictionaries with collocations differ in respect to how they are listed and represented in entries. As such, there is still no general agreement as to which kinds of word combination should be included in dictionaries and how they should be described and classified.

In specialized dictionaries (with the exception of DiCoInfo and DiCoEnviro), collocations can only be accessed by the headword. In other words, if the user is searching for ‘monopolize a market’, it can only be accessed on the basis of the entries for ‘monopolize’ and ‘market’. Apart from allowing access through the headword, DiCoInfo and DiCoEnviro, permit searches directly through the collocation by means of the option *lexical relation*. In consonance with DiCoInfo and DiCoEnviro, an effective terminographic dictionary will contain a wide variety of collocational information and

favor its accessibility by providing different ways of accessing the information depending on user needs (Bergenholtz and 2004, 2010).

Regarding the classification of collocations within an entry, the DTEFC, the DICOFE, and the DA do not offer any type of classification, but just display terms alphabetically. In contrast, the other terminographic resources do provide some kind of organization. Cohen (1986), Caignon (2001), the DAFA, and TERMIUM classify entries according to their grammatical categories. Meynard (2000) and the DAFA also specify the syntactic schema of the base and collocate.

For example, Meynard (2000) describes phraseological units in the following groups: (i) collocate noun + base noun; (ii) collocate verb + base noun; (iii) base noun + collocate verb; (iv) collocate adj. + base noun. The translation into either French or English is also given.

In contrast, the DAFA offers a syntactic classification for verbs according to their actantial structure or semantic roles. Finally, meaning is of utmost importance for the DiCoInfo, DiCoEnviro, DAFA, and Cohen (1986), in that these resources classify collocations according to their sense, as formulated by the lexical functions of Explanatory and Combinatorial Lexicology (ECD) (Mel'čuk et al. 1995; Mel'čuk 1984-1999) (§2.1.3.1). Cohen (1986) adapts the ECD representation and uses paraphrases of a certain number of lexical functions such as *début* [start], *croissance* [growth] and *fin* [end].

Finally, regarding the description of collocations, there are resources such as Cohen (1986) and Caignon (2001) that offer no description at all. Translations of the collocations are provided in Meynard (2000), DTEFC, DICOFE, Termium®, and the DAs. Of the dictionaries that include translations, the DTEFC is the only one that also offers a definition or explanation of almost every collocation. In line with the DTEFC, other resources with detailed definitions for collocations, although they do not contain translations, are the DiCoInfo and DiCoEnviro. Usage examples are also given in a few resources, such as the DTEFC, DAFA, and DA. Finally, synonym and antonym collocations are included in the DAFA and DTEFC.

In our opinion, in order to assure some level of systematicity and coherence in any lexicographic or terminographic resource, a set of criteria should be established beforehand, regarding the following: (i) the types of collocation to be encoded; (ii) the kinds of collocational information offered; (iii) the place for collocations in the micro or macrostructure of the dictionary.

In this regard, the criteria adopted for the description of verbal phrases in EcoLexicon are explained in §5. From the analysis of the various lexicographic and terminographic resources in this section, it has also been inferred that the presentation and arrangement of the information within a lemma is of great importance both for easy access to data as well as for knowledge acquisition on the part of the user. The following guidelines can thus be specified for an ideal phraseological entry in a specialized dictionary:

1. The resource should be available online or in electronic format;
2. It should be bilingual or multilingual, and include correspondences between the phraseological units in the various languages;
3. It should be theoretically based;
4. It should be communicatively and cognitively-based. This means that it should be conceived for both encoding and decoding;
5. The interface and display of the information should not be encoded in a difficult metalanguage;
6. It should allow various ways of searching and accessing collocations, which are suitable to the user and type of situation (Bergenholtz and Tarp 2004, 2010);
7. It should not contain an excessive number of collocations in an entry;
8. It should provide a semantic classification of collocations, based on meaning;
9. It should give a detailed description of the semantic and syntactic patterns associated with the meaning of each collocation.
10. It should provide various usage examples showing the word in context.

Our study focuses on combinations formed by verb + noun or noun + verb within EcoLexicon, which we have referred to as *collocations* in a broad sense. Therefore, in the analysis in this thesis, we concentrate on those combinations of noun + verb or verb + noun, in which the noun is referred to as the *base* and the verb as the *collocate*. These elements frequently occur in combination with each other and in the majority of cases, conserve their meaning. In an EXTREME EVENT frame, the collocate (the verb) is constrained by the meaning of the base (the noun), but at the same time, the collocate constrains the kind of nouns that can combine with it.

3. Theories of Argument Structure

> A theory frames how we view a subject and how we expect to research it. [...] A given theoretical frame will allow us, or occasionally force us, to see certain aspects of a discipline in a certain way while preventing us from seeing other aspects that an alternative theory might naturally bring into view. In this regard, theorists are like tourguides who point out and comment on different sights as the tour travels along its route; one theorist's itinerary and commentary may bear only a passing resemblance to another's (Danaher 2007: 89).

3.1 Theories of argument structure

In the first half of the 20th century, most linguistic theories envisaged combinatorial potential or the subcategorization of verbs primarily from a syntactic perspective. Both Structuralism and Generative Grammar agreed that the study of sentence and word meaning should be postponed, and that sentences should first be analyzed as purely syntactic structures. However, Mairal and Faber (2005: 282) write:

> Linguistic theories that endeavour to account for syntactic structures and leave meaning for later inevitably encounter stumbling blocks because language is not so much about grammatical constructions as about meaning. [...] One might even go so far as to say that the syntax of lexical units would depend on their meaning instead of vice versa since meaning is prior to syntax.

Nevertheless, the importance and impact of Structuralism and Generative Grammar is unquestionable. Both theories helped to systematize linguistic phenomena by means of (i) the decomposition of syntagmas into smaller units, which allowed their comparison; (ii) the formulation of rules that predicted grammar[68].

Little by little, semantics acquired a more important role within linguistics to the extent that today almost all current linguistics theories share the idea that there is a direct relation between syntax, semantics, and pragmatics. In other words, syntax is no longer conceived as separate from meaning and context. On the contrary, sentence structure is usually studied by taking into account the meaning of the words in the sentence as well as the situational context. In the words of Mairal and Faber (2005: 302):

[68] However, the acknowledgement of the close relation between syntax and semantics did not come until somewhat later.

> One of the greatest challenges in linguistics today is to find a kind of representation that will adequately transmit the interface between syntax and semantics. Although syntax is undoubtedly important, it hardly tells the whole story (or even the main part of the story). No linguistic framework can aspire to any sort of adequacy unless it takes a position on the conceptual meaning of lexical units and endeavors to show how this type of meaning is related to their syntactic potential.

With this scenario as a backdrop, various linguistic approaches have emerged that study argument structure. Generally speaking, these approaches can be divided into three categories: (i) formal linguistic approaches; (ii) functional linguistic approaches; (iii) functional-cognitive linguistic approaches. In the following sections, we describe the basic theoretical premises of each type of approach, as well as the main theories that belong to each paradigm.

3.1.1 Formal theories

Formal Linguistics refers to theoretical models that describe languages and explain language acquisition in terms of formalisms based on mathematical logic. The most well-known of these approaches is Chomsky's Generative Grammar. Generative Grammar began with *Syntactic Structures* (Chomsky 1957), and was initially proposed as an alternative to European and American Structural Linguistics, which dominated the first half of the 20th century. However, from the sixties onwards, Generative Grammar became the leading linguistic paradigm.

Structural Linguistics was based on the premise that language is a social institution. In the words of Blackburn (1996: 365):

> The common feature of structuralist positions is the belief that phenomena of human life are not intelligible except through their interrelations. These relations constitute a structure, and behind local variations in the surface phenomena there are constant laws of abstract structure.

In contrast, Generative Grammar conceived language as a faculty which is part of the mind and should be learnt by means of methods and techniques similar to those used in natural sciences, such as physics or biology (Mairal et al. 2012: 54). According to Chomsky ([1964] 1970: 28–56), any linguistic theory or grammatical description should meet three levels of adequacy: (i) observational; (ii) descriptive; (iii) explanatory. A

theory or description achieves *observational adequacy* if it is able to describe which sentences in a language are grammatically well formed; it achieves *descriptive adequacy* if it describes when sentences are well formed and also specifies the grammatical properties that make sentence well formed; it achieves *explanatory adequacy* when descriptive adequacy is reached through the use of principles that can apply to any language. This is the foundation of what has been referred to as *Universal Grammar.* As Meyer (2002: 3) underlines, Generative Grammar prioritizes explanatory adequacy over the other types of adequacy, and thus relegates the representativeness of data and variation to the background:

> Because generative grammar has placed so much emphasis on universal grammar, explanatory adequacy has always been a high priority in generative grammar, often at the expense of descriptive adequacy: there has never been much emphasis in generative grammar in ensuring that the data upon which analyses are based are representative of the language being discussed, and with the notion of the ideal speaker/hearer firmly entrenched in generative grammar, there has been little concern for variation in a language.

This is the reason why Generative Grammar is contrary to the use of corpora for language study. As is well known, corpus linguistics conceives variation as inherent to language, giving priority to descriptive adequacy over explanatory adequacy. As such, corpus linguistics is not in harmony with formal approaches since formal studies do not take usage into account. In the words of Chafe (1994: 21), formal approaches "exclude observations" and rely almost exclusively on "notational devices designed to account for only those aspects of reality that fall within their purview".

Chomsky's approach is referred to as *generative* in the sense that he believes that a finite set of rules *generate* the set of potentially infinite number of sentences of a particular language. In other words, his assertion is that there are a finite number of sounds in each language, but native speakers have a certain intuition of the structure of the language that allows them to understand and transmit an infinite number of sentences (Chomsky 1988: 14).

As such, the basic premises of the initial versions of Generative Grammar were based on the assumption that grammatical mechanisms were regulated by mathematical properties. Grammars were conceived as *theoretical machines* capable of generating and interpreting the expressions of a language in the same way as speakers do (Cabré and Lorente 2005: 435). The metalanguage used by grammars was taken from formal

logic and modern mathematics, and was commonly based on symbols and operations, specified within each grammatical model. This formalization was obtained by means of rules, parameters, constraints, logical functions, etc. The ultimate goal of this process was the establishment of a *universal grammar*, namely, to develop a theory that would explain how the human brain processes language:

> The grammar of a particular language, then, is to be supplemented by a universal grammar that accommodates the creative aspect of language use and expresses the deep-seated regularities which, being universal, are omitted from the grammar itself. Therefore it is quite proper for a grammar to discuss only exceptions and irregularities in any detail. It is only when supplemented by a universal grammar that the grammar of a language provides a full account of the speaker-hearer's competence (Chomsky 1965: 6).

The computational metaphor is at the foundation of all formal grammar approaches (Cabré and Lorente 2005: 435). This means that the same general language subjected to different instructions can generate different outputs. In this sense, language receives different external stimuli, which makes language use possible in the form of a specific language. Generally speaking, the objective of grammar is to adequately describe the expressions generated in each language and at the same time, highlight the nature of a universal grammar and the operations that allow the derivation of specific languages. Apart from Generative Grammar, the Lexicon-Grammar approach (§3.1.1.1), Lexical Syntax (§3.1.1.2), as well as the Linking Rules approach (§3.1.1.3) are currently the most influential formal approaches.

3.1.1.1 Lexicon Grammar[69]

The Lexicon-Grammar is based on the transformational and formal model proposed by Harris (1968, 1970) and afterwards developed by Gross (1968, 1975, 1977, 1986) for French within the *Laboratoire d'Automatique Documentaire et Linguistique (LADL)*, which was later integrated into the *Laboratoire d'Informatique Gaspard-Monge (LIGM)* at the Université Paris-Est Marne-la-Vallée[70]. In contrast to Generative Grammar, this method regards grammar and lexicon as being intimately linked without favoring one or the other. In addition, it does not support the generation of rules since it states that two

[69] The description of this model is a summary of the information provided at the following page: <http://igm.univ-mlv.fr/LIGM> [10/01/2012].
[70] <http://igm.univ-mlv.fr/LIGM/> [10/01/2012].

sentences always have more properties that are different than those that are shared. Even though Gross accepted the importance of introspection in the same way as Generativists, he also pioneered the use of large corpora. Gross's Lexicon Grammar had a major impact on Europe for more than 25 years. The method that he used was extended to a large number of languages and established the theoretical bases of a system of dictionaries and electronic grammars that are currently among the most valued lexical resources for natural language processing[71].

The Lexicon-Grammar method consists of a syntactic analysis and a lexical entry to which syntactic analysis is applied. The result is a lexicon grammar of verbs and other predicate elements in the form of property tables (see Table 37). Each table groups a number of entries judged to be similar in that they share a set of properties. Therefore, each table corresponds to a class that groups the lexical elements of a given grammatical category (verb, adjective, noun, or adverb), which share the same properties referred to as *propriétés définitoires* [defining properties]. Within each table, there is, at least, a *construction de base* [underlying construction].

In this sense, each table contains a significant number of entries and is the result of the application of a set of common linguistic principles: (i) the different senses are carefully distinguished according to transformational and distributional properties; (ii) the *phrase élémentaire* [sentence] is the minimal unit of study; (iii) the sentence is reduced to a predicate and a set of arguments, activated in the form of complements. More specifically there must be principles that allow the separation of essential complements (subjects and objects) from non-essential ones, i.e. adverbial or circumstantial complements. Finally, a syntactic transformation set enables the description of structural variations. Table 37 shows an example taken from the French Lexicon Grammar, which corresponds to *Locatif à corrélat Humain* [locative correlated to human] (Guillet and Leclère 1992: 123):

[71] We will concentrate on the description of the Lexicon Grammar. For more information concerning the rest of resources that the *Equipe Informatique Linguistique* is in the process of developing, see <http://infolingu.univ-mlv.fr/> [30/01/2012].

TABLE 38LH									
No source No destination N1 V N2 V N1		Pfx nég/source Pfx nég/ nv dest	N1=: V-n N0 V N2 (de N1)	Source/destination Prép: = de Autre prep source Prép=: dans Prép=: sur Prép=: contre Prép=: à Prép=: vers	N2 =: V-n Ppv=: y PPv=: en N1 est Vpp N1=: N-hum concret	mot Loc texte idée Loc esprit Nhum Loc Nabs N1=: Qu P	
- - - -	Immiscer	- -	- -	- - - + - - - -	- + - + +	+ + + -	*Max ~ sa soeur dans les affaires de Luc*
- - - -	Impliquer	- -	- -	- - - + - - - -	- + - + -	- - + -	*Max ~ Luc dans un scandale*
- - - +	incarcérer	- -	- -	- - - + - - + -	- + - + -	- - + -	*On ~ Max à la prison de Dax*
- - - +	incorporer	+ -	- -	- - - + - - + -	+ + - + -	- + - -	*On ~ Max dans la marine*
- - - -	infiltrer	- -	- +	- - - + - - - -	- + - + -	+ + - -	*Max ~ un agent dans ce réseau*
- - - -	inhumer	- -	- -	- - - + - - - -	+ + - + -	- + - -	*On ~ Max dans le cimetière*
- - - -	inscrire	- -	- -	- - - + - - + -	- + - + -	- - - -	*Max ~ Ida dans un club de yoga*
- - - -	interner	- -	- -	- - - + - - - -	+ + - + -	- - - -	*On ~ Max dans un asile*
- - - -	introduire	- -	- -	- - - + - - - -	- + - + +	- - - -	*Cette letter ~ Léa auprès de Max*
- - - -	introduire	- -	- -	- - - + - - - -	- + - - -	- - - -	*Le valet ~ Bob dans le boudoir*
- - - -	jeter	- -	- -	- + - - - - - -	- - + + -	- - + -	*Ce malheur ~ Max dans le désespoir*
- - - -	jeter	- -	- -	- - - + - - - -	- + - - -	- - - -	*Le patron ~ Max de son boulot*
- - - -	lever	- -	- -	- + - - - - - -	- - - + +	- - + -	*Max ~ Léa de son lit*
- - - -	libérer	- -	- -	- + - - - - - -	- - + + -	- - + -	*On ~ Max de sa prison*
- - - -	licencier	- -	- -	- + - - - - - -	- - + + -	- - - -	*Cette enterprise ~ 1000 ouvriers*
- - - -	limoger	- -	- -	- + - - - - - -	- - + + -	- - - -	*On ~ Max de son poste*
- + + +	loger	+ -	- -	- - - + + + + -	+ + - + -	+ + - -	*Max loge chez lui des amis*
- - - -	lourder	- -	- -	- + - - - - - -	- - + + -	+ + + -	*On ~ Max de son poste*
- - - -	mander	- -	- -	- - - - - - + -	- + - - -	- - - -	*César ~ Caïus chez lui*
- - - -	masser	- -	- -	- - - + + + + -	- + - + -	- - - -	*Le spectacle ~ les gens sur la place*
- - - -	mener	- -	- -	+ + - + - - + +	- + + - -	- - - -	*Max ~ Luc chez lui*
- - - -	mobiliser	+ +	- -	- - - + - - - -	- + - + -	- - - -	*On ~ Max dans la marine*
- - - +	murer	+ -	- -	- - - + - - - -	+ + - + +	- + + -	*L'éboulement ~ Max dans la grotte*
- - - -	muter	- -	- -	+ + - + - - - -	- + + + -	- - - -	*On ~ Max à Gap*
- - - -	nommer	- -	- -	- - - + - - + -	- + - + -	- - - -	*On ~ Luc à la présidence*
- - - -	noyer	- -	- -	- - - + - - - -	- + - + -	+ + + -	*Max ~ les chatons dans la rivière*
- - - -	parachuter	- -	+ -	- - - + + - + -	- + - - -	+ + - -	*On ~ Max dans cette enterprise*
- - + +	parquer	+ -	- -	- - - + - - - -	+ + - + +	+ + + -	*Max ~ les boeufs dans l'enclos*
- - - -	pelotonner	- -	- -	- - - + + - - -	- + - + -	- - - -	*Ida ~ sa grande taille sur le divan*
- - + -	pendre	+ -	- -	- - - - - - + -	- + - + -	- - + -	*On ~ Max au gibet*

Table 37. Extract of Table 38LH of the Lexicon Grammar

As stated in Guillet and Leclère (1992: 123), the underlying construction within Table 38LH is *N_0 V N_1 Loc N_2*, where N_1 =: Nhum compulsory (thus, N1 =: N-hum is coded −) and where N2 is understood as a place, source, and/or destination. In this sense, sentences such as, 'On a viré Max de son poste' [Max has been fired from his job] is taken into account in this table.

The three basic components of Gross's Lexicon-Grammar are the following: (i) *phrases libres* [free phrases]; (ii) *phrases à verbs support de noms* [support verb

phrases]; and (iii) *phrases figées* [fixed phrases] and *adverbs figés* [fixed adverb phrases][72].

Free phrases were the first ones to be introduced. To date, the French Lexicon Grammar contains about 15,000 free phrases built on distributional verbs. Within these phrases, the verb has the function of predicate and selects a subject and eventually essential complements. Both the subject and the complements have a free distribution, which means that their slots can be filled by a wide range of nouns. These free combinations are described in French in approximately 60 tables. Example (3) lists the free phrases for the lexical entries of the verb 'apprendre' [to learn]:

(3)
Max a appris à Luc qu'il avait réussi (Table 9)
Max a appris (par + de) Luc qu'il avait réussi (Table 10)
Max a appris à Luc à lire (Table 16)
Max apprend ce métier (Table 7) = Max fait l'apprentissage de ce métier (Table DR2)

The second basic component are support verb phrases. The French Lexicon Grammar contains about 50,000 types of *phrases à verbs support de noms* [support verb phrases] described in tables. It should be highlighted that one of Gross's most important contribution was the popularization of the notion of *support verb*. In the course of his work, he realised that sometimes the description of only one element has no meaning unless one takes into account the whole sequence. In the phrases containing a support verb, it is not the verb which has the predicate function, but rather a predicate noun, adjective or adverb (see also Subirats 2001). For example, in 'carry out a plan', the noun 'plan' is the predicative element and the one which imposes selection restrictions, whereas the 'carry out' is the support verb which does not have meaning. The distribution of the subject and eventually essential complements depend on this element.

Apart from combinations with a support verb, there are other cases, such as 'perdre la tête' [go mad], which must be considered as one unit. If 'perdre' and 'tête' were regarded as distintic lexical entries, the sum of their individual meanings would not give the sense of 'perdre la tête' or the distributional and syntactic constraints that such a meaning imposes. These types of complex units are referred to as *phrases figées*

[72] All the examples mentioned regarding the components of Gross's Lexicon Grammar have been taken from the website of *Équipe Informatique Linguistique*: <http://infolingu.univ-mlv.fr/> [11/01/2012].

[fixed verb phrases] or *adverbs figés* [fixed adverbial phrases] in Gross's Lexicon Grammar.

The French Lexicon Grammar contains about 25,000 fixed phrases and about 7,000 fixed adverbial phrases. What is noteworthy is that the systematic study of fixed phrases reflects that they are much more frequent than free combinations (44,000 fixed phrases listed). It is precisely on this category referred to as fixed verb phrases, on the type of constructions this thesis focuses. In fixed phrases, the verb is distributionally collocated with at least the subject or one or several complements. It should also be highlighted that a fixed phrase can be associated with one or more free elements. In other words, these slots can be filled by a wide range of nouns. In French, fixed verb phrases are described in about 50 different tables, and adverb phrases in about 15 tables (Gross 1986). In (4), we list some examples of fixed verb phrases for 'apprendre':

(4)
N_0 apprendre N_1 sur le tas (Table PDETC)
N_0 en apprendre de belles sur N_1 à N_2 (Table CPPN)
N_0 en apprendre un bout sur N_1 à N_2 (Table CPPN)
N_0 en apprendre de drôles sur N_1 à N_2 (Table CPPN)

The Lexicon Grammar method was first applied to Spanish by Subirats (1984, 1987), who classified 4,500 verbs extracted from the *Diccionario de la Real Academia Española* (1979) depending on their argument structure. He established 15 verb classes that could potentially appear with a complement and/or infinitive in Spanish. The full description of his ideas is reflected in his *Sintaxis Léxica* [Lexical Syntax] (§3.1.1.2).

As previously mentioned in the beginning of this section, Gross's Lexicon Grammar is still considered an extremely valid approach to grammar since it provides a very rich syntactic description of the language. However, as highlighted by Tolone, Voyatzi, and Leclère (2010: 321), it has certain limitations regarding its applicability to Natural Language Processing (NLP):

> Lexicon-Grammar tables are a very rich syntactic lexicon for the French language. This linguistic database is nevertheless not directly suitable for use by computer programs, as it is incomplete and lacks consistency. Tables are defined on the basis of features which are not explicitly recorded in the lexicon. These features are only described in literature.

The verb description proposed in this thesis is not based on this approach since our research focuses on semantics and somewhat less on grammar. In this regard, our main goal is to describe verb meaning and combinatorial potential without focusing on their computational representation in NLP. It is true that our ultimate objective is to systematize verb description to further enrich EcoLexicon. However, in this initial stage, this thesis provides a complete linguistic description of the verbs most frequently used in this specialized knowledge frame.

3.1.1.2 Lexical Syntax *(Sintaxis Léxica)*

Lexical Syntax (Subirats 2001) is a linguistic model that studies syntax, based on the relational properties of predicates. According to Subirats (2001: 95-96), "[...] lo que intentamos determinar realmente son *clases de dependencia* y no tipologías de locuciones basadas en clases distribucionales de palabras"[73].

One of the main contributions of Lexical Syntax is the extension of the notion of *predicate* to include both verbal predicates and non-verbal predicates, such as predicate nouns, adjectives, and prepositional phrases. This extension of the notion of predicate is not exclusive of Subirats, but has also been proposed by other linguistic theories such as Dik's Functional Grammar (1989: 161-182), and Gross's Lexicon Grammar (§3.1.1.1). As such, in example (5), 'El presidente hizo un llamamiento a la población civil' (Subirats 2006), the predicate is not the verb 'hacer', but rather 'llamamiento' because from a syntactic point of view, this noun determines the projection of the arguments. From a semantic perspective, 'llamamiento' is the semantically-rich element, whereas 'hacer' is semantically empty. Therefore, 'hacer' can be omitted in the formation of a noun phrase whose head would be 'llamamiento' without any information loss, i.e. 'El llamamiento del presidente a la población civil'.

Apart from extending the notion of predicate, another major contribution of Subirat's Lexical Syntax is the analysis of transformational relations between different classes of predicates. In this sense, starting from relations of dependence and from predicate selection, there are basic sentences from which other sentences can be derived by means of transformations, which do not alter the information in the sentence. As such, dependence hierarchies are projected from basic sentences as well as alternative syntactic realizations of predicate hierarchies of these sentences. These constructions

[73] [what we try to determine are dependence classes and not typologies of locutions based on distributional classes of words].

are semantically paraphrastic in that they transmit the same information as their respective basic sentences. For example, in the sentence, 'Los políticos tuvieron una influencia en la decisión del defensor del pueblo' (Subirats 2006), the predicate noun 'influencia' is not regarded as an autonomous predicate, but rather as derived from the verb 'influir' by means of a nominal deverbal transformation.

Montero (2008) and Montero and Buendía (2010, 2012) also describe this type of transformational analysis. For example, phrases such as 'river bank erosion' can be syntactically explained as a process of predicate nominalization, where the head, 'erosion', is derived from the verb 'erode' and takes 'river bank' as its second argument. The noun phrase, 'river bank', can be regarded as the result of a process of predicate deletion, where the head, 'bank', and the non-head, 'river', have a semantic relation that must be inferred from lexical and contextual information (Montero 2008: 1018).

Because of transformations and dependency relations, the derivatives of a predicate will inherit its same properties. This is also the underlying premise of the Functional-Lexematic Model (and by extension, the Lexical Grammar Model) (§3.1.2.1) and the Lexical Constructional Model (§3.1.3.4). In the realm of specialized language, it has also been applied, among others, by Montero (2002) in her study of phrasemes in the domain of oncology. The use of these transformations in the study of syntax has replaced distributional methodologies. This premise was of great value in our research since it was thus not necessary to describe and formalize all the possible combinations of verbs. Instead, these combinations are defined as the projections of predicate hierarchies formalized in conceptual categories.

3.1.1.3 Linking rules approach

Another approach to the study of verbs from a syntactic-semantic perspective can be found in Levin (1993), Levin and Rappaport (1995), and Rappaport and Levin (1998). The basic hypothesis was that the syntactic properties of verbs are a direct reflection of the underlying meaning:

> The most striking illustration of the role of meaning in the determination of syntax is the tendency for arguments bearing certain semantic roles to be associated with particular syntactic expressions (Levin and Rappaport 1995: 1).

Levin (1993) initially proposed a classification of English verbs on the basis of both intuitive semantic grouping and syntactic alternations. She claims that verbs sharing the same meaning also coincide in their grammatical properties. This assumption makes it possible to predict the behavior of a verb, based on its semantic class.

Levin's taxonomy provides a classification of 3,024 verbs (4,186 senses) into 48 classes and 192 subclasses according to their participation in 79 alternations involving NP and PP complements. Table 38 shows the constructions for VERBS OF CONTACT BY IMPACT, as an example:

1.With/Against Alternation
Paula hit the stick against/on the fence.↔*Paula hit the fence with the stick.*
2. Conative Alternation
Paula hit the fence (with the stick). ↔ *Paula hit at the fence (with the stick).*
3. Body-Part Possessor Ascension Alternation
Paula hit Deirdre on the back. ↔ *Paula hit Deirdre's back.*
4. Together Reciprocal Alternation (transitive)
Paula hit one stick against another. ↔ *Paula hit the sticks together.*
5. Instrument Subject Alternation
Paula hit the fence with the stick. ↔ *The stick hit the fence.*
6. *Through/With Alternation
**Paula hit the stick through/into the fence.* ↔ *Paula hit the fence with the stick.*
7. *Simple Reciprocal Alternation (transitive)
Paula hit one stick against another. ↔*Paula hit the sticks.
8. *Causative Alternations
*Paula hit the fence (with a stick). *The fence hit (with a stick).*
9. *Middle Alternation
Paula hit the fence.↔**The fence hits easily.*
10. Unintentional interpretation available (some verbs)
Reflexive Object: *Paula hit herself on the doorknob.*
Body-Part Object: *Paula hit her elbow on the doorknob.*
11. Resultative Phrase
Paula hit/kicked the door open, Paula banged the window shut.
12. Zero-related Nominal
*a hit/*give a hit/*get a hit in the shoulders, a kick/give a kick/get a kick in the shins.*

Table 38. Levin's verbs of CONTACT BY IMPACT

Accordingly, for example, Levin's class of VERBS OF CONTACT BY IMPACT (class 18.1, headed by the verb 'hit') referring to verbs which describe moving one entity in order to bring it into contact with another entity, and not entailing the contact any effect on the second entity (Levin 1993: 148-150), includes the following verbs: 'bang', 'bash', 'batter', 'beat', 'bump', 'butt', 'dash', 'drummer', 'hit', 'kick', 'knock', 'lash', 'pound', 'rap', 'slap', 'smack', 'smash', 'strike', 'tamp', 'tap', 'thump', 'thwack', 'whack'. The examples in Table 38 show the constructions for VERBS OF CONTACT BY IMPACT classified according to their participation (1-5) or non-participation (6-9) in the following alternations and other constructions (10-12).

In the course of her work, Levin evolved towards a lexical semantic representation based on the formulation of rules that determine argument structure as well as the syntactic projection of these arguments (Levin and Rappaport 1995; Rappaport and Levin 1998). According to the authors, verbs can combine with a wide range of complements, as can be observed in 'sweep' (5) (Rappaport and Levin 1998: 97-98):

(5)
a. Terry swept.
b. Terry swept the floor.
c. Terry swept the crumbs into the corner.
d. Terry swept the leaves off the sidewall.
e. Terry swept the floor clean.
f. Terry swept the leaves into a pile.

The variation observed in (5) not only concerns the number and type of complements, but also the possible combinations of these complements. Accordingly, 'sweep' can simply take a direct object (5b) when the corresponding noun is a surface ('floor'). However, 'crumbs' is only possible as the direct object of 'sweep' when followed by a prepositional phrase referring to transference to a location (5c). Hence, '*Terry swept the crumbs' is not acceptable. Evidently, syntactic variation is closely related to semantic variation. Example (5a) is an activity, whereas (5c) to (5f) describe an accomplishment. More specifically, (5c) and (5d) codify a change of location; (5e) a change of state; and (5f) the creation of an artifact.

Within the context of a lexical rule approach, verbs that can appear with different complements require multiple lexical semantic representations:

> Clearly, what is needed is a theory which will predict the range of argument expressions and meanings a particular verb can be associated with. Then the different lexical entries for individual verbs could be generated from general principles which determine the range of meanings a verb can have, and the existence of these multiple lexical entries would be accounted for in a principled way (Rappaport and Levin 1998: 99).

Nevertheless, the variation in the meaning of a verb meaning is not random, but rather is subject to constraints. According to Rappaport and Levin (1998: 99), "verbs naming similar concepts appear in the same range of syntactic contexts, with verbs in each class showing an unique pattern behaviour" (Rappaport and Levin 1998: 99). For example, the pattern of variation in 'sweep', is the same for the entire semantic class to which 'sweep' belongs. In this regard, 'wipe' (see example 6), which is also a surface contact through motion verb, can appear in the same range of contexts:

(6)
a. Terry wiped.
b. Terry wiped the table.
c. Terry wiped the crumbs into the sink.
d. Terry wiped the crumbs off the table.
e. Terry wiped the slate clean.
f. Terry wiped the crumbs into a pile.

According to Carter (1988) and Levin and Rappaport (1995: 1), *linking regularities* are arguments with certain semantic roles, which are regularly associated with specific syntactic structures. The rules responsible for these associations are *linking rules.* To the extent that the semantic roles of an argument are determined by verb meaning, the existence of linking regularities is evidence that verb meaning determines syntactic structure.

They distinguish between *internally caused* and *externally caused* verbs. In this sense, *internally caused* and *externally caused* are not really properties of verbs, but of the eventualities they describe. However, generally speaking, the labels of *internally* and *externally* are associated with verbs as they are a convenient shorthand.

Externally caused verbs "imply the existence of an 'external cause' with immediate control over bringing about the eventuality described by the verb: an agent,

an instrument, a natural force, or a circumstance" (Levin and Rappaport 1995: 92). Therefore, they permit a causative alternation. Prototypically alternating verbs, such as 'break' and 'open' are found into the externally caused class, as shown in the following example: 'She broke the glass. /The glass broke'.

In contrast, an internally caused eventuality is one in which "some property inherent to the argument of the verb is 'responsible' for bringing about the eventuality" (Levin and Rappaport 1995: 91). Agentive intransitive verbs, such as 'play' and 'speak' are internally caused because the subject argument is the agent of the event. However, verbs do not have to be agentive to be internally caused. Certain non-agentive verbs taking animate arguments, such as 'blush' and 'tremble' are also internally caused because the eventuality comes as a result of bodily physical processes. Even verbs with inanimate arguments can be internally caused. These include verbs of emission, such as 'burble' (sound emission), 'flash' (light), 'stink' (smell), and 'ooze' (substance movement). Internally caused verbs do not alternate as shown in (7) (Levin and Rappaport Hovav 1995: 91):

(7)
a. Mary shuddered. / *He shuddered Mary. / He made Mary shudder.
b. Mary smiled. / *He smiled Mary. / He made Mary smile.

According to Levin and Rappaport (1995), the reason why internally caused verbs cannot alternate is because of a *linking rule*. As previously mentioned, linking rules connect positions in a verb's semantic representation to argument structure positions, which consist of a list of the participants that are expressed as syntactic arguments in the clause headed by the verb. As is well known, in predicate-argument structure, the external argument is normally mapped on to the subject position. When this happens, the first element of the internal arguments is linked to the direct object position, but can also become the subject of the clause. Based on these assumptions, Levin and Rappaport (1995) posit a linking rule that maps the immediate cause to the external argument position, referred to as the *Immediate Cause Linking Rule*. They formulate the following linking rules to account for relations between arguments and their syntactic projection (Levin and Rappaport 1995: 135-154):

- *Immediate Cause Linking Rule*: The argument of a verb that denotes the immediate cause of the eventuality described by that verb is its external argument.
- *Directed Change Linking Rule*: The argument of a verb that corresponds to the entity undergoing the directed change described by the verb is its direct internal argument.
- *Existence Linking Rule*: The argument of a verb whose existence is asserted is its direct internal argument.
- *Default Linking Rule*: An argument of a verb that does not fall under the scope of any of the other linking rules is its direct internal argument.

Levin and Rappaport's approach has made a major contribution to Linguistics. It is frequently used by the NLP community as evidence of the semantic similarity of verbs (Baker and Ruppenhofer 2002: 27). However, since it does not provide a very fine-grained classification, its practical application for NLP is limited. Korhonen and Briscoe (2004: 38) write:

> […] Levin's (1993) taxonomy, mainly deals with verbs taking noun and prepositional phrase complements, and does not provide large numbers of exemplars of the classes. The fact that no comprehensive classification is available limits the usefulness of the classes for practical NLP.

As shall be seen in the description of VerbNet (§3.2.2), new classes have now been added to Levin' classification in order to enhance verb description (Korhonen and Briscoe 2004).

In our opinion, Levin and Rappaport's linking rule approach is very valuable in that it shows the interdependence of syntax and semantics. Accordingly, we agree with some of its basic premises, such as the fact that verbs having the same meaning will presumably have similar grammatical properties and thus, it may be possible to predict behavior based on semantic classes. However, within Levin and Rappaport's approach, argument behavior depends mostly on the verb. No distinction is made between the arguments provided by the verb and the arguments supplied by the construction in which it is inserted. It is true that verbs carry meaning in themselves, but as shall be

seeing in 3.1.3.3, the construction in which a predicate in inserted can also contribute to the activation of a particular meaning.

Baker and Ruppenhofer (2002) compare Levin's classification of verbs with the classification in FrameNet (§3.1.3.2). They observed that although in the linking rule approach there is a link between syntactic and semantic alternations, reliance on syntactic alternations as a main criterion strongly conditions Levin's classification and makes it impossible to account for similar lexemes or on the contrary, link different ones:

> Although the whole thesis of Levin's work is that grouping words according to alternations tends to produce semantically coherent classes, it can also split words that are close in meaning, or lump semantically disparate words (Baker and Ruppenhofer 2002: 30).

In this sense, Baker and Ruppenhofer highlight the fact that the meaning associated with Levin's verb classes is often hard to define.

> [...] many verbs are cross-listed in classes which pick out one aspect of their meaning but do not capture separate senses. An example of this are the hundreds of verbs found in Other Alternating Verbs of Change of State (Baker and Ruppenhofer 2002: 33).

FrameNet's classifications and lexical entries are based on corpus examples (§3.1.3.2). In line with this, Baker and Ruppenhofer (2002: 29) provide evidence that many of Levin's classes have verbs that do not appear in the constructions predicted by Levin. At least, no examples were found in FrameNet corpus data. According to Croft (2003: 50), such problems could be solved to a certain extent if one takes semantics as the starting point:

> Although constructions exist, as cognitive linguists have argued, the relation between constructions (and their meanings) and their component words (and their meanings) is not as neat or as clear as the lexical rule-construction debate implies.

3.1.2 Functional theories

There are many functional and functional-cognitive (§3.1.3) theories of language, which are in opposition to formal theories (e.g. Chomsky's Generative Grammar) (§3.1.1). The objective of functional theories is to describe the use of language in real

communicative situations (Mairal et al. 2012: 222). As an alternative paradigm to Formal Linguistics, Functional Linguistics, and Cognitive Linguistics (§3.1.3) agree on the idea that communicative situations determine meaning and by extension, the structures in them. Therefore, both approaches account for the study of the real uses of languages and variation. In other words, these theories are interested in language as a communicative phenomenon.

The main theoretical premises of Functional Linguistics can be summarized as follows[74] (Cabré and Lorente 2005: 445):

- The basic language function is communication.
- Language is pragmatically motivated, which conditions meaning.
- Pragmatically constrained meaning determines the structures of different languages.
- Linguistic universals are linked to the functions of language.
- Language is a dynamic process.
- The external control of communicative acts explains linguistic variation.

Methodological premises within functional models account for the following (Cabré and Lorente 2005: 445):

- Grammar links the structure of language to communicative situations.
- Pragmatically constrained semantics fulfils a key role in the description and description of linguistic phenomena.
- In order to account for linguistic phenomena, synchronic analysis cannot be separated from diachronic analysis.
- Linguistics should analyse real data and not abstract data.
- Only by carrying out an empirical contrastive analysis of a great number of typologically different languages, can language generalizations be established.

Some of the most influential functional models were developed at the end of the 70s and beginning of 80s. The two most relevant for our purposes are the *Lexical Grammar Model* (§3.1.2.1) and *Role and Reference Grammar* (§3.1.2.2).

[74] The basic tenets of Functional-cognitive Linguistics are described in 3.1.3.

3.1.2.1 Lexical Grammar Model[75]

The Lexical Grammar Model (LGM) (Martín-Mingorance 1984, 1989, 1995; Faber and Mairal 1999), previously called the *Functional Lexematic Model* (FLM), was conceived by Mingorance in the 1980s and further developed by Faber and Mairal (1999). As shall be seen (§3.1.3.4), the *Lexical Constructional Model* has also partially incorporated the theoretical and methodological premises of the LGM.

The LGM integrates Dik's Functional Grammar (Dik 1978a, 1978b, 1989, 1997a, 1997b) and Coseriu's Lexematic Theory (1981) to onomasiologically organize the lexicon in semantic hierarchies which form *lexical domains* and *subdomains*. In other words, what differentiates the LGM from other linguistic models is that it does not represent each lexical unit individually, but rather arrives at syntactic representation based on the semantic representation derived from a structured lexicon. In the words of Faber and Mairal (1999: 66):

> This type of organization is informative because the shared semantic information in these areas of meaning can be used to predict the syntactic behavior of sets of semantically related lexemes.

The LGM focuses on verbs because much of our knowledge consists of events and states, most of which can be linguistically represented by verbs (Faber 1999). Nevertheless, in specialized language and more specifically, in Terminology, verbs have never been given their due importance. In this respect, terminologists have always focused mainly on the description of compound nominals or noun phrases (terms) (Guilbert 1973; Rey 1975; Sager 1990; L'Homme 1998).

The verbal lexicon of the LGM is organized in three axes: (1) the paradigmatic axis; (2) the syntagmatic axis; (3) the cognitive axis. Paradigmatic relations[76] are semantic relations between a set of lexemes in the same area of meaning, whereas syntagmatic relations are relations of combination and refer to the various ways in which elements within the same text may be related to each other (Saussure [1916] 1990). Finally, the cognitive axis combines the other two, and is used as an interface between language and thought. In the words of Langacker (1987: 98), "lexical structure is conceptual structure shaped for linguistic purposes according to the dictates of

[75] This section is based on Faber and Mairal (1999).

[76] Saussure referred to paradigmatic relations as *associative relations* (Saussure [1916] 1990), but Jakobson's term (Jakobson and Halle 1956) is now more commonly used.

linguistic conventions". Therefore, lexical relations codify a map of conceptual relations and provide us with an idea of what a "language of thought" would be like (Pinker 1994: 149). This language of thought could be derived from the lexical meaning of predicates and arguments and also from their configurations within the paradigmatic and syntagmatic axis.

Although the LGM was initially envisaged for general language, in recent years, it has also been successfully applied to the study of specialized language (Tercedor 1999; Montero 2002; Faber and Jiménez 2002). In fact, as shall be seen, some of the premises in this model of lexical-semantic organization have been used for verb description in this thesis concerning the environmental specialized language.

3.1.2.1.1 The concept of lexical domain

A semantic field has traditionally been regarded as a set of lexemes which cover a certain conceptual area and which are semantically related. However, its size, the design of its internal structure, as well as the relations established between the different items is still a matter of debate (Faber and Mairal 1999: 68). In this sense, the LGM has provided a more specific definition of *semantic field* or *lexical domain* (as it is referred to in the LGM).

The LGM conceives lexical domains as somewhat different from semantic fields in that criteria for membership are specified. Furthermore, the internal structure of a lexical domain is mapped out in terms of the codification of syntagmatic and paradigmatic information (Faber and Mairal 1999: 79) or the complementary principles of combination and selection (Saussure [1916] 1990; Lyons 1977: 241). Thus, a lexical domain is defined as: "the set of lexemes which together lexicalize all or part of a conceptual domain" (Faber and Mairal 1999: 59). In other words, hierarchies of lexemes, all of which share the same nuclear meaning and syntax, form a lexical domain.

According to the LGM, the lexicon is divided into twelve lexical domains. Each domain has one or two generic terms or superordinates, in terms of which all the members of the domain are directly or indirectly defined. In this sense, the *genus* or nuclear term of the definition of each lexeme marks the semantic territory covered by a specific domain or subdomain, and thus is the factor that determines lexical domain membership. The *differentiae* comprise the semantic information in the meaning of a lexeme which distinguishes it from others in the same lexical domain or part of a

domain. In Table 39, lexical domains are displayed (in square brackets) as well as their superordinate verbs (italics) (Faber and Mairal 1999: 88):

- (i) *to be* [EXISTENCE]
- (ii) *to become different* [CHANGE]
- (iii) *to have/give* [POSSESSION]
- (iv) *to say* [SPEECH]
- (v) *to feel* [EMOTION]
- (vi) *to do/make* [ACTION]
- (vii) *to use* [MANIPULATION]
- (viii) *to know/think* [COGNITION/MENTAL PERCEPTION]
- (ix) *to move (go/come)* [MOVEMENT]
- (x) *to become aware* (*notice/perceive)* [GENERAL PERCEPTION]
- (xi) *to see/hear/taste/smell/touch* [SENSE PERCEPTION]
- (xii) *to be/stay/put* [POSITION]

Table 39. Lexical domains in the verbal lexicon

It is our assertion that each verb in the lexicon belongs to one of these categories, or in the case of polysemous verbs, they could even belong to different domains. Their membership depends on the most prototypically activated semantic features.

For example, the EXTREME EVENT (§4.1), which is the focus of this thesis, is known for having an important cause/damage component. As such, the verbs most frequently activated in this frame generally belong to the domains of ACTION and CHANGE. In addition, lexical domains can be further subdivided into *subdomains*. Each subdomain focuses on a particular area of meaning and reflects a different specification of its content. This is especially important in terminology, where term meaning depends on the membership of terms in specialized knowledge subdomains.

The most generic terms in the lexicon are initially determined by *Stepwise Lexical Decomposition* (Dik 1978b), and a semantic hierarchy is thus established within each lexical domain. Generic terms can be regarded as possible candidates for the inventory of more basic terms or primitives (Mairal and Faber 2007: 283). This is also in consonance with Mel'čuk's *Decomposition Principle*, which proposes that the definition of a lexical unit must contain only terms that are semantically simpler than the lexical unit itself (Mel'čuk 1988). It is also in consonance with Mel'čuk's *General*

Principle of Lexical Inheritance (1998: 42) which, as previously mentioned (§2.1.3.1), states that when certain lexicographic data are shared by a family of related lexical units, they should be stored just once under the lexical unit in question, from which they would be inherited.

Accordingly, generic terms are the basis for the formulation of the meaning of more specific lexical items. In other words, they are the starter terms for lexical hierarchies that provide the basic structure for the construction of a conceptual network. These generic terms or near primitives correspond to a great extent to those proposed by Wierzbicka's Natural Semantic Metalanguage (NSM) (Goddard 2010; Wierzbicka 1996, 1999; Goddard and Wierzbicka 2002, 2005, 2007). Table 40 shows the semantic primitives or semantic primes identified by the NSM for English (Goddard 2010: 462):

Category	**Wierzbicka's Semantic Primitives**
substantives	I, YOU, SOMEONE/PERSON, PEOPLE, SOMETHING/THING, BODY
relational substantives	KIND, PART
determiners	THIS, THE SAME, OTHER/ELSE
quantifiers	ONE, TWO, SOME, ALL, MANY/MUCH
evaluators	GOOD, BAD
descriptors	BIG, SMALL
mental predicates	THINK, KNOW, WANT, FEEL, SEE, HEAR
speech	SAY, WORDS, TRUE
actions, events, movement, contact	DO, HAPPEN, MOVE, TOUCH
location, existence, possession, specification	BE (SOMEWHERE), THERE IS/EXIST, HAVE, BE
life and death	LIVE, DIE
time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT
space	WHERE/PLACE, HERE, ABOVE, BELOW; FAR, NEAR; SIDE, INSIDE
logical concepts	NOT, MAYBE, CAN, BECAUSE, IF
intensifier, augmentor	VERY, MORE
similarity	LIKE

Table 40. Semantic primitives or primes in the MSN

In the LGM, each category within the conceptual network is organized in terms of meaning parameters forming *lexical dimensions*, which constitute the internal organization of a lexical domain. Lexical dimensions within a field are established by means of the contrasts evidenced in the definitions of lexical units. This signifies that all lexemes within a field can be defined either directly or indirectly in terms of the same nuclear word, while the various lexical dimensions reflect the differences between the lexemes. This is directly related to the notion of *multidimensionality* in Terminology (Kageura 1997; Bowker 1997; Meyer, Eck, and Skuce 1997; Wright 1997; León 2009).

Multidimensionality occurs when a concept can be classified differently depending on the conceptual characteristics used as the basis of the subclassification (Bowker and Meyer 1993: 123). Its origin lies in the different needs and different levels of perception and cognition derived from our categorization process, which configures categories of concepts as dynamic (Faber 2011; Buendía 2012). Thus, conceptual representation can be organized from different perspectives. As shall be seen in the description of verbs that are activated within the context of the EXTREME EVENT (§4.1), the basic meaning of each verb profiles the meaning of each concept in different ways, and thus provides a way to access its multidimensionality. Thus, for example, action verbs, such as 'hit', focalize the causes/consequences of natural disasters as well as the location where they take place, and leave other characteristics such as the origin, phases, types, and other attributes in the background.

3.1.2.1.2 The paradigmatic axis

The paradigmatic axis codifies how elements are configured on the axis of selection by organizing them onomasiologically in a hierarchy of domains and subdomains, and is also a determining factor in their syntax (Faber and Mairal 1999: 80). As shown in Table 41, the initial lexical organization in the LGM is based on definitional analysis. In this sense, in order to obtain the meaning components by means of which the definition of each lexical unit is specified, various dictionaries were consulted, but no corpus was used. This is a weakness of the model, which was conceived when corpus analysis techniques and lexical analysis software programs were still in their initial stages.

For the paradigmatic description of the verbs within the EXTREME EVENT, definitions provided by specialized dictionaries and other reference material, along with

the information extracted from a multilingual corpus of texts were used (§4.2.3). In the same way as the analysis of dictionary definitions, corpus analysis offers the possibility of using linguistic clues to extract conceptual information regarding semantic relations (see 4.2 on design, compilation and description of corpus). This information is semantically classified and analyzed so that the concept can be placed in the underlying conceptual framework of a knowledge domain event (Faber et al. 2007).

Accordingly, in the LGM, definitions are then elaborated to assure structured meaning definitions and the mutual coherence of lexemes in the same meaning area. In this regard, each verb has a *genus* which corresponds to the superordinate term in the hierarchy. The more specific the hierarchy becomes, the more focalized the information will be. Hyponymy is thus the basis of hierarchical relationships. Although the lexicon has primarily a hierarchical type of organization, non-hierarchical relations are also taken into account (Faber and Mairal 1999: 63).

As an example, Table 41 shows the possibilities of selection on the paradigmatic axis for the lexical dimension *to come against sb/sth with sudden force* within the lexical domain of ACTION. As can be seen in Table 41, the subdomain of *to come against sb/sth with sudden force*, includes verbs such as 'hit', 'strike', 'collide', 'bump', 'crash', 'clash', 'knock', 'chatter', 'pound', 'thump', 'bang', 'thud', 'plop', '(pitter-) patter'. These verbs are defined in terms of the superordinate verb of the domain (i.e. 'hit'). As such, for example, 'chatter' is defined in terms of 'knock', which in turn is defined in terms of 'hit', the generic term of the subdomain.

The reason for choosing ACTION verbs was because ACTION is one of the most salient lexical domains within the EXTREME EVENT (§4.1). As such, almost every concept in the frame, (e.g. HURRICANE, EARTHQUAKE, TSUNAMI, TORNADO, AVALANCHE, LANDSLIDE, DROUGHT, FLOOD or FIRE) can produce a strong impact [ACTION] that affects another entity. This type of force is lexicalized in the form of 'hit' or one of its hyponyms.

To come against sb/sth with sudden force
hit1 to come against sb/sth with sudden force.
strike1 to hit against sth/sb else.
collide to hit sb/sth accidentally while moving (of two moving things).
bump1 to hit against sth solid and heavy esp. by accident.
crash1 to hit against sth noisily and violently (esp. causing destruction).
to collide with sth
clash to hit against sth else with a loud, harsh noise
knock1 to hit against sth, making a repeated sharp noise.
chatter to knock together repeatedly because of cold or fear (of teeth).
pound1 to hit against sb/sth repeatedly and noisily.
thump1 to hit against sb/sth hard, making a dull, heavy sound.
bang1 to hit against sth violently, making a loud noise.
thud to hit a surface, making a dull, punctual sound (of sth heavy).
plop to hit a surface gently, making a soft dull, punctual sound (of sth light).
(pitter-)patter to hit a surface, making short, light tapping sounds (of rain, footsteps).

Table 41. ACTION lexical domain. Lexical dimension of *to come against sb/sth with sudden force*

3.1.2.1.3 The syntagmatic axis

As previously mentioned, one of the basic premises of the LGM is that the English verbal lexicon is organized in sets of semantic classes and that a verb's membership in a lexical set is a determining factor in its syntactic realization. Each lexical domain is thus a repository of both syntactic and semantic information, and as such, can be regarded as indicative of the interface between syntax and semantics. As Faber and Mairal (1999: 121):

> We might even go so far as to say that a subdomain represents a kernel lexical grammar, which encapsulates all the linguistic properties of its members.

On the syntagmatic axis, each verb is associated with its inventory of complementation patterns or different activations of its argument structure. These patterns are instantiations of an underlying schema indicative of its quantitative and qualitative valency. The quantitative valency refers to the number of arguments,

whereas the qualitative valency provides the semantic characterization of arguments, i.e. their semantic roles, and selection constraints.

At this point, it is necessary to define what a semantic role is since every linguistic theory that endeavors to account for verb meaning must also include a description of verb arguments (i.e. the participants) as well how they relate to the predicate. Semantic roles generally express the set of properties that a verb entails for a given argument. Although almost every linguistic theory makes use of semantic roles in some form, there is considerable disagreement as to their number, nature or function. The two roles in which almost all theories coincide is in the role of AGENT and PATIENT. Generally speaking, an AGENT is an entity that causes an action. Either a NATURAL AGENT (e.g. a natural force) or a HUMAN AGENT carries out the action with intentionality. A PATIENT refers to the entity that is affected by an event or process. In this research, we propose an inventory of semantic roles that are typical of the EXTREME EVENT frame, and which are activated by verbs that are frequently used in environmental discourse in regard to natural disasters (§4.1).

The syntagmatic axis specifies the complementation patterns for each verb. It is not just an inventory of syntactic patterns, but illustrates the extent to which semantic information on the paradigmatic axis is relevant to the form and function of verb complementation (Faber and Mairal 1999: 114-115). As such, the syntagmatic axis deals with combinatorial relations.

For example, 'hit' is a predicate with a relatively simple syntax: NP *hit* NP. When it is used transitively, it has two arguments, both of which are noun phrases (NPs). The first one can be an AGENT or FORCE and the second is a PATIENT since there is an entity affected by the action. The semantic characteristics of these arguments can vary, and depending on whether they are more or less prototypical, the base meaning of the verb varies accordingly. In this sense, Table 42 shows the range of prototipicity in the inventory of selection constraints typical of the arguments of 'hit' (to come against something with sudden force).

As can be observed in Table 42, the most prototypical activation coincides with the realization of the verb when the AGENT is either a human agent or force and the PATIENT, a natural entity, artifact, or abstract entity. All the verbs within the dimension *to come against sth/sb with sudden force*, described above in Table 41, would contain a similar number and type of arguments as those of 'hit'.

More prototypical ↓ Less prototypical	**CONCRETE**	**AGENT/NATURAL FORCE**	**PATIENT**
	I hit the ball.	human agent	artifact
	My car hit the tree.	vehicle driven by a human agent	natural entity
	The hurricane hit the city.	natural disaster	geographic area
	FIGURATIVE		
	Employment has been hit by the crisis.	abstract event	abstract situational social entity
	She has been hit by her breakup.	abstract event	abstract situational emotional entity

Table 42. 'hit' inventory of selection restrictions

3.1.2.1.4 The cognitive axis

The third component in the LGM lexicon is the cognitive axis, which represents the intersection point of language and thought. In other words, "lexical structure on both the paradigmatic and syntagmatic axes reflects cognition" (Faber and Mairal 1999: 271). In this sense, the information codified in the paradigmatic and syntagmatic axes of the lexicon combines to form what the authors call a *predicate schema*. The notion of predicate schema in the LGM is linked to Langacker's schemas (Langacker 1987: 371) (§3.1.3.1). Predicate schemas are derived from the common properties of a set of predicates and are thus compatible with all members of the category and are defined as follows:

> [...] a modular, dynamic characterization that subsumes linguistic symbolic units obtained through the activation of lower-level schemas. These schemas are linguistically motivated and reflect our perceptions of reality (Faber and Mairal 1999: 213).

The lexical organization in the LGM is configured in hierarchies of predicate schemas at the levels of lexeme, subdomain, and domain. They have the following characteristics (Faber and Mairal 1999: 213-214): (i) they are cognitive constructs obtained by the convergence of both paradigmatic and syntagmatic features; (ii) they are modular since they were organized hierarchically within lexical domains; (iii) they are linguistic in that the units which embody them are obtained by means of lexical

structure; (iv) they are dynamic, in the sense that they established new connections with other schemas, a process which is at the base of metaphors and metonymy.

Faber and Mairal (1999: 233) are skeptical about the possibility of discovering any set of absolute rules valid for entire lexical domains. However, they claim that there is clear evidence of recurrent organizational patterns and regularities, which justify the type of lexical productivity aiming at systematizing information. In this sense, the objective of introducing the notion of *predicate schema* in the lexicon was to be able to account for the systematic inclusion of the recurrent sets of syntactic and semantic properties.

Predicate schemas operate from both an inter-and an intra-domain perspective. From an intra-domain perspective they account for recurrent semantic and syntactic parameters that constrain syntactic projections and which prompt other types of information by a set of inheritance mechanisms. From an inter-domain perspective, predicate schemas reflect the connection with other lexemes in the lexicon. The result is a *semantic macronet* which show how domains are related and how lexical dependency is linked to conceptual dependency (Faber and Mairal 1999: 273).

Nevetheless, the notion of predicate schema has evolved dramatically towards what is currently known as *lexical templates* as conceived by the Lexical Constructional Model (§3.1.3.4). Indeed, as previously mentioned, the LGM has been partially integrated into the theoretical and methodological premises of the Lexical Constructional Model (§3.1.3.4), more specifically the mechanism for describing lexical domains paradigmatically by means of hierarchies based on semantic similarities and differences, and the underlying idea of lexical templates. In this sense, as shall be seen (§3.1.3.4), lexical templates are formal constructs that contain the relevant semantic and syntactic information of each predicate in a lexical domain. In the words of Mairal and Faber (2007: 138) a lexical template is defined as a "formal representation of a lexical unit and the world knowledge elements which affect its syntactic representation".

Our research is largely based on the theoretical principles of the LGM and, by extension, on those of the Lexical Constructional Model (§3.1.3.4). Accordingly, the LGM allowed us to establish the conceptual structure of the ENVIRONMENTAL EVENT and more specifically, of the EXTREME EVENT (§4.1). It also provided us with the basic premises for paradigmatically describing the lexical domains involved in the EXTREME EVENT in terms of hierarchical structures according to meaning and for the elaboration

of lexical templates that capture the shared characteristics of verbs belonging to the same lexical dimensions. These templates are inherited by all subordinate dimensions. The ultimate goal is to create rules that can be computationally systemized to enhance and automate verb argument extraction.

Despite the usefulness of the LGM for our research, we agree with Butler (2009: 9) in that the LGM does not provide a systematic account of the mapping of semantics on to syntax since it focuses on and from the lexicon:

> [Faber and Mairal (1999)] paved the way for an integration of the semantic aspects of lexical structure with the syntactic aspects, in terms of the linkage between semantically-based hierarchies and syntactic complementation patterns. However, the model had no explicit component which provided a fully systematic account of the mapping of semantics on to syntax. Unfortunately, the parent model FG was being criticized for its own lack of a syntactic level.

Within this scenario, there is another functional theory known as *Role and Reference Grammar* (§3.1.2.2) which contains a clearly defined syntax. At the semantic level, it includes the logical structures of predicates that form the core of the entry for a given predicate in the lexicon:

> The combination of the FLM-derived proposals with the account of predicate argument structure in terms of the logical structures of RRG provides, in principle, a powerful means of relating semantics to syntax in this area (Butler 2009: 9).

The following section describes the basic tenets of Role and Reference Grammar, a linguistic theory that has also been integrated in the Lexical Constructional Model (§3.1.3.4) and on which some of the theoretical premises of our verb description are also based. More specifically, the set of semantic roles used in this study are largely based on the thematic roles proposed by Role and Reference Grammar and its further generalization of thematic roles into the macroroles of ACTOR and UNDERGOER (§3.1.2.2.3).

3.1.2.2 Role and Reference Grammar

Role and Reference Grammar (Van Valin and LaPolla 1997; Van Valin 2005), henceforth RRG, is a functional linguistic model. It was conceived with the objective of

describing the interaction of syntax, semantics and pragmatics in different grammatical systems (Van Valin 2005: 1). According to this theory, grammar is to a great extent determined by semantics and pragmatics, which means that syntax is no longer autonomous. In other words, pragmatics and semantics are thought to be powerful components of language that constrain syntax.

RRG focuses on function, meaning, and language use. It stresses the importance of *function* in that it considers that function and meaning are elements that condition form, but not the other way around. It stresses *meaning* because it believes that grammar is structured by semantic and conceptual content; and finally it stresses *language use*, since it does not separate the study of language from its communicative function.

As shown in Figure 39 (Mairal and Cortés 2006: 102), the architecture of RRG consists of the following: (i) a *lexicon component* where predicates and their corresponding meaning representations are stored, i.e. semantic representation; (ii) a *syntactic inventory* that contains the inventory of syntactic constructions of the grammar in a specific language, i.e. the syntactic representation. Since our approach to verb argument structure starts from semantics to describe syntax, this section only describes the semantic component of RRG.

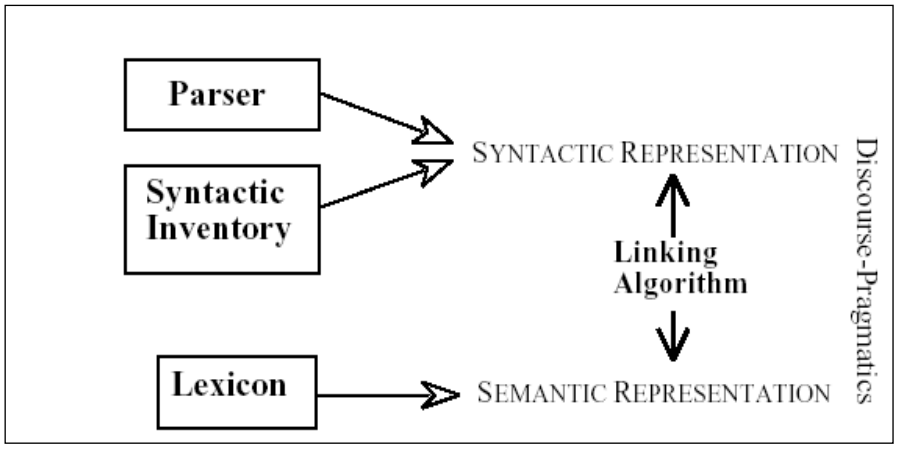


Figure 39. The architecture of RRG

As shown in Figure 39, the syntactic and semantic representations are mapped by means of a bidirectional linking algorithm, which goes from syntax to semantics and from semantics to syntax since the meaning of a predicate presumably determines to a great extent its syntactic behavior. The range of morphosyntactic patterns that a

predicate has can thus be predicted from its meaning. Accordingly, the architecture of RRG also includes a set of rules that account for the syntax-semantics interface.

3.1.2.2.1 Aktionsart types

As previously mentioned, the *lexicon* refers to the place where words with their corresponding semantic representations are stored. In this sense, it is like a mental dictionary. Speakers describe events that occur and the entities involved in these events by means of linguistic expressions called *sentences*. In RRG, such events are referred to as *states of affairs (SoA)*. In other words, sentences are the lexical expression of SoAs, where elements of the sentences (mainly verbs, but also other predicative constructions) can express four different SoAs: (i) *situations* which are static; (ii) *actions*, which are dynamic; (iii) *processes*, involving a change that takes place over a bounded period of time; (iv) *events*, which happen instantly. As shall be seen, the verbs that combine with the concepts activated in the EXTREME EVENT frame in our study mainly describe the process events in which they can participate. Therefore, our description of predicates focuses on processes and events.

Following Vendler's classification of verbs (1967) and a modified version of Dowty's (1979) representation system, RRG classifies verbs[77] according to their *Aktionsart*, which is the German label for *SoAs*. Table 43 describes the four different types of SoAs and their corresponding Aktionsart, along with their definition and verb examples (Mairal and Cortés 2006: 106). It is important to highlight that each type of Aktionsart has a causative variant, when the action expressed by the verb is induced. As shown in Table 43, the difference between the various types of Aktionsarts can be established in terms of three parameters (Mairal and Cortés 2006: 106):

- [± static]: the action implies a change in the ACTOR's location.
- [± telic]: the action goes on until a terminus point, which is necessary for the action to fulfil its meaning.
- [± punctual] the process cannot be divided into different stages.

[77] Whenever we refer to verbs in this section, we refer to verbs and all other predicative elements.

State of Affair	Corresponding Aktionsart	Definition	Defining parameters	Examples
Situation	State	static, non-dynamic state of affairs, involving the location of a participant, temporally unbounded	[+static],[-telic], [-punctual]	know, believe, have, be sick, love.
Event	Achievement	happen instantly, changes of state and changes of activities as well, inherent terminal point	[-static],[+telic], [+punctual]	pop, explode, shatter collapse
Process	Accomplishment	involve change and take place over time, bounded.	[-static],[+telic], [-punctual]	melt, freeze,(intrans.), dry, learn
Action	Activity	dynamic in which a participant does something, no inherent terminal point	[-static], [-telic], [-punctual]	march, walk, roll (intrans.), think, rain

Table 43. Description of Aktionsart types in RRG (Mairal and Cortés 2006: 106)

Generally speaking, Aktionsart codifies the kind of action that a verb carries out. In this sense, a verb will have one basic Aktionsart interpretation although more interpretations are possible depending, for example, on the expression or omission of certain arguments. The use of definitive NP objects with verbs of consumption and creation and the use of directional PPs with verbs of movement also may carry different interpretations of the verb's Aktionsart. This leads to a new category referred to as *active accomplishment* and, by extension, *causative active accomplishment*. Another category added to the classification of Vendler was that of *semelfactive*, which includes verbs such as *bang*, *sneeze*, *flash*, *knock* that encode punctual events, not conducive to a result state.

Consequently, RRG, as based on Vendler's Aktionsart distinctions, divides verbs into states, activities, achievements, accomplishments, semelfactives and active accomplishments, as well as their respective causative variants.

3.1.2.2.2 Logical structures

Each Aktionsart type is represented formally by means of a logical structure (LS), using a metalanguage valid for all languages. LSs constitute the nucleus of the semantic representation of lexical units in the lexicon. In this regard, Van Valin and LaPolla (1997: 91) state:

> [...] arriving at a decompositional system is a compromise between the demands of semantics (make all the necessary distinctions relevant to meaning) and those of syntax (make syntactically relevant distinctions that permit the expression of significant generalizations) [...] so from the point of view of syntactic theory the most desirable system of decomposition is one which is just fine-grained enough (and no more) to make the distinctions necessary for capturing linguistically significant generalizations about syntax, semantics and their interaction.

The basic LS associated with each Aktionsart type is displayed in Table 44 (Van Valin 2005: 45). As can be deduced, states and activities are primitives, whereas accomplishments and achievements consist of either a state or activity plus a BECOME, an INGR, or a SEML operator.

Aktionsart type	**Logical Structure**
State	**predicate'** (x) or (x,y)
Activity	**do'** (x, [**predicate'** (x) or (x,y)]
Achievement	INGR **predicate'** (x) or (x,y), *or* INGR **do'** (x, [**predicate'** (x) or (x,y)]
Accomplishment	BECOME **predicate'** (x) or (x,y), *or* BECOME **do'** (x, [**predicate'** (x) or (x,y)]
Semelfactive	SEML **predicate'** (x) or (x,y), *or* SEML **do'** (x, [**predicate'** (x) or (x,y)]
Active accomplishment	**do'** (x, [**predicate$_1$'** (x, (y))] & INGR **predicate$_2$**' (z,x) or (y)
Causative	α CAUSES ß where α, ß are LS of any type

Table 44. Basic RRG logical structures

As shown, there are three types of elements in this metalanguage: (i) elements in boldface followed by a prime, which are constants (usually predicates) and are regarded as semantic primes; (ii) elements in normal typeface (e.g. x, y), which are variables and express positions that are to be filled by expressions of particular languages when building the semantic representation of individual sentences; (iii) elements in capitals, which are operators that modify constants.

3.1.2.2.3 Thematic roles and macroroles

Once the different Aktionsart types of predicates and their logical structures are specified, the next step in the semantic description of a predicate in RRG is to account for the semantic relations between the arguments of a logical structure and the verb in

question, i.e. its semantic roles. RRG establishes semantic roles at two different levels: (i) *thematic roles*, which are directly related to the type of predicate that they occur with; (ii) *macroroles*, which are generalized semantic functions. As shall be seen, the verb entry description proposed in this thesis also includes semantic roles, and in line with RRG, we distinguish between thematic roles and macroroles.

RRG does not provide a closed list of thematic roles. They are determined by its position in the logical structure of the predicate that they occur with. Van Valin and LaPolla (1997:115) list some of the most common thematic roles together with their definitions:

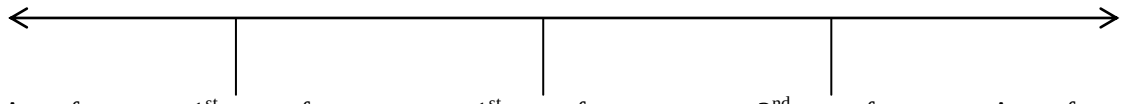
State predicates:		
A. One-argument predicates:		
1. State or condition	**broken'** (x)	x= PATIENT
2. Existence	**exist'** (x)	x = ENTITY
B. Two-argument predicates		
1. Perception	**hear'** (x,y)	x= PERCEIVER y= STIMULUS
2. Cognition	**know'** (x,y)	x= COGNIZER y= CONTENT
3. Desire	**want'** (x,y)	x= WANTER y= DESIRE
4. Possession	**have'** (x,y)	x= POSSESSOR y= POSSESSED
5. Internal experience	**feel'** (x,y)	x=EXPERIENCER,Y=SENSATION
Activity predicates:		
A. One-argument predicates:		
1. Light emission	**do'** (x, [**shine'** (x)])	x= L-EMITTER
2. Sound emission	**do'** (x, [**gurgle** (x)])	x= S-EMITTER
3. Static motion	**do'** (x, [**spin'** (x)])	x= ST-MOVER
4. Motion	**do'** (x, [**walk'** (x)])	x= MOVER
B. Two-argument predicates		
1. Performance	**do'**(x,[**sing'** (x,(y))])	x=PERFORMER y=PERFORMANCE
2. Consumption	**do'** (x, [**eat'** (x, (y))])	x= CONSUMER y=CONSUMED
3. Directed Perception	**do'** (x, [**see'** (x, (y))])	x= OBSERVER y=STIMULUS
4. Use	**do'** (x, [**use'** (x, (y))])	x= USER y=IMPLEMENT

Table 45. Sample of some thematic roles in RRG

As shown, PATIENT, ENTITY, PERCEIVER, STIMULUS, COGNIZER, CONTENT, WANTER, DESIRE, POSSESSOR, POSSESSED, EXPERIENCER, SENSATION, L-EMITTER, S-EMITTER, ST-MOVER, MOVER, PERFORMER, PERFORMANCE, CONSUMER, CONSUMED, OBSERVER, STIMULUS, USER or IMPLEMENT are some of the thematic roles proposed in RRG. These thematic roles correspond to the Aktionsart types of states and activities. In consonance with this, the thematic roles of the rest of Aktionsart types, i.e. achievements,

semelfactives, accomplishments and active accomplishments are the same as the corresponding activity or state predicate that is part of their LS, since the operators BECOME, INGR AND SEML do not affect the argument structure in the formula. As previously mentioned, only activities and states are responsible for defining the types of thematic relations.

Despite the apparent diversity of thematic roles, RRG establishes only five relevant distinctions, which correspond to the five possible argument positions in the logical structures. Figure 40 displays the thematic roles continuum, depending on their position in the LS (Van Valin 2005: 58):



Arg of DO	1st arg. of **do'** (x,...	1st arg. of **pred'** (x,y)	2nd arg. of **pred'** (x,y)	Arg. of state **pred'** (x)
AGENT	EFFECTOR	LOCATION	THEME	PATIENT
	MOVER	PERCEIVER	STIMULUS	ENTITY
	ST-MOVER	COGNIZER	CONTENT	
	L-EMITTER	WANTER	DESIRE	
	S-EMITTER	JUDGER	JUDGMENT	
	PERFORMER	POSSESSOR	POSSESSED	
	CONSUMER	EXPERIENCER	SENSATION	
	CREATOR	EMOTER	TARGET	
	SPEAKER	ATTRIBUTANT	ATTRIBUTE	
	OBSERVER		PERFORMANCE	
	USER		CONSUMED	
			CREATION	
			LOCUS	
			IMPLEMENT	

Figure 40. Thematic roles continuum according to argument positions in LS

Within RRG, the thematic role of AGENT is very restrictive. In this regard, one argument will be assigned the function of AGENT, only if the semantics of the predicate that it occurs with requires the presence of a volitional entity that acts intentionally and controls his/her actions.

Apart from the thematic roles listed, there are others of great importance that cannot be defined in terms of the five positions in Figure 40, such as the role of INSTRUMENT, related to both EFFECTORS and IMPLEMENTS, as well as the roles of SOURCE, GOAL and RECIPIENT.

Along with thematic roles, RRG recognizes another type of semantic role referred to as *macroroles*. Macroroles are generalizations across different argument types that have significant grammatical consequences (Van Valin and LaPolla 1997: 139). RRG distinguishes between two different types of macroroles: ACTOR and UNDERGOER. Generally speaking, the group of thematic roles that are subjects in transitive active sentences and prepositional complements in passive sentences have the macrorole of ACTOR, whereas those that behave as direct objects in active sentences and as subject in passives have the macrorole of UNDERGOER. As such, the most prototypical ACTORS are AGENT arguments and the most common UNDERGOERS are PATIENTS. According to the ACTOR-UNDERGOER hierarchy in the LS of a predicate with two arguments, the leftmost argument is the ACTOR and the rightmost one is the UNDERGOER. Although this is the default situation, two different realizations of the same LS can assign different macroroles to the same argument. This is the case of the *dative shift* alternation, which is typical of predicates such as *give.* As described by Mairal and Cortés (2006: 128), the THEME-POSSESSED argument, i.e. *clavichord*, is assigned the macrorole of UNDERGOER (8a), whereas in (8b), the UNDERGOER macrorole corresponds to the RECIPIENT, i.e. *Philip*:

(8)
[**do'** (Gwendolyn, Ø)] CAUSE [BECOME **have'** (Philip, clavichord)]
a. Gwendolyn[Actor] gave a clavichord [Undergoer] to Philip
b. Gwendolyn [Actor] gave Philip [Undergoer] a clavichord

In this sense, both RECIPIENTS and THEMES can be UNDERGOERS when there is a choice. Consequently, PATIENTS will never alternate since they are always in state structures with a single argument. Regarding ACTORS, when there is more than one possible candidate susceptible to be assigned the macrorole of ACTOR, AGENT-EFFECTORS are higher ranking argument than INSTRUMENTS and the first EFFECTOR has priority over the second EFFECTOR for the role of ACTOR when there is a casual chain.

The default macrorole assignment principles are related to their number of arguments and their nature (Van Valin and LaPolla 1997: 152-153). The number of macroroles that a predicate takes is generally predictable from its LS. This number is always less than or equal to the number of arguments in its LS. Therefore, the only three possibilities are 0, 1 or 2 macroroles. Regarding the nature of verbs that only take one

macrorole, if a verb has an activity predicate in its LS, the macrorole is ACTOR, whereas if it does not have an activity predicate in its LS, the macrorole is UNDERGOER.

RRG is one of the most valuable models in contemporary linguistics since it can be the basis for cognitive models (§3.1.3). The Lexical Constructional Model (LCM, §3.1.3.4) proposes the enhancement of LSs with lexical templates (§3.1.3.4.1). Lexical templates in the LCM provide a finer-grained semantic decomposition of state and activity predicates by using a universal semantic metalanguage that consists of a number of primitives (or undefinable items) and a list of operators (or lexical functions) that account for the semantic and pragmatic parameters relevant to the meaning of a predicate.

3.1.3 Functional-cognitive theories

Functionalist approaches (§3.1.2) tried to provide the answer to issues that more syntactically-oriented theories were unable to address. However, there were phenomena that functionalists could not describe either, such as the unexpected differences in argument structure between apparently similar verbs or on the contrary, coincidences between verbs belonging to different lexical domains. Although functional linguistic theories try to make explicit the relationship between syntax and semantics, there is also the need to explore the underlying reasons for the existence of what many have called the *interface* between syntax and semantics (Faber and Ureña 2012: 75). This led to the emergence of *functional-cognitive approaches* in linguistics.

Cognitive Linguistics arose in the United States in the 1970s. Its basic premises were developed by Lakoff (1987), Langacker (1987, 1991), and Fillmore (1977, 1982, 1985; Fillmore and Atkins 1992), among others. Although each of these linguists had a somewhat different approach to language description and linguistic theory, they were all interested in the relation between language and mind. Their work was a reaction to Chomsky's Generative Grammar, which was the prevailing linguistic model at the time. In consonance with functionalists (§3.1.2), they consider meaning (and thus semantics) to be central to language, and thus, the main focus of study. What differentiates cognitive linguistics from functional and formal theories and is the reason why this new trend was referred to as *cognitive* is explained by Langacker (2008: 8):

> "What is **cognitive** about Cognitive Grammar? Or about cognitive linguistics in general?" These do not merit the label merely by proclaiming that language is part of cognition and

> that linguistic investigation contributes understanding the human mind—that much is shared by many approaches, both formal and functional. Then what links to cognition do distinguish cognitive linguistics from either formal linguistics or other strains of functionalism? Within functionalism, linguistics stands out by emphasizing the semiological function of language. It fully acknowledges the grounding of language in social interaction, but insists that even its interactive function is critically dependent on conceptualization. Compared with formal approaches, cognitive linguistics stands out by resisting the imposition of boundaries between language and other psychological phenomena. Insofar as possible, linguistic structure is seen as drawing on other, more basic systems and abilities (e.g. perception, memory, categorization) from which it cannot be segregated. Rather than constituting a distinct, self-contained entity (a separate "module" or "mental faculty"), language is viewed as an integral facet of cognition.

The publication of *Women, Fire and Dangerous Things* (Lakoff 1987) further affirmed Cognitive Linguistics as a new approach to language[78]. Along with Lakoff's work, Langacker's monograph *Foundations of Cognitive Grammar: Theoretical Prerequisites* (1987, 1991) is also one of the first formal proposals in Cognitive Linguistics. Fillmore's ideas developed into *Frame Semantics* (§3.1.3.2) and in collaboration with others, into *Construction Grammar* (§3.1.3.3). Leonard Talmy is also known for his pioneering work in cognitive linguistics. His work is based on the relationship between semantic and formal linguistic structures and that of semantic typologies and universals (Talmy 2000ab)

Along with this work, Rosch's (Rosch 1973ab; Rosch and Lloyd 1978) research on categorization also had a direct influence onto the development of Cognitive Linguistics. As is well known, categorization is a mental process by which humans group entities into cognitive categories. Rosch provided evidence that categorization does not lead to clearly delimited categories of elements with shared properties, but rather to graded categories with fuzzy boundaries. In other words, categories have a prototype, generally represented by a schematic set of features or by an exemplar at the center, which possesses the features that are the most typical of category members. Therefore, within a category, there are entities that have a greater resemblance to the prototype (and are thus at the center of the category) and others that have less of a resemblance (and are thus closer to the periphery. In this sense, Rosch's proposal challenged the classical view of categorization, which is based on necessary and

[78] Lakoff is also well known for his work on metaphor (Lakoff 1993; Lakoff and Johnson 1980, 1999, *inter alia*).

sufficient conditions. As such, it paved the way to the definition of categories by means of features and relations of graded similarity.

As a theoretical proposal, Cognitive Linguistics includes insights from different fields, such as psychology, philosophy, anthropology, neurology, artificial intelligence and linguistics. All of these domains underline the importance of cognition in language, and share certain principles regarding the cognition processing and its consequences for human behavior. An important feature of Cognitive Linguistics is that its main premises are not derived from a theory, but emerge from empirical data confirmed by other disciplines within Cognitivism. Geeraerts (1995: 111-112) describes what could be included under the label of *Cognitive Linguistics* and implies that Cognitive Linguistics is closely linked to functional linguistics, psycholinguistics, linguistic description, pragmatics, and discourse studies:

> Because cognitive linguistics sees language as embedded in the overall cognitive capacities of man, topics of special interest for cognitive linguistics include: the structural characteristics of natural language categorization (such as prototypicality, systematic polysemy, cognitive models, mental imagery and metaphor); the functional principles of linguistic organization (such as iconicity and naturalness); the conceptual interface between syntax and semantics (as explored by cognitive grammar and construction grammar); the experiential and pragmatic background of language-in-use; and the relationship between language and thought, including questions about relativism and conceptual universals.

We now offer a brief overview of the most relevant Cognitive Linguistic models for our research purposes. The following sections describe relevant aspects of Langacker's Cognitive Grammar, Frame Semantics (as applied in FrameNet), Construction Grammar, and finally the Lexical Constructional Grammar.

3.1.3.1 Langacker's Cognitive Grammar[79]

As previously mentioned, Cognitive Grammar, henceforth CG, is the result of the work of Langacker (Langacker 1987, 1991, 2008). The original name given to this grammar in the 1970s was *Space Grammar* (Langacker 1982). Although CG is functional, it also provides explicit diagrams of language structure. This is the reason why CG is regarded by some functionalists as a formal model. However, Langacker (2008: 8) humorously observes that "formalists tend not to make that mistake".

[79]The content of this section is based on Langacker (2008).

The main assertion of Cognitive Grammar is that "grammar is symbolic in nature" (Langacker 2008: 5). As is well known, symbolism lies in the pairing between a semantic structure and a phonological structure. Grammar is thus conceived as an abstract symbolic structure that links a semantic structure (i.e. conceptualizations exploited for linguistic purposes, especially as the meaning of expressions) with a phonological structure (i.e. sounds, gestures or orthographic representations), so that one is able to evoke the other. The basic premise of CG is that nothing beyond symbolic structures needs to be activated for the characterization of complex expressions (Langacker 2008: 16):

> Of course, a higher-level symbolic structure is itself capable of entering into a combinatory relationship, producing a more elaborate symbolic assembly, as shown in diagram (c) ['moonless night']. We can say that a series of structures like (a) ['moon'], (b) ['moonless'], and (c) exhibit progressively greater **symbolic complexity**. Through repeated combination, at successively higher levels of organization, assemblies having any degree of symbolic complexity can be formed.

In this sense, it can be deduced that CG conceives lexicon and grammar as forming a *continuum*. A direct consequence of this position is that all grammatical descriptions, i.e. elements such as nouns or subjects, must carry meaning. In other words, "grammar is meaningful" (Langacker 2008: 3):

> […] grammar is meaningful. This is so in two respects. For one thing, the elements of grammar—like vocabulary items—have meanings in their own right. Additionally, grammar allows us to construct and symbolize the more elaborate meanings of complex expressions (like phrases, clauses, and sentences). It is thus an essential aspect of the conceptual apparatus through which we apprehend and engage the world. And instead of being a distinct and self-contained cognitive system, grammar is not only an integral part of cognition but also a key to understanding it (Langacker 2008: 3-4).

Broadly speaking, meaning consists of both conceptual content and a particular way of construing that content (Langacker 2008: 43):

> Indeed, the meaning of many linguistic elements—especially those considered "grammatical"—consists primarily in the construal they impose, rather than any specific content. Yet every element evokes some content (however schematic it might be), and conversely, any content evoked is construed in some fashion.

In CG, in order to refer to content, the term *domain* is adopted. In line with this, an expression can invoke various domains as the basis of its meaning. The group of domains that an expression activates is referred to as its *matrix*. In addition, CG highlights the existence of still higher-level conceptions than those activated by the domains, similar to Fillmore's frames (§3.1.3.2), and he refers to them as *non-basic domains.* As previously mentioned, for our analysis of EXTREME EVENT verbs, we will also classify verbs in terms of domains (the lexical domains within the Lexical Grammar Model, §3.1.2.1) and in terms of Fillmore's frames (§3.1.3.2). In contrast, Langacker's concept of domain is more fluid and imprecise. A domain can be the product of any immediate context, and in that sense, resembles the online cognition mentioned by Barsalou (2003).

Construal, which in many ways is similar to multidimensionality in Terminology, can be defined as our ability to conceive and portray the same situation in alternate ways, as illustrated in Figure 41 (Langacker 2008: 43):

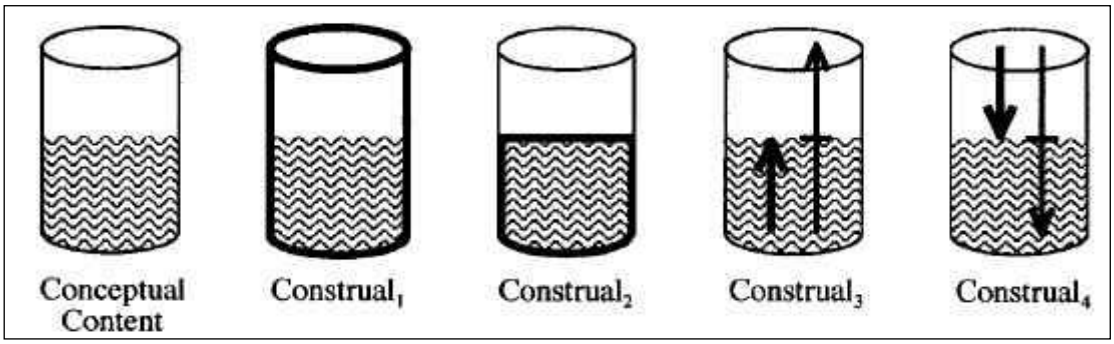


Figure 41. Construal illustration for a glass containing water

As shown, the conceptual content refers to the conception of a glass containing water, which occupies half of its volume. Once it is linguistically encoded, a certain construal is necessarily imposed. In this sense, (1) designates the container (i.e. the glass with water in it); (2) refers to the liquid it contains (i.e. the water in the glass); (3) designates the relationship by means of which the volume of the liquid is just half of its potential volume (i.e. the glass is half-full); and (4) refers to the relationship by means of which the volume of the water is just half of its volume (i.e. the glass is half-empty).

Construal depends on various factors, namely, on (i) how closely we examine the scene (i.e. specificity); (ii) what we choose to look at (i.e. focusing), including the notions of *foreground* and *background*: "the selected content is foregrounded relative to

unselected content" (Langaker 2008: 57); (iii) where we view it from (i.e. perspective); and (iv) which elements we pay most attention to (i.e. prominence).

As far as *prominence* is concerned, in order to understand it, it can be said that "within a category, the prototype has greater prominence than its various extensions" (Langacker 2008: 66), existing two types of prominence: (i) *profiling*; (ii) *trajector/landmark alignment*. Generally speaking, an expression's profile can be understood as the specific focus of attention within its immediate scope (Langacker 2008: 66):

> As the basis for its meaning, an expression selects a certain body of conceptual content. Let us call this its conceptual **base**. Construed broadly, an expression's conceptual base is identified as its maximal scope in all domains of its matrix (or all domains accessed on a given occasion). Construed more narrowly, its base is identified as the immediate scope in active domains—that is, the portion put "onstage" and foregrounded as the general locus of viewing attention. Within this onstage region, attention is directed to a particular substructure, called the **profile**. Thus an expression's profile stands out as the specific **focus** of attention within its immediate scope

Accordingly, very often various expressions that activate the same conceptual content differ in their meaning because they profile different parts of the common base. At this point, it should be underlined that an expression can profile either a thing or a relationship. For instance, concerning our description of verbs within the EXTREME EVENT in its sense of natural disaster, the difference in the meaning of 'damage' and 'devastate' lies in the fact that 'damage' profiles the action of affecting something negatively, whereas 'devastate' adds the notion of damaging completely. This example provides evidence that difference in meaning can sometimes be found in alternate choices of profile within the same conceptual base.

However, CG also states that it is possible to find expressions that despite having the same conceptual base and profiling the same relationships, differ in their semantic content. In order to be able to account for this phenomenon, an additional kind of prominence is needed: the *trajector/landmark alignment. Alignment* refers to the coding relation between the clause and the event. The *trajector* can be said to be the most prominent participant in a relationship. It is the entity "construed as being located, evaluated, or described" (Langacker 2008: 70). In this sense, it can be referred to as being the primary focus of the profiled relationship. If a secondary focus appears in the

relationship, then it is referred to as the *landmark.* In this regard, Langacker (2008: 70) stresses:

> Expressions can have the same content, and profile the same relationship, but differ in meaning because they make different choices of trajector and landmark.

Accordingly, examples in 9 referring to the domain of CHANGE within the EXTREME EVENT represent alternate *construals* of the same event, more specifically concerning the profiling and trajector/landmark organization. 9a highlights that it was the *hurricane* that destroyed the city, whereas in the 9b, emphasis is placed on the entity destroyed by the hurricane, i.e. *the city*:

(9)
a. The hurricane destroyed the city.
b. The city was destroyed by the hurricane.

In our analysis, we have found that the frame (§3.1.3.2) is the element that constrains the potential relations holding between the concepts. In this sense, the construals that the frame allows are only a subset of the construals allowed by the arguments. For example, the argument 'cyclone', in 'the cyclone hit the city', constrains the possible construals of the predicate 'hit' and viceversa. In other words, the meaning of the verb is imposed by the meaning of its subjects, but at the same time, the verb also constrains the arguments that can combine with it. As such, as previously mentioned in 2.1.4 in 'fire burns', the predicate 'burn' only admits noun phrases designating combustible entities. In addition, 'fire' requires a verb designating a combustion process ('burn').

Another important assumption of Langacker's Cognitive Grammar is that grammar is found in "schematized patterns of conceptual structuring and symbolization" (Langacker 2008: 27). In this sense, he introduces the notion of *conceptual archetypes,* which are similar to predicate schemas previously described within LGM (§3.1.2.1) and, consequently, to lexical templates within the Lexical Constructional Model (§3.1.3.4). Conceptual archetypes can be a physical object, the human body, the human face, a whole and its parts, a physical container and its contents, seeing something or holding something. They are all experientially grounded

concepts in our everyday lives and act as prototypes for clausal elements, which are a crucial factor that determines their configuration.

In regards to clause structure in CG, a verb is regarded as any expression that profiles a process, whereas clauses are conceived as the basic means of talking about the world (Langacker 2008: 354):

> Usually a discourse consists primarily of a series of clauses; only rarely does it comprise a series of nominals. The main reason for evoking nominal referents is to describe their participation in relationships.

In this sense, a key dimension in clausal organization is that of the *role* of nominal participants with respect to the verb, which refers to both *semantic roles* (i.e. such as AGENT, PATIENT, INSTRUMENT) and how these semantic roles map onto *grammatical roles* (e.g. subject and objects). Langacker refers to roles as *archetypal roles*. Archetypal roles are treated by CG not as linguistic roles, but rather as pre-linguistic conceptualizations. At one level, the roles of SETTING, LOCATION and PARTICIPANT are distinguished, and more specifically, within PARTICIPANT roles, those of AGENT, PATIENT, INSTRUMENT, MOVER, EXPERIENCER, and ZERO are included.

Since archetypal roles represent conceptual archetypes, Langacker (2008: 355ff) proposes different archetypes to account for clause structure. One archetype is the organization of a scene into a *global setting* (e.g. a room, a building, a geographical region) with a number of participants, which can be found at a specific location. In this sense, the location is part of the setting, but is not itself a participant.

The *stage model* is another basic archetype proposed by Langacker, which compares the way we apprehend the world to the way that we watch a play (Langacker 2008: 356):

> We cannot see everything at once, so viewing the world requires the directing and focusing of attention. From the maximal field of view, we select a limited area as the general locus of attention (the analog of looking at the stage). Within this region, we focus our attention specifically on certain elements (analogous to actors and props).

More specifically, in order to account for the appropriateness of clause structure to archetypal roles, Langacker uses the archetypes of *action change* and *billiard-ball model*, which are closely linked to each other. In this sense, in a clause structure, one

entity acts as an energy source and transfers energy to a second entity through impact. This energy is received and consumed by the second entity. For example, in the clause 'the hurricane hit the city', 'hurricane' is the energy source that transfers its energy to *city*. The 'city' is regarded as the energy sink since it consumes all of this energy without passing any of it on to another entity.

Langacker describes this phenomenon by means of the archetype of *action chain*, which states that an action starts with the transmission of energy from an entity (the head) to the next entity and so on, until it reaches a last entity (the tail) which consumes the energy and does not transmit it anymore. It is similar to a game of billiards in which some objects supply the necessary energy and others transmit or absorb it.

A final group of archetypes concern the *speech event* itself (e.g. speaking, listening, and engaging in a social interaction); the *basic speech acts* (e.g. stating, ordering, asking, or promising), and the *default viewing arrangement***,** that of two interlocutors being together in a fixed location, using a shared language to describe occurrences in the world around them (Langacker 2008: 357).

The different archetypes described are interconnected and can be combined. A prototypical combination is what Langacker refers to as the *canonical event model*. It is formed by an AGENT, which acts on a PATIENT to induce a change of state. This event is organized around a global setting. It is the focus of attention within the onstage region, and is apprehended from offstage by a viewer, not involved in the event:

> The stage model contributes the notion of an event occurring within a setting and a viewer (V) observing it from an external vantage point. Inherited from the billiard-ball model is the minimal conception of an action-chain, in which onediscrete object transmits energy to another through forceful physical contact. Moreover, the action-chain head is characterized as an agent, and its tail as patient that undergoes a resultant change of sate […]. In sum, the canonical event model represents the normal observation of a prototypical action (Langacker 1990: 286).

These archetypes can be applied to the assignment of syntactic functions to semantic roles. The *action change* and *billiard game model*, for example, can provide an answer to why in clause structures formed by prototypical verbs (i.e. profiling an AGENT PATIENT interaction) the AGENT corresponds to the energy head and the PATIENT, to the tail. The reason is because the AGENT initiates the chain energy and the PATIENT receives it. Consequently, syntactically speaking, AGENTS are frequently subjects, and

PATIENTS are objects. In chains of three elements, the situation is basically the same though the intermediate role corresponds to that of INSTRUMENT. INSTRUMENTS and LOCATIONS are normally expressed by prepositional phrases, and SETTINGS by adverbial phrases.

In connection with the action change archetype and billiard ball model, an AGENT is the energy source and is thus defined as an individual who willfully initiates and carries out an action, typically a physical action affecting other entities. In contrast, a PATIENT is an inanimate and non-volitional entity which undergoes an internal change of state as the result of being affected by outside forces. It is thus the 'energy sink' and the final participant in an action chain. An INSTRUMENT is something used by an AGENT to affect another entity. Thus, as previously mentioned, it acts as the intermediary in the transfer of energy from AGENT to PATIENT. EXPERIENCER is a role that refers to mental experience, whatever its nature, intellectual, perceptual, or emotional. Therefore, it is generally an animate entity. In contrast, a MOVER is defined as anything that moves (i.e. changes position in relation to its external surroundings). Finally, the role ZERO is adopted for participants whose role is conceptually minimal and non-distinctive. This is the neutral or baseline role of participants that merely exist, occupy some location, or exhibit static properties.

However, this description of roles corresponds to a prototypical situation within the canonical event. In this regard, not all situations are prototypical nor are all events canonical. If this were the case, there would be no variation in the semantic nature of grammatical roles (e.g. a subject will simply be the AGENT). As Langacker highlights (2008: 358), an AGENT-PATIENT interaction is just one occurrence, but there are other pairings with different semantic import.

Similarly to the RRG macroroles of ACTOR and UNDERGOER (§3.2.1.3.2), CG distinguishes between *agent orientation alignment*, on the one hand, and *theme orientation alignment*, on the other hand. It is mainly a matter of whether, in the baseline situation, a trajector is conferred on the most AGENT-like participant or on the most THEME-like participant. Generally speaking, THEME includes what can be referred to as passive semantic roles, such as the roles of PATIENT, MOVER, EXPERIENCE and ZERO. These two orientations are illustrated in Figure 42, which shows how clause structures map onto archetypal events with one or two participants (Langacker 2006: 123):

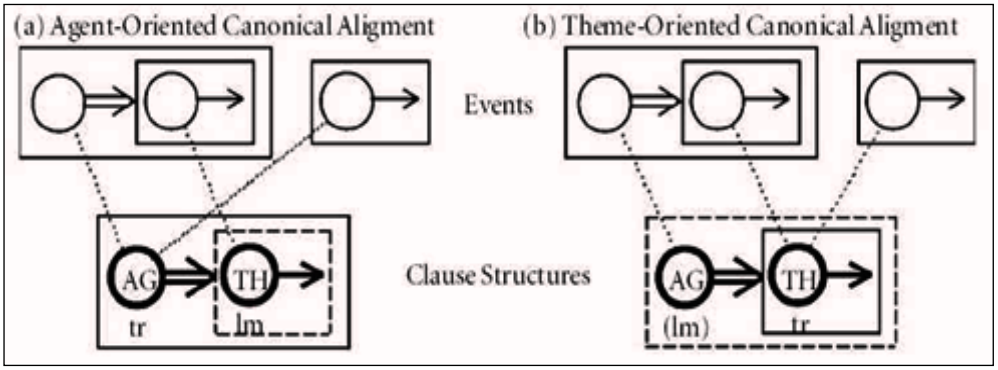


Figure 42. AGENT and THEME-oriented alignment

Basically, Langacker emphasizes that in order to provide a detailed description of clause structure, it is necessary to analyze the verb along with its arguments and their semantic roles. His notion of *construal* is of paramount importance for the representation of conceptual content. In this sense, in our description of the EXTREME EVENT verbs, the frame is the element that constrains the potential relations between the concepts, and the construals that the frame permits are only a subset of the construals sanctioned by the head argument.

3.1.3.2 Frame Semantics and FrameNet

Frame Semantics is an approach to the study of lexical meaning based on the work of Fillmore (Fillmore 1977, 1982, 1985; Fillmore and Atkins 1992). Frame Semantics stems from Fillmore's Theory of Case Grammar (Fillmore 1968). The Theory of Case Grammar was initially a contribution to Generative-Transformational Grammar. It stated that syntactic deep structures were best expressed as configurations of 'deep cases', which were assigned general semantic-role names, such as AGENT, PATIENT, GOAL, etc.

The idea was that a verb should first be described in terms of the semantic roles forming its meaning, and secondly, in terms of the rules needed to convert them into grammatically realized constituents, e.g. subjects, objects, and oblique phrases (Fillmore, Johnson, and Petruck 2003: 240). In subsequent contributions, Fillmore argued that a small fixed set of 'deep case' roles was not sufficient to describe the complementation properties of lexical items. As such, the frame elements currently used in Frame Semantics currently are frame-specific situational roles, rather than semantic roles as first proposed in Case Grammar (Fillmore 1968).

The basic premise of Frame Semantics is that word meanings should be described within the context of conceptual scenarios, referred to as *frames*. In a first approximation, Fillmore (1977: 63) described frames strictly from a linguistic perspective:

> A frame is any system of linguistic choices –the easiest being collections of words, but also including choices of grammatical rules or linguistic categories– that can get associated with prototypical instances of scenes.

As the theory shifted towards a more cognitive orientation, the notion of frame also evolved accordingly. Therefore, frames are currently understood as "schematic representations of the conceptual structures and patterns of beliefs, practices, institutions, images, etc. that provide a foundation for meaningful interaction in a given speech community" (Fillmore, Johnson, and Petruck 2003: 235). In this sense, regarding Frame Semantics, Fillmore, Johnson, and Petruck (2003: 241) affirm:

> The frame provides the conceptual underpinnings for related senses of a single word and semantically related words. With this approach to lexical semantic analysis and description, it is possible to characterize all categories of words, as well as phrases and expressions, using the same apparatus – the frame

The practical application of Frame Semantics is FrameNet. FrameNet[80] is a computational lexicography project developed at the International Computer Science Institute in Berkeley (California), which extracts information pertaining to the linked semantic and syntactic properties of words from large electronic text corpora, using both manual and automatic procedures. It was initially conceived as a frame-based lexicon aimed at covering the general vocabulary of English. The goal of the first phase of the project was to develop a set of tools for lexicon building. During the second phase, which began in 2000, work continued on the database. More specifically, word coverage was increased, and sample sentences were annotated. FrameNet was extended to different languages, and this gave rise to Chinese FrameNet (You and Liu 2005), FrameNet Brasil[81], German FrameNet[82], Japanese FrameNet[83], Swedish FrameNet[84],

[80] <https://framenet.icsi.berkeley.edu> [03/01/2012].
[81] <http://www.framenetbr.ufjf.br/> [03/01/2012].
[82] <http://www.laits.utexas.edu/gframenet/> [03/01/2012].
[83] <http://jfn.st.hc.keio.ac.jp/> [03/01/2012].

and Spanish FrameNet[85]. For example, Spanish FrameNet is freely available, and contains more than 1,000 lexical items in Spanish, which are representative of a wide range of semantic domains.

The basic units of lexical analysis in FrameNet are the following: (i) frame; (ii) lexical unit; (iii) valence. A *lexical unit* (LU) (Cruse 1986) can be defined as a pairing of a word with one of its senses. Accordingly, the separate senses of a word generally match different semantic frames (Ruppenhofer et al. 2006: 5). In our description, the same verb can also belong to different domains or subdomains in line with the Lexical Grammar Model (§3.1.2.1), and therefore, can also evoke different frames. The question of polysemy, however, has been easily solved in our analysis since only the verb sense related to the EXTREME EVENT in its sense of natural disaster was taken into account.

Generally speaking, *verb valency* or *valence* refers to the number of arguments activated by a verbal predicate. For example, the verb 'hit', in a normal active sentence, requires a subject that expresses the 'hitter', and a direct object that expresses the 'hit' entity (Fillmore, Johnson, and Petruck 2003: 235). In FrameNet, information about valence is specified in both syntactic and semantic terms (see Table 46). Syntactic valence information is usually specified in terms of complement type (e.g. noun phrase, prepositional phrase, etc.) and grammatical function (e.g. subject, object, etc.). Semantic valence properties describe the kinds of entities that can take part in a frame, i.e. *frame elements* (FEs). FEs resemble the argument variables in predicate-argument structure but have substantial differences since frames are much more complex than logical predicates. In other words, FEs are frame-specific participant roles.

For example, within the TRANSFER frame, the FEs are DONOR, THEME and RECIPIENT. Some of the verbs activated by the TRANSFER frame are 'give' or 'receive', which as shall be seen, differ in the perspectives that they impose on TRANSFER events, as can be observed in the syntactic expression of the FEs. In English, the predicate *give* can be found in two basic complementation patterns (Fillmore, Johnson, and Petruck 2003: 237):

(10)

a. The teacher **gave** the student a book.

b. The teacher **gave** a book to the student.

---

[84] <http://spraakbanken.gu.se/eng/swefn> [03/01/2012].
[85] <http://gemini.uab.es:9080/SFNsite> [02/01/2012].

In both sentences, the DONOR role is expressed by the noun phrase (NP) 'the teacher', and the THEME role is expressed by the NP 'a book'; the RECIPIENT by the NP 'the student' in the first sentence, and by the prepositional phrase (PP) 'to the student' in the second sentence. Grammatically, 'the teacher' is the subject of 'gave' in both sentences (in Frame Semantics subjects are generally referred to as *external arguments* (Ext)). In the first sentence, 'the student' is the grammatical object (Obj) of 'gave', and 'a book' is what is called an NP *complement* (Comp). In the second sentence, 'a book' is the Obj of 'gave', and 'to the student' is a PP Comp. These patterns of grammatical realization are referred to as *valence patterns*. Valence patterns are represented in FrameNet as triples of FE names, phrase types (PTs) and grammatical functions (GFs), as shown in Table 46 (Fillmore, Johnson, and Petruck 2003: 238):

give	FEs:	Donor	Theme	Recipient
	PTs:	NP	NP	NP
	GFs:	Ext	Comp	Obj
give	FEs:	Donor	Theme	Recipient
	PTs:	NP	NP	PP-to
	GFs:	Ext	Obj	Comp

Table 46. Valence pattern representation in FrameNet

In contrast, in the case of 'receive', the RECIPIENT is expressed as the external argument (Ext), and the DONOR optionally in the form a PP Comp headed by 'from'. In this sense, FrameNet authors claim that differences between the LUs in a frame provide evidence that general thematic roles are not the best way of describing LUs:

> Thematic roles are meant to capture regularities about *linking*, i.e. the relation between semantic participant roles and the grammatical functions of constituents that express them. A thematic role analysis might claim that the verbs *give* and *receive* have different sets of thematic roles: 'give' (Agent Theme Recipient) and 'receive' (Recipient Theme Source). This analysis preserves generalizations about linking – in particular, that Agents tend to be expressed as subjects (Exts), and that Recipients are more likely to be subjects than are Sources. However, in doing so it misses the meaning relation that is captured in the frame semantic analysis. In the Transfer frame, it is apparent from the description that the DONOR has properties of both an Agent and a Source, emphasized by the verbs *give* and *receive* respectively (Fillmore, Johnson, and Petruck 2003: 238–239).

This is the reason why the FEs used in Frame Semantics are designated in terms of frame-specific situational roles rather than general semantic roles. The fact that FrameNet defines each role as *frame-specific* means that the same role does not appear in different frames. However, since different annotators tend to assign a human-understandable name to each role in a rather systematic manner (Matsubayashi, Okazaki, and Tsujii 2009: 21), sometimes the same role name is shared in different frames, even though it corresponds to a different role.

In our opinion, these FEs, though informative, have certain disadvantages. On the one hand, the list of FEs is open-ended and unconstrained. In other words, each time a new frame is described, new FEs are created. On the other hand, the tagging of FEs relies entirely on the intuition of the analyzer. Bearing this in mind, for our analysis, we decided to characterize the various linguistic realizations of the same argument activated by each verb in terms of the kinds of generalization for a set of terms that are thought to have a similar semantic and syntactic behavior. They can be compared to the FEs as proposed by FrameNet, but with the difference that the set of categories is established beforehand and their boundaries are clearly stated. (See §5.4.2.1.1 for a description of how this semantic labeling was performed).

Regarding syntax, we used an adapted version of the valence pattern schema adopted in FrameNet, more specifically, by specifying for each category linked to a thematic role, and a macrorole, the phrase type associated with it (NP, PP, etc.). (See §5.4.2.1.1, for a more detailed description of phrase type assignment).

Because of the information contained, FrameNet is conceptually best described as a *lexical database* and an *annotation database* (Fillmore, Johnson, and Petruck 2003: 240). The *lexical database* contains the features necessary to characterize LUs. It provides information about frames, FEs, lemmas, lexemes, word forms, and their parts of speech. It also defines hierarchical relations between frames, i.e. frame-to-frame relations, namely, *inheritance*, *subframe*, *perspective_on*, *using*, *causative_of*, *inchoative_of*, *see_also*, and relations among roles or FE-to-FE relations. In this sense, it seemed that this hierarchy could serve as a starting point for generalizing roles in FrameNet, given that more specific roles should be able to inherit the properties of more general ones. Although the role hierarchy in individual frames was expected to help to generalize semantic roles, this does not seem to have happened (Baldewein et al. 2004). The generalization of semantic roles across frames has thus become a critical issue for FrameNet (Gildea and Jurafsky 2002; Shi and Mihalcea 2005; Giuglea and

Moschitti 2006). In this sense, an increasing number of studies have presented machine-learning approaches to generalize semantic roles across different frames (c.f. Matsubayashi, Okazaki, and Tsujii 2009).

The lexical database in FrameNet can be accessed either by frames, through the *Frame Index* and visualized by means of *Frame Reports* (see Figs. 43, 44 and 45) or by lexical units through the *Lexical Unit Index,* visualized in *Lexical Entry Reports* (see Fig. 46 and 47). Lexical entry reports, apart from being available through the lexical unit index, can also be accessed through the frame report to which they belong.

The annotation database stores the annotated sentences, along with the subcorpora from which sentences were selected for annotation. Sentences are annotated by using different colors to identify FEs, phrase types, and grammatical functions. The information contained in the annotation database can be found in the *Annotation Report* for each LU (see Fig. 48).

We will now give an example of the information in FrameNet for the IMPACT frame as instantiated in the lexical unit 'hit'. The reason for choosing this particular frame and lexical unit is because our research focuses on the analysis and description of verbs belonging to the EXTREME EVENT in its sense of natural disaster. In this regard, the IMPACT subframe is one of the most salient, and 'hit' is one of the most prototypical verbs within the subframe.

Accordingly, Figure 43 displays an extract of the IMPACT frame report in FrameNet, which provides the definition of the frame. This definition is the schematic representation of a situation type which includes the FEs specifically defined for the frame (Fillmore et al. 2003: 305). As shown in Figure 43, the core and non-core FEs are designated using a color code. Core FEs include obligatory objects and complements of the headword (see Figure 44), whereas non-core or periphery FEs refer to expressions, such as time, place or manner, modifications that could be added to almost any situation (see Figure 45). As shown in Figure 43, the description of the IMPACT[86] frame includes the core elements of IMPACTOR (in red), IMPACTEE (in dark blue) and IMPACTORS (light blue). Along with the description of the frame, an example sentence which instantiates the frame is also provided (Figure 43):

[86] All the screenshots of the IMPACT frame were extracted from:
<https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=frameIndex> [05/01/2012]

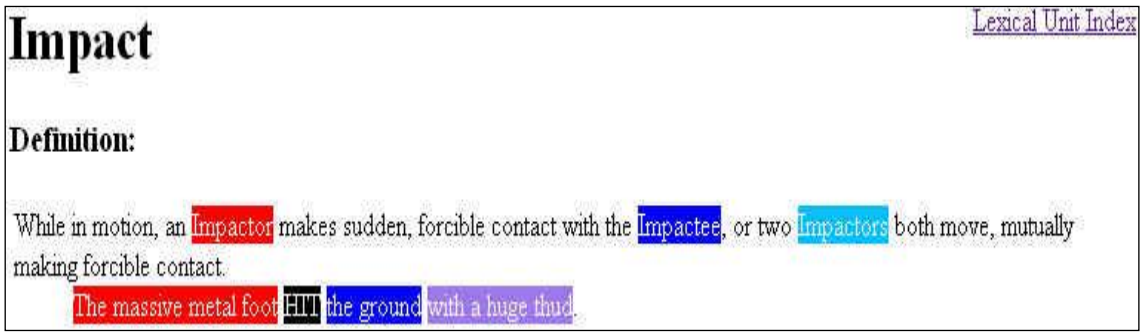


Figure 43. Extract of the IMPACT frame report. Definition

In each frame report, after the definition of the frame, the core and non-core elements are clearly specified by using different colors. A definition is given as well as a sentence in which the FEs in question are labelled. As shown in Figure 43, the core FEs in the IMPACT frame are the IMPACTEE (in dark blue), the IMPACTOR (in red) and the IMPACTORS (in light blue). This information is also highlighted in the definition (Figure 43).

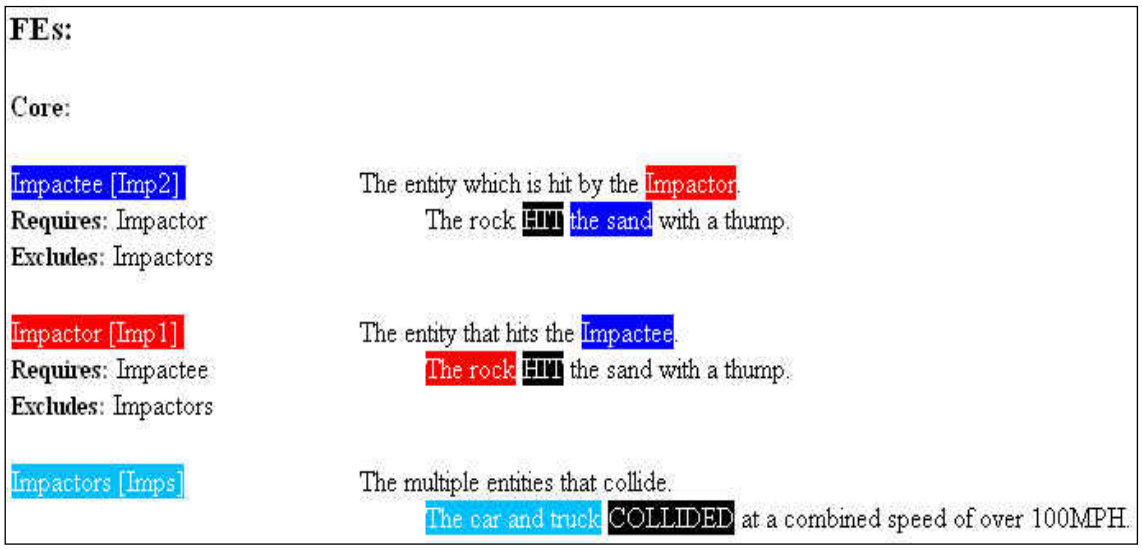


Figure 44. Extract of the IMPACT frame report. Core frame elements

The non-core elements in the IMPACT frame are CAUSE, DEPICTIVE, FORCE, MANNER, PERIOD_OF_ITERATIONS, PLACE, RESULT, SPEED, SUBREGION, TIME (Figure 45):

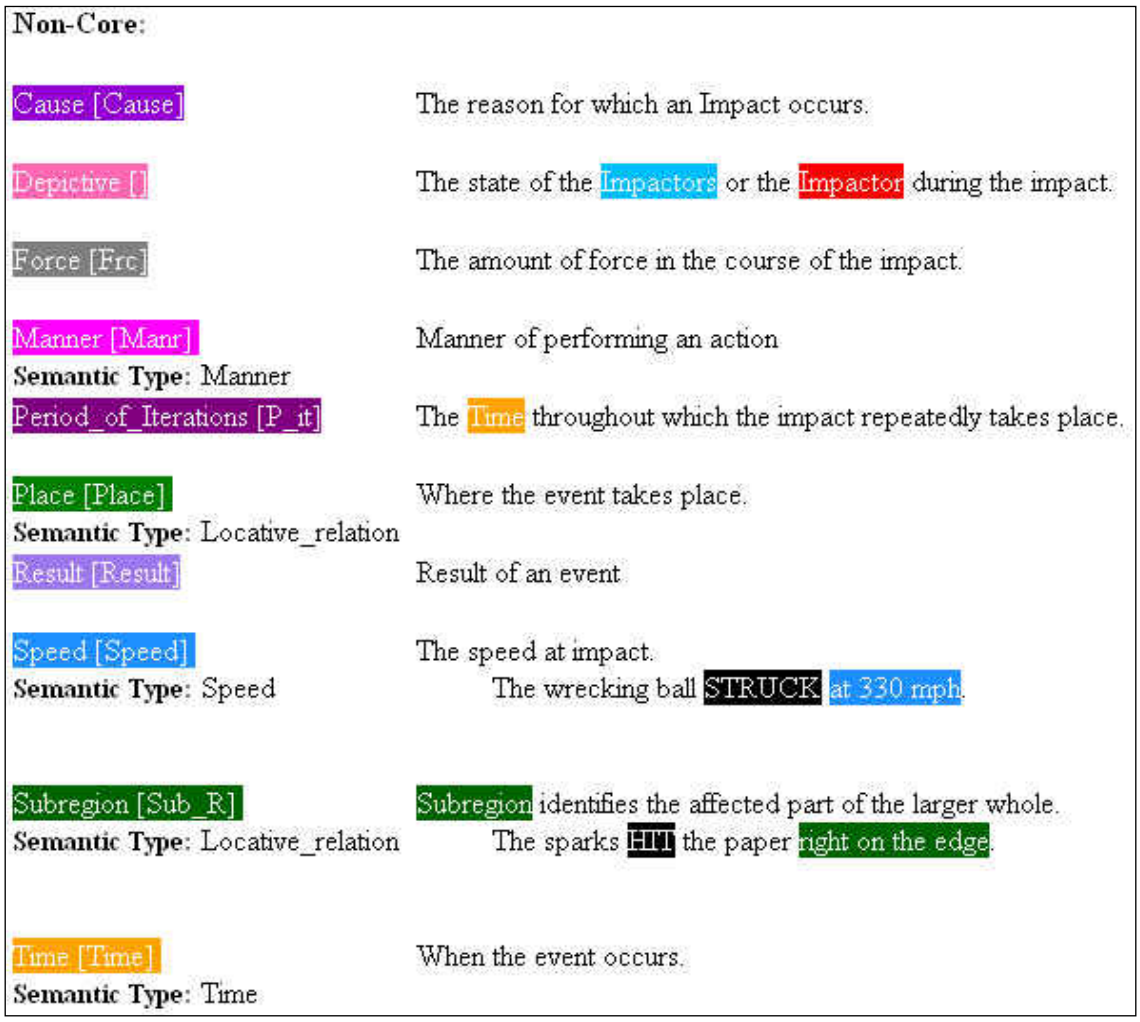


Figure 45. Extract of the IMPACT frame report. Non-core frame elements

Once the core and non-core elements within a specific frame are described, a frame report provides the frame relations. For instance, in this particular example, the frame report states that IMPACT *is_used_by* FRICTION and refers users to the frame of MAKE_NOISE. Along with the relations between frames, the frame report supplies the lexical units activated within the frame. As such, the IMPACT frame is evoked by English lexical units, such as the nouns 'collision', 'crash', 'hit', 'impact'; the verbs, 'bang', 'brush', 'bump', 'chatter', 'clang', 'clash', 'clatter', 'click', 'clink', 'clunk', 'collide', 'crash', 'crunch', 'graze', 'hiss', 'hit', 'impact', 'knock', 'patter', 'plash', 'plop', 'plough', 'plunk', 'rap', 'rattle', 'run', 'slam', 'slap', 'smack', 'smash', 'strike', 'thud', 'thump', 'tinkle', 'touch'; and the adjective, 'glancing'.

FrameNet provides a lexical entry report and an annotation report for each LU. It also includes an index that enables searches for lexical units. Figures 46 and 47 display extracts of the lexical entry report for the LU 'hit'[87] (verb):

hit.v

Frame: Impact

Definition:

COD: (of a moving object or body) come into contact with (someone or something stationary) quickly and forcefully.

Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

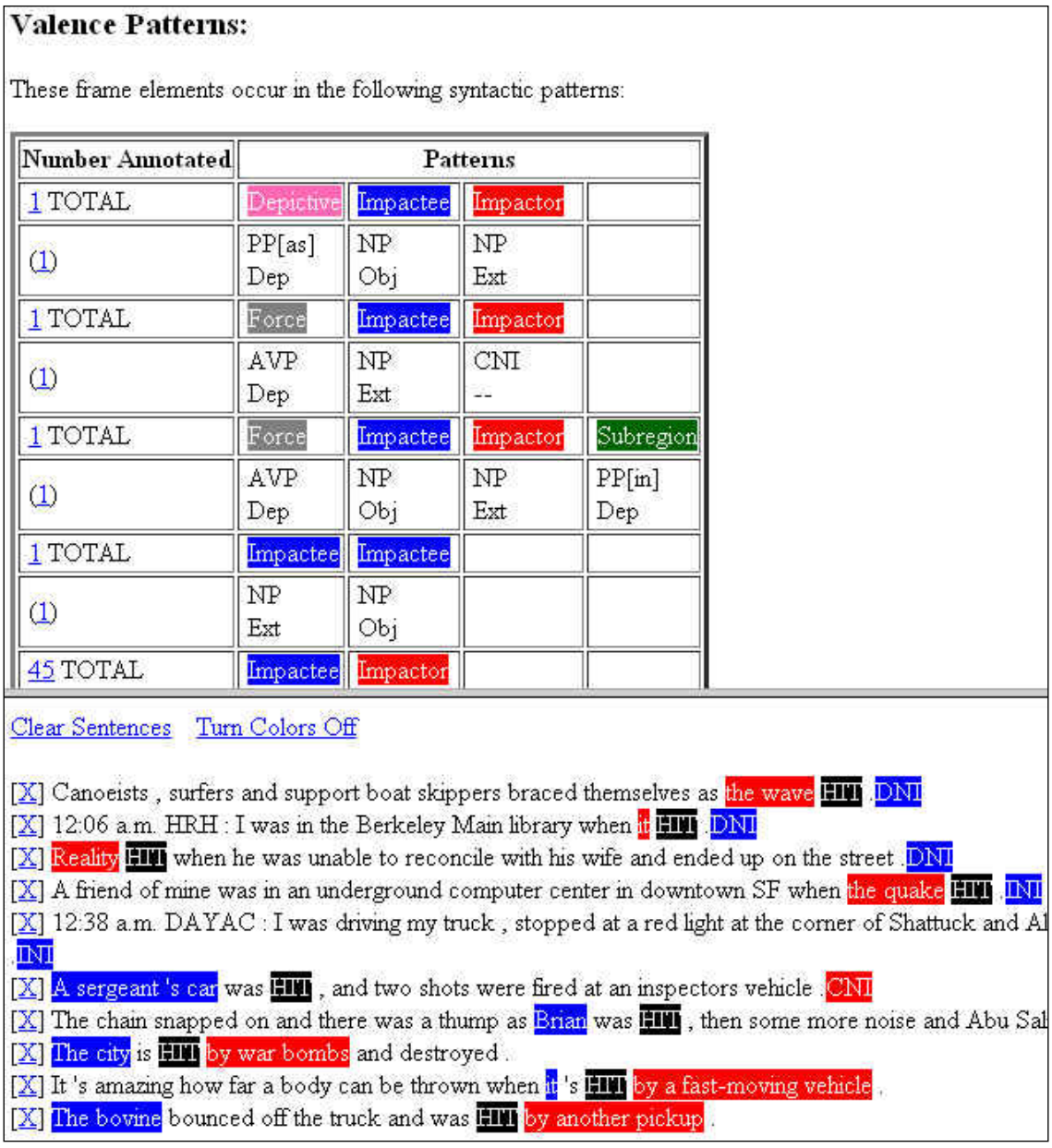
Frame Element	Number Annotated	Realization(s)
Depictive	(1)	PP[as].Dep (1)
Force	(2)	AVP.Dep (2)
Impactee	(67)	NP.Obj (48) NP.Ext (14) DNI.-- (4) INI.-- (2)
Impactor	(66)	NP.Ext (52) CNI.-- (3) PP[by].Dep (9) PPing[by].Dep (1) DNI.-- (1)
Manner	(2)	PP[like].Dep (1) PP[with].Dep (1)
Place	(4)	PP[at].Dep (1) PP[in].Dep (1) PP[from].Dep (1) PP[on].Dep (1)
Result	(4)	PP[in].Dep (2) PP[with].Dep (2)
Subregion	(5)	PP[in].Dep (5)
Time	(2)	PP[during].Dep (1) Sub.Dep (1)

Figure 46. Extract of the lexical entry report of 'hit' (verb)

Each lexical unit report first states the frame to which the lexical unit belongs (IMPACT frame), followed by the definition of the lexical unit and the frame elements activated along with their syntactic realizations. Finally, it gives the valence patterns (Figure 47, below). By clicking on the numbers shown in the columns *Number annotated* and *Realizations*, the user has access to the annotated sentences that

[87] <https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=frameIndex> [05/01/2012].

instantiate the pattern in question. Figure 47 shows annotated sentences that illustrate the valence patterns of 'hit'. As previously mentioned, valence patterns are grammatical realizations represented in FrameNet as *triples* of FE names, phrase types and grammatical functions. In this sense, the pattern formed by the FEs DEPICTIVE IMPACTEE IMPACTOR is realized by the phrase types PP (as) NP and NP and the grammatical functions of Dep, Obj and Ext. The pattern IMPACTEE + IMPACTOR is among the constructions with the most realizations (45 recorded annotated examples). By clicking on the number, the user can access the annotated sentences instantiating the pattern in question.



Valence Patterns:

These frame elements occur in the following syntactic patterns:

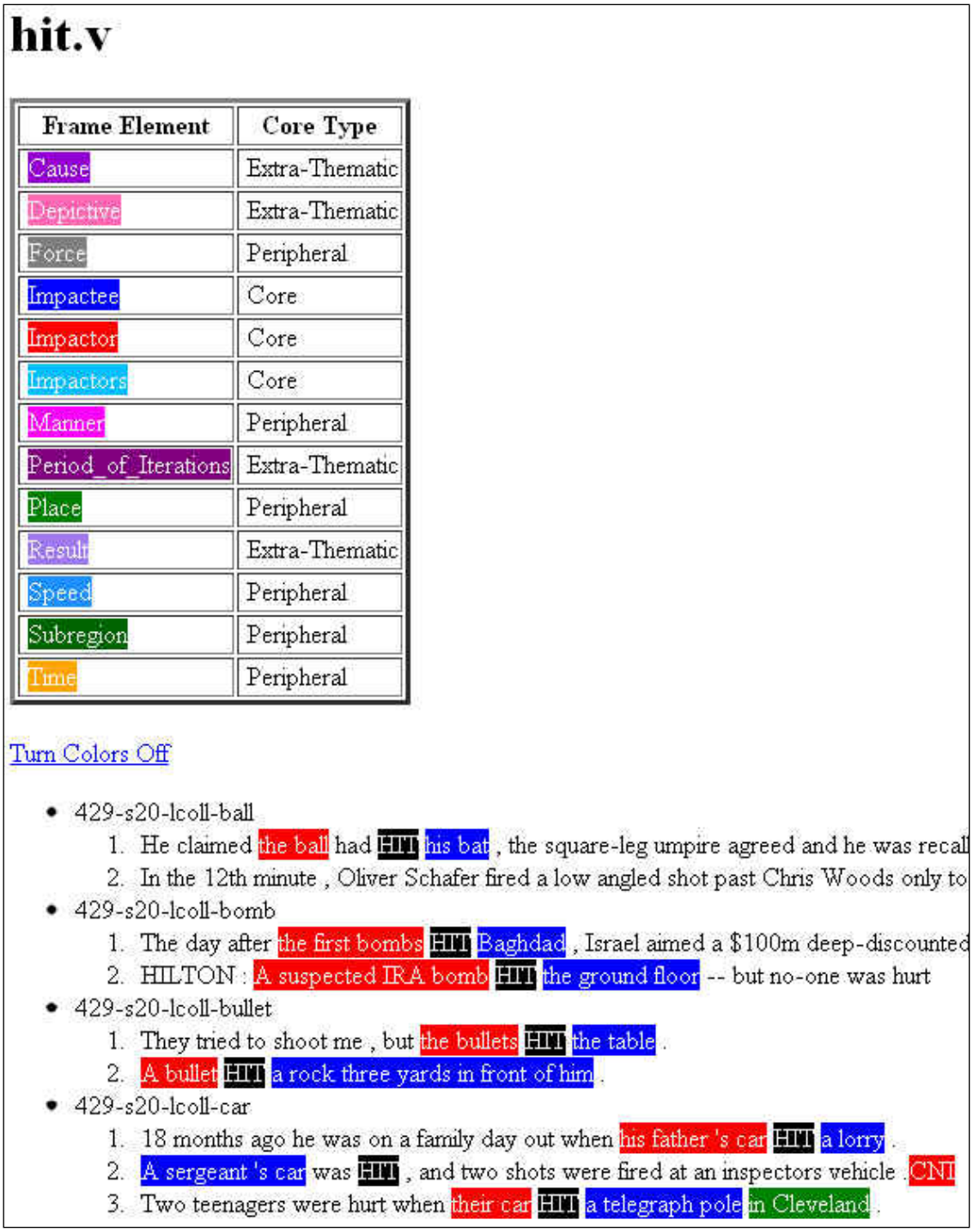
Number Annotated	Patterns			
1 TOTAL	Depictive	Impactee	Impactor	
(1)	PP[as] Dep	NP Obj	NP Ext	
1 TOTAL	Force	Impactee	Impactor	
(1)	AVP Dep	NP Ext	CNI --	
1 TOTAL	Force	Impactee	Impactor	Subregion
(1)	AVP Dep	NP Obj	NP Ext	PP[in] Dep
1 TOTAL	Impactee	Impactee		
(1)	NP Ext	NP Obj		
45 TOTAL	Impactee	Impactor		

Clear Sentences Turn Colors Off

[X] Canoeists , surfers and support boat skippers braced themselves as the wave HIT . DNI
[X] 12:06 a.m. HRH : I was in the Berkeley Main library when it HIT . DNI
[X] Reality HIT when he was unable to reconcile with his wife and ended up on the street . DNI
[X] A friend of mine was in an underground computer center in downtown SF when the quake HIT . DNI
[X] 12:38 a.m. DAYAC : I was driving my truck , stopped at a red light at the corner of Shattuck and Al . DNI
[X] A sergeant 's car was HIT , and two shots were fired at an inspectors vehicle . CNI
[X] The chain snapped on and there was a thump as Brian was HIT , then some more noise and Abu Sal
[X] The city is HIT by war bombs and destroyed .
[X] It 's amazing how far a body can be thrown when it 's HIT by a fast-moving vehicle .
[X] The bovine bounced off the truck and was HIT by another pickup .

Figure 47. Extract of the lexical entry report of 'hit' (verb)

The Annotation Report for the predicate 'hit'[88] is shown in Figure 48. It includes the list of the FEs and the selected sentences that are tagged with the appropriate FE labels:



hit.v

Frame Element	Core Type
Cause	Extra-Thematic
Depictive	Extra-Thematic
Force	Peripheral
Impactee	Core
Impactor	Core
Impactors	Core
Manner	Peripheral
Period_of_Iterations	Extra-Thematic
Place	Peripheral
Result	Extra-Thematic
Speed	Peripheral
Subregion	Peripheral
Time	Peripheral

Turn Colors Off

- 429-s20-lcoll-ball
  1. He claimed the ball had HIT his bat , the square-leg umpire agreed and he was recall
  2. In the 12th minute , Oliver Schafer fired a low angled shot past Chris Woods only to
- 429-s20-lcoll-bomb
  1. The day after the first bombs HIT Baghdad , Israel aimed a $100m deep-discounted
  2. HILTON : A suspected IRA bomb HIT the ground floor -- but no-one was hurt
- 429-s20-lcoll-bullet
  1. They tried to shoot me , but the bullets HIT the table .
  2. A bullet HIT a rock three yards in front of him .
- 429-s20-lcoll-car
  1. 18 months ago he was on a family day out when his father 's car HIT a lorry .
  2. A sergeant 's car was HIT , and two shots were fired at an inspectors vehicle . CNI
  3. Two teenagers were hurt when their car HIT a telegraph pole in Cleveland .

Figure 48. Extract of the annotation report of 'hit' (verb)

[88] <https://framenet.icsi.berkeley.edu/fndrupal/index.php?q=luIndex> [05/01/2012].

Finally, FrameNet, apart from permitting searches by means of a frame or lexical unit index, has recently developed a visualization tool called *FrameGrapher*[89] that shows the relations between frames and their FEs. FrameGrapher allows the user to choose a starting frame and systematically explore the connections between frames and their frame elements. Users can select the types of frame relation that they wish displayed and the number of links that are shown at one time. Clicking on a node expands the graph out from the new node. Figure 49 displays a screenshot showing the relation of the IMPACT frame with other related frames:

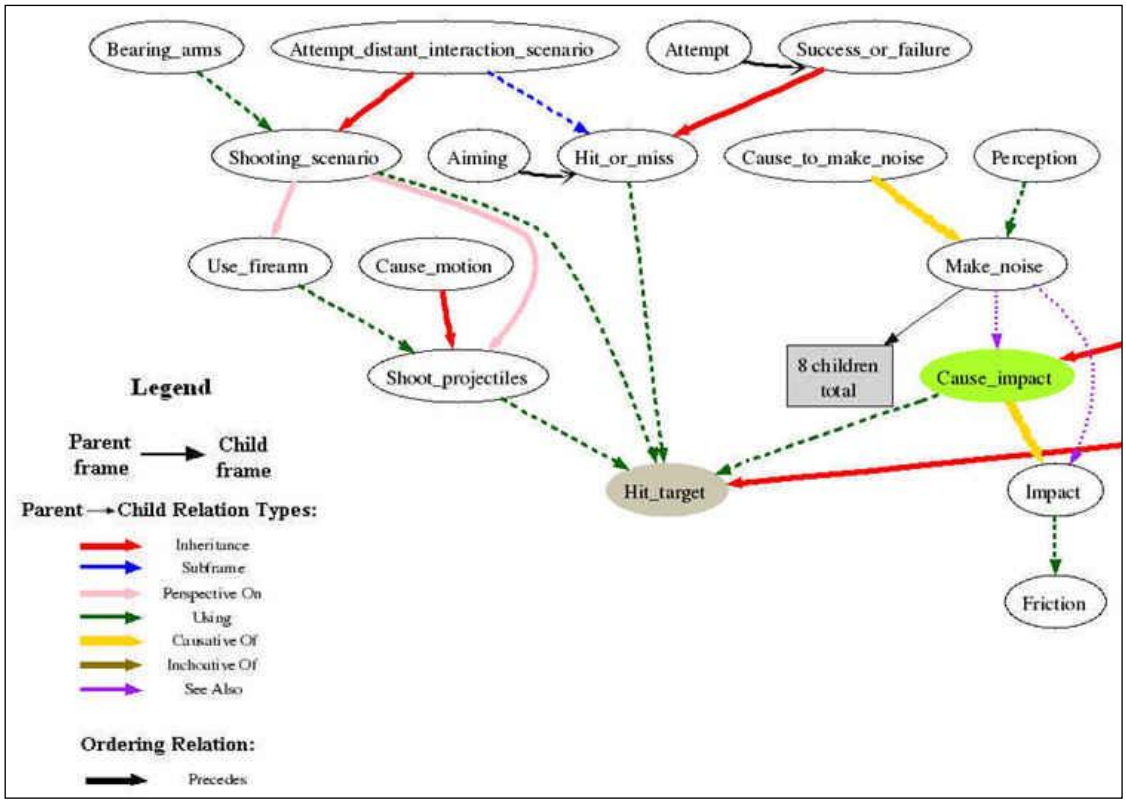


Figure 49. Relations of the CAUSE_IMPACT frame in FrameGrapher

Over the last few years, the theoretical and methodological premises of FrameNet have been extended to specialized language as reflected in the *Kicktionary*[90] (Schmidt 2009), a multilingual electronic dictionary of soccer terms, and *BioFrameNet* (Dolbey, Ellsworth, and Scheffczyk 2006), which focuses on the domain of molecular biology. As previously mentioned, the basic principles of FrameNet have also been partially incorporated in Frame-based Terminology (Faber 2009, 2011, 2012) (§1.4.3) and by extension, in its practical application, the environmental knowledge base *EcoLexicon*

[89] <https://framenet.icsi.berkeley.edu/fndrupal/FrameGrapher> [05/01/2012].
[90] <http://www.kicktionary.de/> [04/02/12].

(§1.4.3.2). Since this research focuses on enhancing the linguistic information in EcoLexicon with verb phrase collocations, it is based on basic FrameNet principles, especially its assertion that the understanding of word meaning is based on knowledge of the semantic frames or conceptual structures that underlie its usage.

The inventory of semantic roles in this thesis coincides with the most general roles in FrameNet (§5.4.2.1.2). However, instead of the FrameNet conception of roles as frame elements, we have opted for the thematic relations within RRG, and established a further categorization in terms of conceptual categories. However, the syntactic description in our study is similar to that in FrameNet in terms of phrase types, but not regarding grammatical functions.

As described throughout this section, FrameNet is a huge database that provides very valuable information for any linguist. However, apart from the open-ended number of FEs or roles, FrameNet also has other limitations as pointed out by Jiménez (2008), which are the following: (i) some frames provide all the senses for a predicate, such as *believe*, but others only include more peripheral ones, such as *antagonize*; (ii) there is no explanation of how lexical-semantic information can condition the syntactic realization of a predicate; (iii) there is no proposal of a metalanguage, which would enhance its consistency.

3.1.3.3 Construction Grammar

Construction Grammar is a theory closely related to Fillmore's Frame Semantics (§3.1.3.2). It was initially proposed by Fillmore (Fillmore, Kay, and O'Connor 1988; Fillmore and Kay 1993) and subsequently developed by Goldberg (1995, 2006). Croft (2001) also proposed an approach to Construction Grammar, known as *Radical Construction Grammar*.

Construction Grammar rejects a strict division between syntax and semantics. Even though lexical and syntactic constructions differ in internal complexity, both pair form and meaning. It also rejects a division between semantics and pragmatics. Its basic premise is that constructions are considered to be the basic units of language. *Constructions* are defined as form-meaning correspondences that exist independently of particular verbs. In other words, constructions are thought to carry meaning in themselves, independently of the words in the sentence (Goldberg 1995: 1). Formally speaking, a construction is described by Goldberg (1995: 4) as follows:

> C is a CONSTRUCTION iff$_{def}$ C is a form-meaning pair <F_i, S_i> such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions.

Therefore, in a constructional approach to argument structure, systematic differences in meaning between the same verb in different constructions are attributed directly to the particular constructions (Goldberg 1995: 4). Nevertheless, even though constructions are regarded as having meaning independently of verbs, Goldberg (1995: 1) highlights the fact that grammars do not work entirely top-down, with constructions simply imposing their meaning on verbs, but rather that the analysis must be both top-down and bottom-up:

> There is no question that a large amount of information is contributed by individual lexical items [...] An entirely lexically-based, or bottom-up, approach fails to account for the full range of English data. Particular semantic structures together with their associated formal expression must be recognized as constructions independent of the lexical items which instantiate them.

In other words, it is the combination of a specific verb in a specific construction what entails certain differences in the general meaning of the whole group. In this sense, Goldberg (1995: 49-50) states that constructions must specify in which ways verbs combine with them; and constructions need to be able to constrain the class of verbs that can be integrated with them and specify the way in which the event type designated by the verb is integrated into the event type designated by the construction.

Argument structure constructions are a special subclass of constructions that provide the basic means of clausal expression in a language. For English, Goldberg distinguishes five types of argument structure constructions, namely, ditransitive, caused-motion, resultative, intransitive motion, and conative. Table 47 summarizes the semantic structure of each construction as well as its syntactic realization, and gives an example.

1. Ditransitive	X causes Y to receive Z	Subj V Obj Obj_2 Pat faxed Bill the letter.
2. Caused motion	X causes Y to move Z	Subj V Obj Obl Pat sneezed the napkin off the table.
3. Resultative	X causes Y to become Z	Subj V Obj X_{comp} She kissed him unconscious.
4. Intrans. Motion	X moves Y	Subj V Obl The fly buzzed into the room.
5. Conative	X directs actions at Y	Subj V Obl_{at} Sam kicked at Bill.

Table 47. Typology of constructions within Construction Grammar (Goldberg 1995: 3)

As shown, each argument structure in Table 47 designates a general and abstract "semantic frame" (Goldberg 2010: 50): the transfer of something from one person to another (for the ditransitive); causation of something to a new location (for the caused-motion); causation of a state change (for the resultative); direct motion or action along a path (intransitive motion and conative, respectively). In this sense, argument structure constructions are considered to be generalizations over a large number of verbs that have a basic underlying frame (Goldberg 2010: 50):

> In fact, in the cases of argument structure constructions, the frames involved are basic and very commonly experienced; otherwise they would not be frequent enough to be considered argument structure constructions. Argument structure constructions are generalizations over multiple verbs; this insures that they are fairly general.

An example, which is in consonance with the verbs activated within the EXTREME EVENT frame, is 'hit'. As is well known, 'hit' is very frequently used with a causative interpretation. However, there is a general consensus that many verbs are not causative verbs in themselves (Fillmore 1971; Talmy 1976; Randall 1983; Goldberg 1995, *inter alia*). It is when the verbs are used in the causative motion construction that they become causative. This is the case of 'hit'. In this sense, in 'John hit the ball', no causation is assumed. However, when 'hit' is used in the causative motion construction, ('Joe hit the ball across the field'), it is interpreted causatively since Joe caused the ball to move across the field (Goldberg 1995: 165).

Goldberg distinguishes between roles associated with the verb and roles supplied by the construction. *Participant roles* are associated with a particular verb and *argument roles* with the construction. The reason of making this distinction is that verbs are associated with frame-specific roles in line with FrameNet (§3.1.3.2). For example, the

verb 'hand' is associated with three participant roles: HANDER, HANDEE and HANDED (Goldberg 1995: 51), whereas constructions are linked to more general roles, such as AGENT, CAUSE > RECIPIENT, EXPERIENCER > INSTRUMENT > PATIENT, THEME > LOCATION, SOURCE, GOAL (Goldberg 1995: 57). In other words, "participant roles are instances of the more general argument roles and capture specific selectional restrictions as well" (Goldberg 1995: 43).

It is important to emphasize here that participant roles and argument roles can be fused (Goldberg 1995: 50):

> If a verb is a member of a verb class that is conventionally associated with a construction, then the participant roles of the verb may be semantically fused with argument roles of the argument structure construction.

The fusion of participant and argument roles is determined by two principles (Goldberg 1995: 50):

- The *Semantic Coherence Principle*, by which only roles that are semantically compatible can be fused. In other words, two roles R1 and R2 will be semantically compatible if either R1 can be construed as an instance of R2, or R1 can be construed as an instance of R1.
- The *Correspondence Principle*, by which each lexically profiled participant role must be fused with a profiled argument role of the construction. When a verb has three profiled participant roles, then one of them may be fused with a nonprofiled argument role of a construction.

In the example just described concerning the construction with 'hand', the three profiled participants of 'hand' (HANDER, HANDEE, HANDED) can be put into a one-to-one correspondence with the profiled argument role of the ditransitive construction:

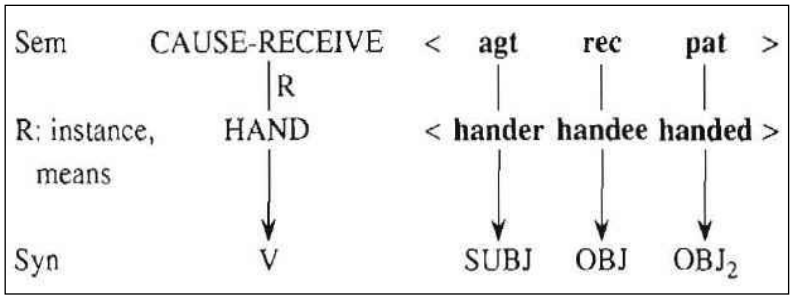


Figure 50. Role fusion in the ditransitive construction with 'hand' (Goldberg 1995: 51)

As previously mentioned, for our description of verbs within the EXTREME EVENT, we make use of what in Goldberg's terminology would be referred to as *argument roles*, i.e. the general roles of AGENT, PATIENT, INSTRUMENT, etc. For our purposes, general roles are better to describe our predicate-argument structures since they are generalizations that allow the user to predict relations between verbs and nominal forms. The reason for not including participant roles is that the list of participant roles proposed by Goldberg, in line with the frame elements in FrameNet (§3.1.3.2), is open-ended and relies exclusively on the intuition of the user. As previously mentioned, we propose an inventory of semantic categories which are similar to participant roles, but with the difference that they are established beforehand.

One of the main advantages of Construction Grammar is that it avoids implausible verb senses (Goldberg 1995: 9). For instance, in 'He sneezed the napkin off the table' (an example of the caused motion construction), 'sneeze' has three arguments despite the fact that it is usually an intransitive verb. To account for this phenomenon, a lexicosemantic theory would have argued that 'sneeze' has a three-argument sense. Goldberg, however, states that this is not the case. What happens is that in this example, *sneeze* appears in a three-argument construction, which signifies that the direct object is licensed by the construction, not by the verb. A constructional approach thus can describe situations in which a verb appears in an unusual construction. This eliminates the need to postulate unnecessary or improbable senses for the verb in question as an explanation of its atypical use.

Accordingly, Construction Grammar also avoids circularity derived from the traditional claim of defining verbs, based on the type and number of arguments they need (Goldberg 1995: 10-12). She argues against those claims that an ordinary verb such as 'kick' can be lexicalized in at least eight different argument structure situations. In Construction Grammar, constructions are associated with a family of closely related senses rather than a single, fixed abstract sense. In this sense, the different argument structures lexicalized by 'kick' can be seen as a case of constructional polysemy: the same form is paired with different but related senses:

> By accounting [...] in terms of constructional polysemy, as opposed to positing a collection of lexical rules, for example, we can capture the relations between the different senses in a natural way. In particular, a polysemous analysis allows us to recognize the special status of the central sense of the construction (Goldberg 1995: 33).

Finally, another advantage of Construction Grammar, according to Goldberg, is the preservation of compositionality. It is widely acknowledged that compositionality refers to the meaning of an expression in a language as the meaning of its immediate constituents and the syntactic rules that govern them:

> By recognizing the existence of contentful constructions, we can save compositionality in a weakened form: the meaning of an expression is the result of integrating the meanings of the lexical items into the meanings of constructions. In this way, we do not need to claim that the syntax and semantics of the clause is projected exclusively from the specifications of the main verb (Goldberg 1995: 16).

The contribution of Goldberg's Construction Grammar to linguistics has been enormous. Thanks to her work, no one questions the importance of constructions in any predicate-argument analysis. However, since Construction Grammar focuses on the description of constructions, it has not paid sufficient attention to the representation of verbs and to the kind of constraints that determine how the various constructions are fused with each specific verb. In line with this, Croft (2003: 49) carried out a comparative analysis between Levin's lexical rules (§3.1.1.3) and Goldberg's constructions and concluded that both perspectives can be complementary[91]:

> Verbs appear to change their meaning when put into particular constructions. For example, *bake* means 'bake *x* and give *x* to someone' in the ditransitive construction. Two alternative analyses have been proposed: to derive the meaning of *bake* in the ditransitive construction by a lexical rule, or to have the meaning of *bake* in the ditransitive construction be predictable from the semantics of the construction. An analysis of the behavior of different verb classes in the ditransitive construction indicates that both analyses are partly right.

Needless to say, any approach to argument structure should take into account the constructions that underlie their meaning. Accordingly, Lexical Constructional Grammar (§3.1.3.4), on which our theoretical premises are largely based, integrates premises of Construction Grammar. However, it differs from Construction Grammar in that it aims at describing both lexical and constructional semantics without prioritizing constructions as a better way to describe meaning. In our opinion, the best way of approaching argument-structure analysis is to consider both the meaning provided by the verb and that provided by the construction.

[91] For a complete comparison between lexical rules and construction, cf. Croft (2003)

3.1.3.4 Lexical Constructional Model

The Lexical Constructional Model, henceforth LCM, is a meaning construction model that aims to account for the relationship between syntax and all facets of meaning construction, including traditional implicature, illocutionary meaning, and discourse coherence (Ruiz de Mendoza and Mairal 2008: 355). Its founders identified the weaknesses of both functionalist and constructionist approaches, and realized that the best option was to develop a new model that would incorporate premises of both approaches:

> [...] functionalist approaches do not pay sufficient attention to the importance of constructions in determining morphosyntactic structure, while on the other hand constructionist approaches do not offer detailed accounts of the constraints on combining particular lexical entries with particular constructions (Butler 2009: 22).

In this sense, LCM integrates premises from functional models of language, especially Role and Reference Grammar (§3.1.2.2), and Cognitive Linguistics, mainly Goldberg's Construction Grammar (§3.1.3.3) and Lakoff's (1988) Cognitive Semantics. As shall be seen, it also includes insights into lexical decomposition from the Lexematic Functional Model and the Lexical Grammar Model (§3.1.2.1). In order to enrich semantics, it makes use of the role of Natural Semantic Metalanguage (Wierzbicka 1999; Goddard and Wierzbicka 2002, 2005, 2007) and the Meaning Text Theory (§2.1.3.1). Figure 51 shows the genesis of the LCM (Butler 2009: 3):

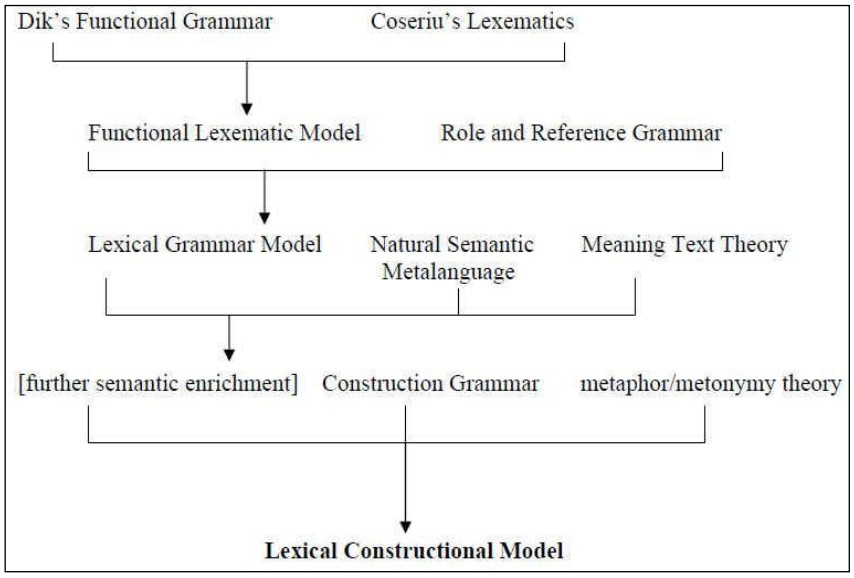


Figure 51. Genesis of LCM

The LCM states that a theory dealing with semantic interpretation should be constructed on the basis of the following: (i) two representational mechanisms, i.e. lexical (§3.1.3.4.1) and constructional templates; (ii) two basic cognitive operations, i.e. subsumption and conceptual cueing.

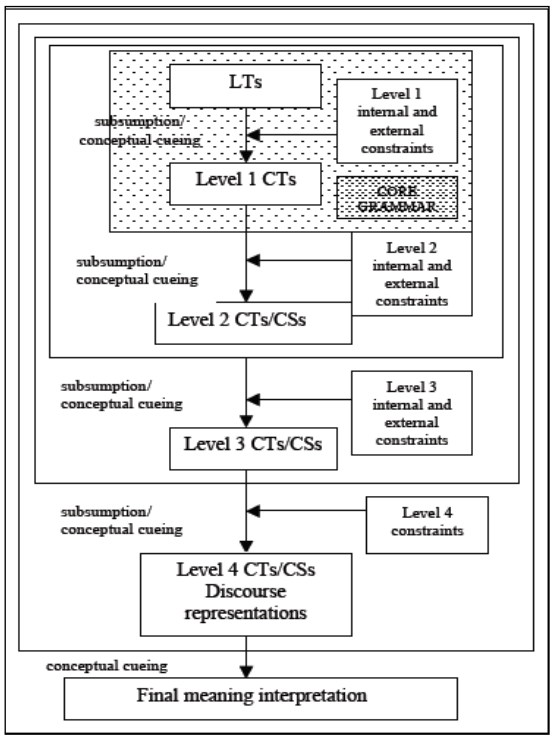


Figure 52. Architecture of the LCM (Ruiz de Mendoza and Mairal 2008: 358)

As shown in Figure 52, the model has four levels. The first level, also referred to as the *core grammar*, contains the notions of lexical and constructional template, more specifically it describes the subsumption of lexical templates into level 1 constructional templates. This level also accounts for some inferential activity, such as conceptual cueing. The level 2 module is a pragmatic module that deals with low-level inferential aspects of linguistic communication. Level 3 accounts for high-level inferences, such as illocutionary force. Finally, level 4 includes discourse aspects, focusing on cohesion and coherence and therefore, with implicit, logical, temporal and conceptual relations between propositions[92].

3.1.3.4.1 Lexical templates

[92] The non-core levels of the model (i.e. levels 2, 3 and 4) are still somewhat programmatic, as pointed out by Butler (2009: 33).

Lexical templates are formalized constructs that work as meta-entries and contain the semantic and syntactic relevant information of each predicate that forms a verbal or lexical domain. They are a central part in the LCM because the rest of the meaning levels (pragmatic, illocutionary and discursive meaning) are developed form lexical templates. Mairal and Faber (2007: 138) define a lexical template as a "formal representation of a lexical unit and the world knowledge elements which affect its syntactic representation". A lexical template contains three types of information: syntactic, semantic and pragmatic. The only compulsory information is syntactic unless the semantic component appears with the immediate hyperonym. In this case, the syntactic component is inherited from that of the hyperonym in question. In the same way, if the lexical template does not contain any semantic component, it will be inherited from the predicates used in the syntactic component. Lexical templates have the following format:

lexical template:
<pragmatic aspects> [SEMANTIC MODULE] [SYNTACTIC MODULE]

Table 48. Format of a lexical template

As shown (Table 48), semantic and syntactic components are delimited by square brackets and pragmatic aspects can come at the beginning of the lexical template between <> signs. To differentiate the semantic module from the syntactic one, the templates use two different variables: external and internal. External variables are those used in the logical structures in RRG (x, y, z) (§3.1.2.2) and they are always projected onto syntax. In contrast, internal variables appear in the semantic module as numbers (1, 2, and 3). The default correlation between the two variables is x = 1, y = 2, z = 3. Therefore, they are not explicitly included in lexical templates, unless they differ from this correlation.

Pragmatic information can add the following functions to a template (Martín-Mingorance 1998: 96–97): (i) informative, where lexemes are marked according to the type of discourse in which they appear (technical, legal, religious); (ii) psychological, where subjective elements are expressed, such as ironic, appreciative or pejorative elements; (iii) social, referring to the circumstances of the communicative exchange, such as geographical areas, social norms. These features are displayed in small letters at the beginning of the lexical template. When more than one feature exists, they are

separated by commas. Pragmatic features accepted within the Lexical Constructional Model are the following (Bernardos, Jiménez, and Pérez 2011: 55–56):

Pragmatic features	
Formal	fml
Informal	infml
Colloquial	col
Old/obsolete	old/ant
Literary	lit
Emphatic	emph/enf
Philosophical	ph/fil
Humorous	hum
Technical	tech/téc

Table 49. Inventory of pragmatic features in the LCM

The semantic module contains the semantic parameters that underlie predicate meaning. Consequently, they differentiate each of the various predicates that form a same lexical class. To this end, the semantic module makes use of an inventory of semantic primes and lexical functions. Semantic primes or undefinable items, correspond to the archilexemes or generic terms identified through lexical factorization in the Lexical Grammar Model (§3.1.2.1)[93], a model partially integrated into the LCM, especially those aspects concerning the paradigmatic and syntagmatic organization of predicates in lexical domains. These primes combine to express the more specific hyponyms of the same lexical domain with a catalogue of operators based on the notion of *lexical function* as proposed in Mel'čuk's Explanatory and Combinatorial Lexicology (§2.1.3.1). More specifically, the semantic portion of a lexical template is composed of the following: (i) a hyperonym; (ii) lexical functions; (iii) selection restrictions.

The *hyperonym* is a predicate that denotes the lexical class of the element being described and from which it inherits its lexical template. It is represented by the most immediate superordinate term, i.e. the semantic prime in question or a predicate already defined to which features have been added by means of one or several functions. All of this information is provided at the beginning of the semantic component within each lexical template. When more specifications are needed, they are provided immediately afterwards, headed by function "&". Very occasionally, there are predicates with more than one hyperonym, or predicates that do not have an explicit hyperonym.

[93] These primes are mainly based on Natural Semantic Metalanguage (Wierzbicka 1999; Goddard and Wierzbicka 2002, 2005, 2007).

LCM has adapted Mel'čuk's lexical functions to account for lexical domain-specific relations and to encode those semantic parameters that are not evident in syntax, e.g. manner, purpose, means, social status, speaker's attitude, urgency of a request, *inter alia* (c.f. Faber and Mairal 1999; Mairal and Faber 2007). It was thus necessary to add new lexical functions to describe certain types of predicates, such as *verbs of feeling* (Jiménez 2008). This is the reason why in the LCM, lexical functions are referred to as *semantic functions.* To date, the semantic functions accepted in the LCM are displayed in Table 50 (adapted from Bernardos, Jiménez, and Pérez 2011: 58).

As shown (Table 50, following page), each function is displayed in capital letters and must always take either an argument or feature. These arguments or features can be: (i) internal variables, which as already mentioned, use numerical subindices (1, 2, 3); (ii) other functions; (iii) some of the elements that can appear as hyperonyms.

A lexical template can contain several independent functions which are separated with "&" or "/". Functions are interpreted from left to right, based on their position within the lexical template, taking into account the following (Bernardos, Jiménez, and Pérez 2011: 58-59):

- If it is a binary function and has a numerical subscript, the numerical subscript is interpreted as one of its arguments. The second argument in the lexical template is what is to the right of the function. When the binary function does not contain any subscript, the arguments of the function are everything that appears both to the right and to the left of the function.
- If it is a tertiary function and there are two numerical subscripts separated by a comma, those subscripts are their arguments and the third argument is the rest of the information that appears to the right of the function in question.
- When the function is unitary, everything to the right of the function is its only argument.
- Whenever this order must be altered, brackets are used to group the elements as desired. If a function, for instance, has different subscripts as arguments, these subscripts are separated by commas (,). However, it is also possible to combine different subscripts by means of "&" or "/" in order to form one argument of the function.

Unitary functions		**Binary functions**	
Name	Description	Name	Description
ANTI	Antonym/negation	ABLE	Ability to
CONT	Continuity/duration	CAUS	Cause
CULM	The highest point of []	BECAUSE	Reason
DEGRAD	To get worse	INSTR	Instrument/ by means of
FIGUR	Figuratively, metaphorically	INVOLV	Subactivities implied by the predicate
INCEP	The beginning of []	RESULT	The sub-activity is a direct, non-cancellable, result of the main predicate
LOC	Temporal location	LOC_{in}	Spatial location with directionality "in"
MAGN	Intense(ly), very [intensifier], to a very high degree	LOC_{ad}	Spatial location with directionality "to"
MINUS	Less of []	LOC_{tf}	Temporal location future
OBSTR	To function with difficulty	LOC_{tp}	Temporal location past
PLUS	More of	MANIF	Showing in appearance
PROPT	Because of	MANNER	Manner
Tertiary functions (in study)		PROP	Propositional argument
CONV	Conversives of the predicate	LOC	Temporal location present
		PURP	Purpose
		SYMPT	Physical symptoms
		&	And
		/	Or
		=	Equal

Table 50. Inventory of semantic functions in the LCM

As an example, the lexical template of 'procure' is shown below (Bernardos, Jiménez, and Pérez 2011: 59):

(10)

procure: <fml> OBSTR [*obtain*]

Example: 'He therefore persuaded a friend to procure him a ticket without disclosing his identity'

As previously mentioned, lexical class inheritance presupposes that hyponyms inherit the features of their more immediate hyperonym. In the lexical template in (10) only the hyperonym has been specified (obtain) along with the formal pragmatic feature that distinguishes this verb from others in the same lexical class (<fml>). The function OBSTR is included to express the difficulty involved in fulfilling the event.

The syntactic module corresponds to RRG logical structures (§3.1.2.2.2). As previously mentioned, in RRG verbs are classified as states, activities, achievements, accomplishments, semelfactives and active accomplishments, as well as their respective causative variant. States and activities are primitives, whereas accomplishments and achievements consist of either a state or activity plus a BECOME, an INGR, or a SEML operator. However, in order to obtained a better description of verbal predicates, in the LCM the terms in boldface are not those used in RRG, but rather belong to the previously described inventory of primes (Ruiz de Mendoza and Mairal 2008: 363).

Therefore, the syntactic module of a lexical template contains: (i) information concerning the argumental structure (compulsory number of participants or external variables); (ii) the verbal class or lexical inheritance; and (iii) the interface with syntax, i.e. how external variables are projected into syntax. The verbal class or lexical inheritance is represented by the predicate in a given language that corresponds to a semantic prime (written in bold lower case letters and with an apostrophe at the end) or cuasi-prime, following the format of RRG logical structures, but in italics. In the words of Ruiz de Mendoza and Mairal (2008: 362):

> Lexical templates provide enhanced semantic representation and consequently allow us to account for those properties which go beyond those aspects of the meaning of a word that are grammatically relevant.

In this sense, lexical templates can be considered as lower-level constructional characterizations that share crucial features with higher-level

representations referred to as *constructional templates* (Ruiz de Mendoza and Mairal 2008: 362).

The LCM is currently considered one of the most coherent approaches to predicate-argument structure. However, as highlighted by Butler (2009: 138):

> The LCM, for all its history in a diverse range of functional, cognitive and/or constructionist approaches, is still very much in its infancy, and there remain a number of questions to be answered, and challenges to be faced.

More specifically, LCM faces the need to provide the model with a syntax and with clear mechanisms for linking the syntax to its rich semantic description (Butler 2009: 145-146). As stated throughout this chapter, our methodology approach is based on the premises of LCM because it is a linguistic model capable of providing a formalization of meaning that can be used in NLP. Although the LCM is still being developed, the results so far are very promising. In our opinion, it will eventually become a reference point for models of predicate argument structure. It is true that we have not specified lexical templates in this thesis, but we have implemented definition rules that could become the basis for lexical templates associated with specialized subdomains, and which could be used to formalize this area of specialized knowledge.

3.2 Argument Structure as codified in Lexical Resources

After an overview of some of the main linguistic theories that deal with argument structure, we now look at some of the most important lexical resources that have verb entries. This section thus describes the English resources, WordNet (§3.2.1), VerbNet (§3.2.2), PropBank (§3.2.3) and the Spanish resources, ADESSE (§3.2.4) and SenSem (§3.2.5), and explains which aspects of these resources are interesting for our verb entry proposal.

PropBank, VerbNet, WordNet, and FrameNet (§3.1.3.2) are currently being integrated into a project referred to as *SemLink*[94], which is in the process of linking different meaning-based resources by mappings in order to enhance inferencing as well as semi-automatic extraction tasks, among others. The current unified version of SemLink can either be downloaded from its website or accessed online by means of a

[94] < http://verbs.colorado.edu/semlink/> [11/01/2012].

unified verb index[95]. However, this section presents each resource separately in order to better discuss how each resource approaches verb description.

3.2.1 WordNet

WordNet is an online lexical database for English that can be either browsed online[96], or freely downloaded. It was conceived by George Miller in 1986, at the Cognitive Science Laboratory of Princeton University, where it continues to be developed. Miller (1986) attempted to reflect the hierarchical Network Model of Semantic Memory by Collins and Quillian (1969) into a lexicographic model of lexical organization, i.e. conceptually based, rather than alphabetically.

WordNet groups English words into sets of synonyms referred to as *synsets*. Each member of a given synset represents the same concept, though not all synset members are interchangeable in all contexts (e.g. 'car', 'automobile'; 'hit', 'strike'; 'big', 'large') (Fellbaum 2006: 665). In this sense, for WordNet, two words that share at least one sense are synonymous, whereas a word having more than one sense appears in different synsets, reflecting the word's polysemy.

WordNet describes four syntactic categories of words, namely, nouns, verbs, adjectives, and adverbs. According its statistics[97], WordNet's latest version (WordNet 3.1) contains 155,287 words organized in 117,659 synsets, comprising more than 82,115 noun synsets, 13,767 verb synsets, 18,156 adjective synsets, and 3,621 adverb synsets. Each synset further contains a short defining *gloss*, which is a brief definition, and/or one or more example sentences illustrating the usage of the synonyms. Synsets are linked to other synsets by means of different semantic relations. Along with semantic relations, which describe relations between synsets, WordNet also describes lexical relations, i.e. the relations between words (e.g. antonyms).

Verbs are organized in hierarchies by means of *lexical entailment relations* (Fellbaum and Miller 1990; Fellbaum 1998). Lexical entailment relations between verbs are established when a verb V_1 entails V_2, i.e. if V_2 cannot be done unless V_1 is or has been done. For example, 'snore' lexically implies 'sleep' since the sentence 'he is snoring' implies that 'he is sleeping' (Moreno 1998). In this sense, verbs in WordNet are organized by several entailment relations, such as troponymy, which describes a

[95] <http://verbs.colorado.edu/verb-index/> [11/01/2012].
[96] <http://wordnet.princeton.edu/wordnet/> [11/01/2012].
[97] <http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html#toc2> [11/01/2012].

manner of doing something, i.e. V_1 is to V_2 in some particular manner (*march-walk*), backward entailment (*divorce-marry*); presupposition (*buy-pay*), or cause relations (*show-see*) (Fellbaum 2006: 667). The most prominent lexical entailment relation is that of troponymy, which could be regarded as a parallel relation to that of hyponymy used for describing nouns. In this sense, troponymy builds hierarchies of several levels of specificity for verbs, in basically the same way as hyponymy does for nouns.

In this regard, WordNet differentiates 15 base types for verbs, namely, BODILY CARE, CHANGE, COGNITION, COMMUNICATION, COMPETITION, CONSUMPTION, CONTACT, CREATION, EMOTION, MOTION, PERCEPTION, POSSESSION, SOCIAL INTERACTION, STATIVE and WEATHER. These base types could be regarded as general categories similar to the twelve lexical domains proposed by the Lexical Grammar Model (§3.1.2.1). It is noteworthy that the four types of lexical entailment relations are more usefuly applied to certain kinds of verbs than to others. For example, troponymy relations are mainly used within verbs of CREATION, COMMUNICATION, COMPETITION, MOVEMENT and CONSUMPTION; backward entailment is useful to organize verbs of STATE and verbs denoting CHANGE; whereas cause relations are particularly found within verbs of MOVEMENT (Moreno 1998).

For example, the entry for ‘hit’ in WordNet shows seven senses for ‘hit’ as a noun, and 16 for it as a verb. These senses are displayed in order of frequency (Figure 53):

Noun

- S: (n) **hit** *"he came all the way around on Williams' hit"*
- S: (n) **hit**, hitting, striking *"repeated hitting raised a large bruise"; "after three misses she finally got a hit"*
- S: (n) **hit**, smash, smasher, strike, bang *"that song was his first hit and marked the beginning of his career"; "that new Broadway show is a real smasher"; "the party went with a bang"*
- S: (n) collision, **hit** *"the collision of the particles resulted in an exchange of energy and a change of direction"*
- S: (n) **hit**
- S: (n) **hit** *"it has all the earmarks of a Mafia hit"*
- S: (n) **hit** *"WordNet gets many hits from users worldwide"*

Verb

- S: (v) **hit** *"hit a ball"*
- S: (v) **hit**, strike, impinge on, run into, collide with *"The car hit a tree"; "He struck the table with his elbow"*
- S: (v) **hit** *"He hit her hard in the face"*
- S: (v) reach, make, attain, **hit**, arrive at, gain *"We hit Detroit by noon"; "The water reached the doorstep"; "We barely made it to the finish line"; "I have to hit the MAC machine before the weekend starts"*
- S: (v) **hit**, strike *"We were hit by really bad weather"; "He was stricken with cancer when he was still a teenager"; "The earthquake struck at midnight"*
- S: (v) shoot, **hit**, pip
- S: (v) stumble, **hit** *"I stumbled across a long-lost cousin last night in a restaurant"*
- S: (v) score, **hit**, tally, rack up *"The home team scored many times"; "He hit a home run"; "He hit .300 in the past season"*
- S: (v) **hit**, strike, come to *"Panic struck me"; "An interesting idea hit her"; "A thought came to me"; "The thought struck terror in our minds"; "They were struck with fear"*

Figure 53. Extract from the entry for 'hit' in WordNet

As shown, in the first sense for nouns, the synset is only composed of the search word 'hit', whereas in the third sense, the synset is formed by five words, namely, the search word, along with 'smash', 'smasher', 'strike' and 'bang'. In regards to verbs, the senses that are most relevant for our purposes are the following: (i) the second sense, formed by the verbs 'hit', 'strike', 'impinge on', 'run into', and 'collide with'; (ii) the fifth sense, formed by 'hit' and 'strike'.

When users click on label 'S' (in blue, before each sense), they access the synset semantic relations associated with the sense in question. Figure 54 displays the second sense relation of the verb 'to hit':

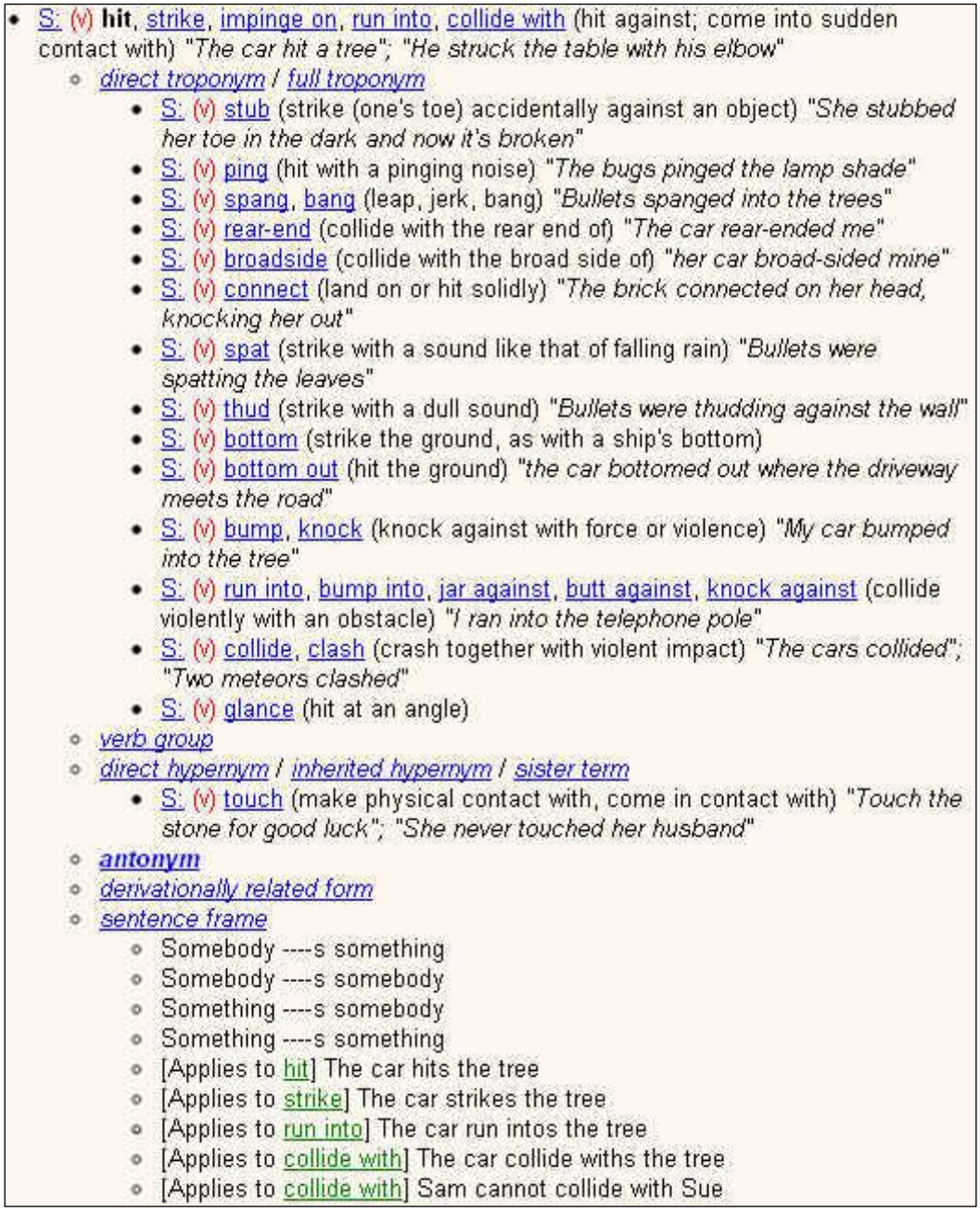


Figure 54. Extract from the entry for 'hit' within WordNet (second sense)

As shown, the second sense synset is composed of five words, namely, the verb 'hit' along with 'strike', 'impinge on', 'run into' and 'collide with'. The gloss is provided subsequently, which corresponds to the definition. Usage examples are also included ('hit against'; 'come into sudden contact with', 'the car hit a tree'; 'he stuck the table with his elbow'), along with the various relations[98].

[98] Since the figure was too large to fit on one page, only the main relations (i.e. troponym and sentence frame) are completely displayed.

As can be observed, WordNet first shows the troponym hierarchy for 'hit': ('stub'; 'ping'; 'spang'; 'bang'; 'rear-end'; 'broadside'; 'connect'; 'spat'; 'thud'; 'bottom'; 'bottom out'; 'bump', 'know'; 'run into', 'bump into'; 'jar against', 'butt against', 'knock against'; 'collide', 'clash'; 'glance'), followed by the verb group, the direct hyperonym ('touch'), the inherited hyperonym and the sister term (i.e. words with the same immediate hyperonym). Also included is the antonym relation; a relation labeled as *derivationally related forms,* which refers to terms in different syntactic categories that have the same root form and are semantically related; and the *sentence frame,* which describes the arguments of the verbs and which can be also interesting from a phraseological point of view since it provides the verbs together with the prepositions or phrases with which they collocate (e.g. 'run into', 'collide with').

It is our assertion that providing synonyms in a resource is very useful for text generation activities such as translation. However, offering verb synonyms in synsets has certain limitations since the user has no means of discriminating to what extent the verbs in the same synset are totally interchangeable in all contexts. For our description of verbs within the EXTREME EVENT, we use troponym relations as proposed by WordNet. More specifically, our verbs are hierarchically arranged by means of troponymy relations in lexical domains and subdomains in line with the Lexical Grammar Model (§3.1.2.1). Each verb has a *genus* that corresponds to the superordinate term in the hierarchy. The more specific the hierarchy becomes, the more focalized the information will be. In this sense, verbs within the same subdomain or domain can be considered to be synonymous to a certain degree, but the final user will always know how the verb in question relates to the main hyperonym or the verb heading the domain. In addition, we provide the user with the definition and specific contexts of each verb in the hierarchy (by means of clicking on each verb). Another striking disadvantage is the fact that WordNet does not provide any description of the potential arguments that can be used with each verb.

Other limitations described by Fellbaum (2006: 668) are the following: (i) it does not contain any syntactic information, when as described by Levin (Levin 1993) studies have provided evidence that verbs must be defined on the basis of their semantic and syntactic behavior; (ii) it does not describe syntagmatic relations; (iii) semantic roles of nouns functioning as arguments of specific verbs are not encoded; (iv) although the information in WordNet is currently based on web information, most of the data are not

corpus-based, but rather based on the creator's intuition since in 1986, digital corpora were not available.

Despite these limitations, WordNet is a huge database that has proved to be a very useful tool for improving recall, inferencing and enhancing the expression in language generation and translation by means of providing alternative words and paraphrases. In this sense, WordNet has been used in a wide range of NLP applications, such as word sense disambiguation and information retrieval. In addition, the information contained in WordNet along with its structure has been reused in several projects to build ontologies, such as the ontology SUMO[99]. However, it describes general language and does not cover specialized language. This makes it unsuitable as a database for NLP applications in specific domains that are not covered by WordNet.

Since 1990, the methodology applied to the English WordNet has been extended to other languages. Accordingly, the EuroWordNet project[100] was developed. EuroWordnet (Vossen 1998), funded by the European Commission, is a multilingual database with wordnets for several European languages, such as Dutch, Italian, Spanish, German, French, Czech or Estonian[101]. Along with European language wordnets, wordnets for other languages have also been developed. In this scenario, the Global WordNet Association[102] was created, whose principal aim is to coordinate, share and connect wordnets for all languages in the world. Currently, there are wordnets in more than 50 languages, including Arabic, Bantu, Basque, Chinese, Bulgarian, Estonian, Hebrew, Icelandic, Japanese, Kannada, Korean, Latvian, Nepali, Persian, Romanian, Sanskrit, Tamil, Thai, Turkish, and Zulu.

3.2.2 VerbNet

VerbNet[103] (Kipper 2005) can be regarded as the largest on-line verb lexicon currently available for English. VerbNet is organized into verb classes derived from Levin's classes (§3.1.1.3), but has refined them and incorporated new classes and subclasses in order to account for syntactic and semantic coherence between the different members of a class. Accordingly, VerbNet integrated 57 new classes from Korhonen and Briscoe (2004) and 53 new classes from Korhonen and Ryant (2005). These new incorporations

[99] < http://www.ontologyportal.org/> [17/01/2011].
[100] <http://www.illc.uva.nl/EuroWordNet/#EuroWordnet> [13/01/2012].
[101] Unfortunately, these wordnets are not freely available.
[102] <http://www.globalwordnet.org/> [13/01/2012].
[103] <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> [13/03/2012].

helped to overcome a fundamental limitation of Levin's verb classification, especially the fact that Levin focuses on the description of alternations involving NP and PP complements, but does not provide a complete description of adjective, adverbial, predicative, control and sentential complements, thus excluding many verbs that are highly frequent in English (Kipper et al. 2006: 1028). In this sense, the current version of VerbNet contains 274 first-level classes, further divided in subclasses, 3,769 lemmas, and 5,257 verb senses. In addition, VerbNet provides mappings between its verbs and WordNet senses (§3.2.1), and between its verbs and FrameNet frames (§3.1.3.2). Its syntactic coverage was mapped onto the Proposition Bank (§3.2.3) and VerbNet syntactic frames were able to account for over 84% exact matches to the frames found in PropBank (Kipper 2005: 4).

In each class, three main sections are distinguished: (i) *members*, which lists the verbs that belong to the class; (ii) *thematic roles*, which contains the roles associated with the arguments in the class, along with the selection restrictions on the arguments; (iii) *frame*, where a fine-grained description of the various frames realized by the verbs belonging to the class are provided. When needed, another section is added for *subclasses*, which follows exactly the same organizational configuration for classes.

In line with the examples offered in the description of the rest of theoretical models and resources in this chapter, we describe the information contained in a VerbNet entry for the HIT verbal class. Figures 55 and 56 display two extracts corresponding to the entry for the HIT class[104]. Figure 55 accounts for the members and thematic role description, whereas Figure 56 focuses on the frame counterpart analysis.

[104] For layout reasons, it is not possible to offer the two extracts in one caption.

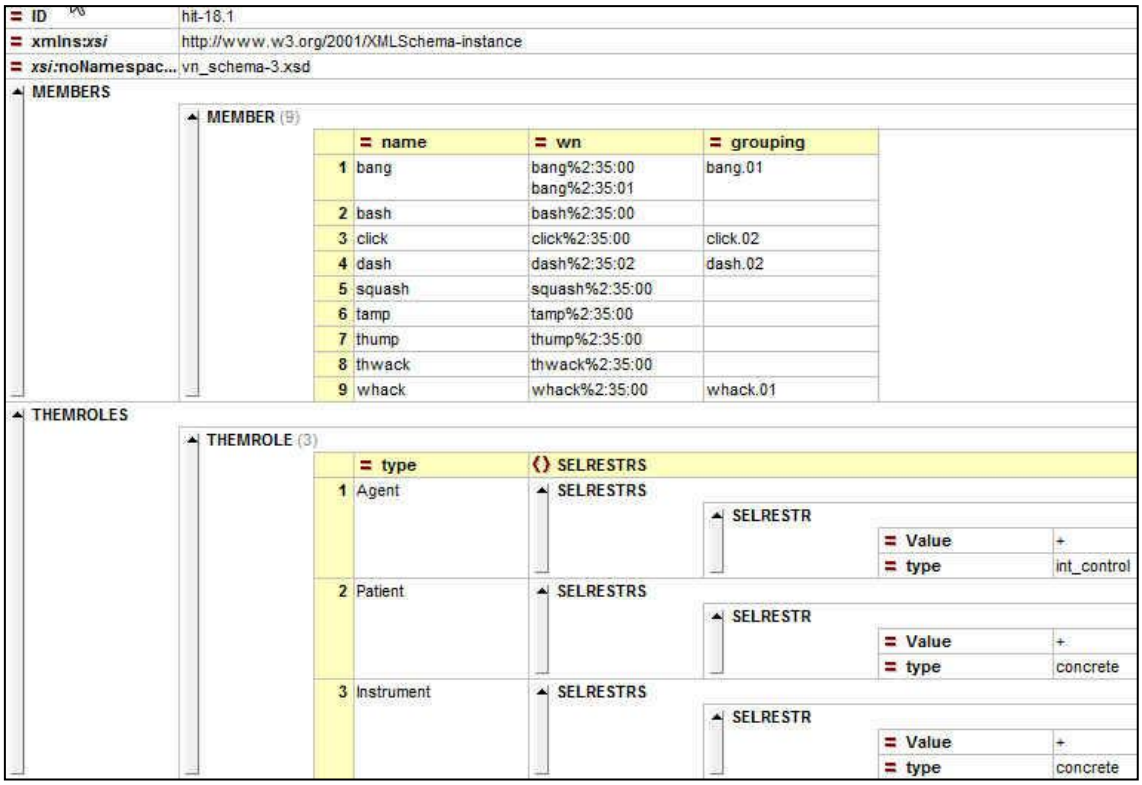


Figure 55. Extract of the HIT class in VerbNet focusing on members and thematic role description

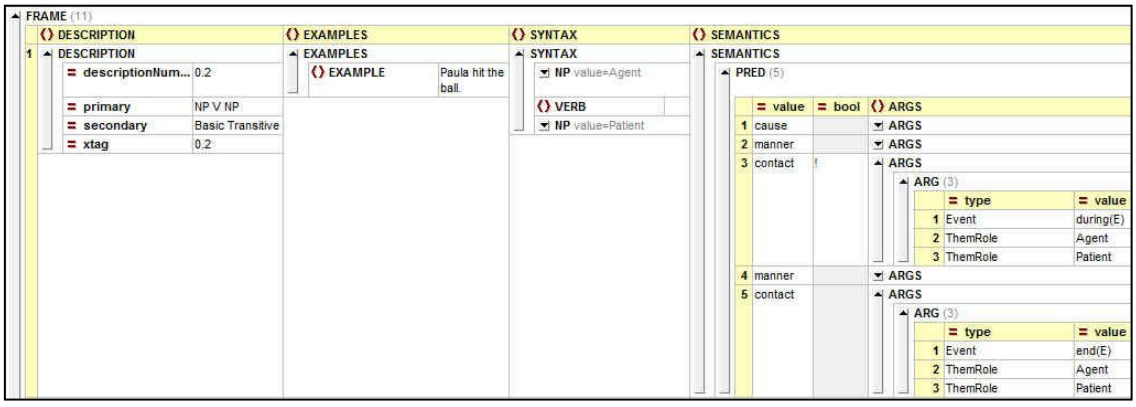


Figure 56. Extract of the HIT class in VerbNet focusing on the frame description

Figure 55 shows the verbs that belong to the HIT class and which share the same basic meaning and have similar syntactic behavior (e.g. 'bang', 'bash', 'click', 'dash', 'squash', 'tamp', 'thump', 'thwack', 'whack'). A complete description of the thematic roles of the arguments is also provided. In this sense, VerbNet establishes 23 thematic roles that are summarized in Table 51, along with their description[105]:

[105] The description of the set of semantic roles used by VerbNet was taken from: <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> [13/03/2012].

Thematic role	Description
ACTOR	Used for some communication classes when both arguments can be considered symmetrical (pseudo-agents).
AGENT	Generally a human or an animate subject. Used mostly as a volitional agent, but also for internally controlled subjects, such as forces and machines.
ASSET	Used for the sum of money alternation with *currency* as a selectional restriction.
ATTRIBUTE	Attribute of PATIENT/THEME refers to a quality of something that is being changed.
BENEFICIARY	The entity that benefits from some action. Generally introduced by the preposition *for*, or double object variant in the benefactive alternation.
CAUSE	Used mostly by classes involving Psychological Verbs and Verbs Involving the Body.
LOCATION, DESTINATION, SOURCE	Used for spatial locations.
DESTINATION	End point of the motion, or direction towards which the motion is directed. Used with a *to* prepositional phrase by classes of change of location, and verbs of sending and carrying. Also used as location direct objects in classes where the concept of destination is implicit (and location could not be SOURCE).
SOURCE	Start point of the motion. Usually introduced by a source prepositional phrase (mostly headed by *from* or *out of*).
LOCATION	Underspecified destination, source, or place, in general introduced by a locative or path prepositional phrase.
EXPERIENCER	Used for a participant that is aware or experiencing something.
EXTENT	Used to specify the range or degree of change
INSTRUMENT	Used for objects (or forces) that come in contact with an object and cause some change in them. Generally introduced by a `with' prepositional phrase.
MATERIAL AND PRODUCT	Used in the Build and Grow classes to capture the key semantic components of the arguments. Used by classes from Verbs of Creation and Transformation that allow for the Material/Product Alternation.
MATERIAL	Start point of transformation.
PRODUCT	End result of transformation.

PATIENT	Used for participants that are undergoing a process or that have been affected in some way. Verbs that explicitly (or implicitly) express changes of state have Patient as their usual direct object.
PREDICATE	Used for classes with a predicative complement.
RECIPIENT	Target of the transfer. Used by some classes of Verbs of Change of Possession, Verbs of Communication, and Verbs Involving the Body. The selection restrictions on this role always allow for animate and sometimes for organization recipients.
STIMULUS	Used by Verbs of Perception for events or objects that elicit some response from an experiencer.
THEME	Used for participants in a location or undergoing a change of location.
TIME	Class-specific role, used in Begin-55.1 class to express time.
TOPIC	Topic of communication verbs to handle theme/topic of the conversation or transfer of message. In some cases, like the verbs in the Say-37.7 class, it would seem better to have `Message' instead of `Topic', but we decided not to proliferate the number of roles.

Table 51. Description of the roles in VerbNet

As shown in Figure 55, the thematic roles of the HIT class are AGENT, PATIENT and INSTRUMENT. The selectional constraint for AGENT is that it must be an internally controlled subject *(int-control)*. PATIENT and INSTRUMENT must be concrete entities. Figure 57 (below) displays some of the selectional restrictions applied to the various thematic roles in VerbNet[106].

In reference to HIT verbs, Figure 56 displays the first frame of the eleven in which the class of HIT verbs participates. Firstly, a schematic description of the frame is given, which provides the possible realizations of the argument structure for constructions, such as transitive, intransitive, prepositional phrase, resultative and the different diathesis alternations. In this respect, the description of the frame in Figure 56 is NP V VP, corresponding to a basic transitive. An example of the frame is also given (e.g. 'Paula hit the ball'). Also included is a syntactic description of the frame (i.e. the syntactic frame) that links each syntactic category to its semantic role. In this sense, it can be seen that in the construction NP V NP, the first NP is associated with the role of AGENT and the second with that of PATIENT. Finally, each frame has semantic information, expressed by means of Boolean semantic predicates (e.g. MOTION,

[106] <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> [14/03/2012].

CONTACT, CAUSE, etc.), the semantic roles of the frame, and the addition of a temporal function or event variable (E) which allows predicates to specify when in the event the predicate is true (e.g. *start(E)*, for the preparatory stage; *during(E)*, for the culmination stage; and *end(E)*, for the culminating stage). In this regard, when verbs within the HIT class activate the meaning of CONTACT, it can be seen that the predicate can be true during the event, as the value associated with the event in the first CONTACT predicate is *during(E)*, or in the culmination stage of the event (*end(E)*).

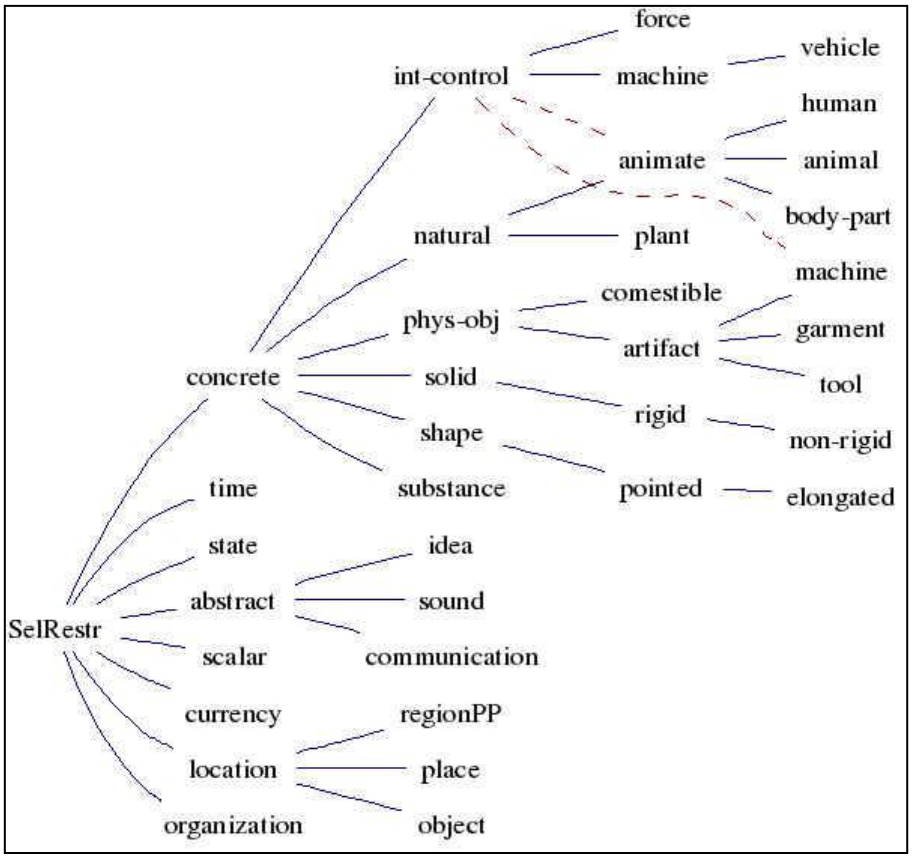


Figure 57. Selectional restrictions associated with thematic roles in VerbNet

VerbNet provides an effective way of mapping the syntax and semantics of a verb and at the same time, offers rich and valuable semantic information. In line with VerbNet, we also match the syntactic category of each conceptual category linked to an argument to its semantic role and macrorole. However, VerbNet also has certain gaps. As pointed out by Perera (2012), certain verb usages are not covered by the frame to which the verb belongs. The most remarkable limitation in VerbNet is that some frames have no corpus instantiations when contexts of use and examples are crucial information for text generation and one of the principal aspects to bear in mind for a verb entry design. In this sense, as pointed out by Pazienza, Pennacchiotti, and Zanzotto

(2006: 1373) since the representation of complex verb relations and the connection to a reference corpus are not the focus of VerbNet, a mapping to other resources such as PropBank (described below) can be very useful.

3.2.3 PropBank

Proposition Bank or PropBank[107] (Palmer, Gildea, and Kingsbury 2005) was conceived with the idea of adding semantic information to the syntactic structures of the Penn English Treebank II[108], by means of annotating each occurrence of a verb with its semantic roles. It shared with FrameNet (§3.1.3.2) the objective of documenting the syntactic realization of arguments of the predicates of the general English lexicon by annotating a corpus with semantic roles. However, contrary to FrameNet, which provides a detailed description of all the possible syntactic realizations of all semantic classes, PropBank is aimed at offering a broad-coverage of annotated corpus which could be useful for the development of language understanding systems, including text understanding, machine translation and fact retrieval (Palmer, Gildea, and Kingsbury 2005: 75). In the words of the Palmer, Gildea, and Kingsbury (2005: 73):

> By annotating semantic roles for every verb in our corpus, we provide a more domain-independent resource, which we hope will lead to more robust and broad coverage natural language understanding systems. […] The Proposition Bank allows us for the first time to determine the frequency of syntactic variations in practice, the problems they pose for natural language understanding, and the strategies to which they may be susceptible.

In this sense, PropBank has led to the creation of (i) a semantically annotated corpus, and to (ii) a verb lexicon, which contains about 3,500 verbs (Kingsbury and Palmer 2002: 1992). Figure 58 displays an extract of the verb 'hit' in the PropBank lexicon:

[107] <http://verbs.colorado.edu/~mpalmer/projects/ace.html> [20/03/2012].
[108] < http://www.cis.upenn.edu/~treebank/> [20/03/2012].

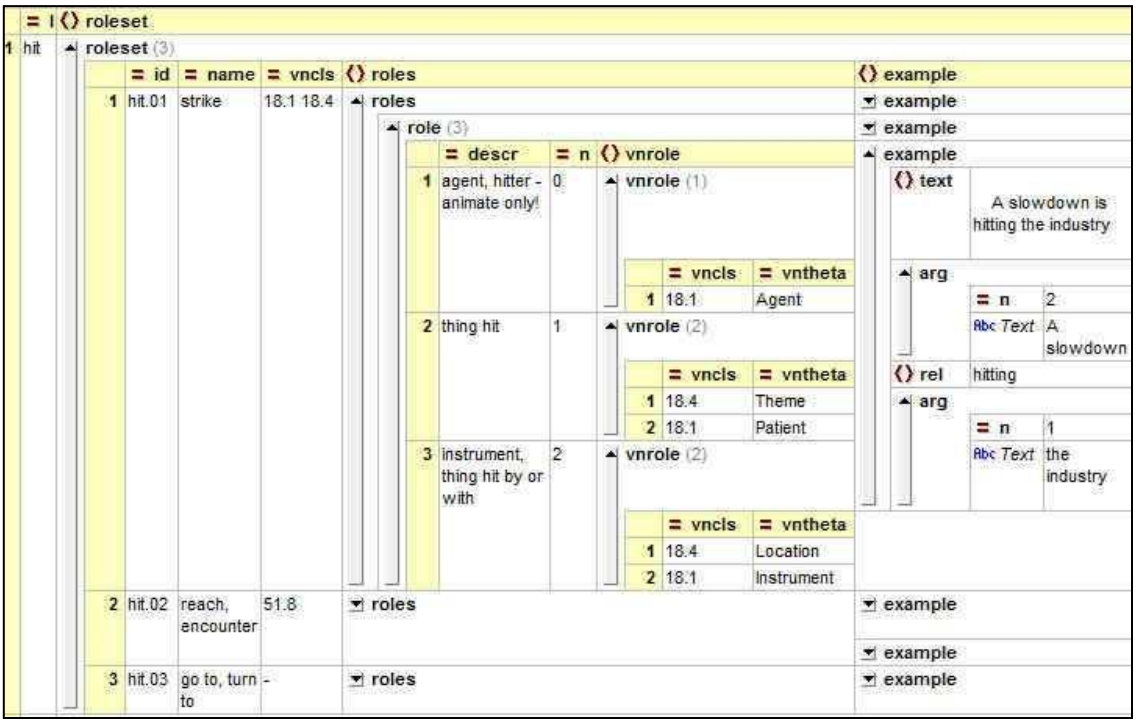


Figure 58. Extract from the frame for 'hit' in the PropBank lexicon

Each verb in PropBank is represented by a *frame*, composed of one or various *framesets* that account for the different senses of a verb. In this sense, as shown, the 'hit' frame has three different framesets: (1) the first refers to the meaning of strike (the meaning in which we are interested in for the sake of coherence with all the examples displayed in this chapter); (2) the second refers to the idea of reach or encounter; (3) the third activates the idea of going to or turning to.

Different framesets for the same frame are established in order to account for the number of arguments that the same predicate can be associated with:

> The criteria for distinguishing framesets are based on both semantics and syntax. Two verb meanings are distinguished as different framesets if they take different numbers of arguments (Palmer, Gildea, and Kingsbury 2005: 77).

This is contrary to constructional approaches, such as Construction Grammar (§3.1.3.3), which state that the realization of a different number of arguments associated with the same verb does not lead to a different sense. What happens is simply that the verb in question appears in a construction that requires a different number of arguments. It is thus the construction and not the verb that licenses a particular argument in an atypical construction.

PropBank also seeks to relate semantically similar verbs in a consistent way, based on VerbNet classes for determining semantic relatedness. In this sense, each PropBank frameset is manually linked, when possible, to the VerbNet class expressing its syntactic-semantic behavior (Kipper, Palmer, and Rambow 2002; Kipper, Snyder, and Palmer 2004). However, not all framesets in PropBank can be associated with VerbNet classes, since a verb sense can be present in PropBank but not in VerbNet. Furthermore, a PropBank frameset can be mapped onto more than one VerbNet class. What is missing in PropBank is a link to WordNet. Therefore, according to Pazienza, Pennacchiotti, and Zanzotto (2006: 1373), it is not possible to directly use PropBank knowledge in conjunction with WordNet information on verb relations. Because of all these similarities between WordNet, VerbNet, FrameNet and PropBank, the project SemLink emerged as a way to link and unify them.

Regarding the description of the verb entry or frame entry in PropBank, as shown in Figure 58, each frameset has a name (which refers to its sense), a number which links the particular sense to the sense in VerbNet, a complete description of the roles in that sense, as well as example sentences from the corpus which show all the possible syntactic realizations of the frameset. This means that sometimes a particular realization is not found in the Penn Treebank corpus. Should this occur, an invented sentence is used, usually headed by *John* (Palmer, Gildea, and Kingsbury 2005: 72). Our methodology for providing example contexts in our verb entry is similar. If good examples are not retrieved from the bilingual comparable corpus compiled for this research, we provide examples from the web, following the web as corpus techniques (§4.2.2.2).

PropBank defines semantic roles on a verb-by-verb basis. Accordingly, semantic roles in PropBank are more specific than thematic roles in VerbNet. In other words, whereas in VerbNet, roles are general and valid across different classes, in PropBank they are strictly tied to a specific roleset. As a consequence, VerbNet has only 20 thematic roles, while PropBank has more than 1,400 roles (Pazienza, Pennacchiotti, and Zanzotto 2006: 1373). Each PropBank roleset is mapped, when possible, onto the corresponding VerbNet thematic role. However, not all roles in PropBank have a mapping to a thematic role.

In this sense, PropBank's roles can be compared to the open-ended frame element roles depicted by FrameNet (§3.1.3.2). As previously mentioned, our inventory of roles is more restrictive since for our purposes if the inventory of roles is open-ended and

unconstrained, then the roles have no general descriptive adequacy. Accordingly, the set of semantic roles in this study largely coincides with the thematic relations and macroroles provided by Role and Reference Grammar (§3.1.2.2).

The set of semantic roles associated with each frameset in PropBank is referred to as *roleset*. In this sense, in every frameset, there is a *description* field for each role, such as *hitter, thing hit, instrument, thing hit by or with* (see Figure 58), which aims at helping during the annotation and as documentation, but which does not have any theoretical background (Palmer, Gildea, and Kingsbury 2005: 76). Each argument is then numbered, starting with 0, 1, 2, etc. Accordingly, for instance, Arg0 is generally matched to the role of AGENT, while Arg1 is prototypically a PATIENT or THEME. However, no generalizations can be made across verbs because of the large number of arguments labeled despite the fact that efforts were made to consistently define roles across members of VerbNet classes (Palmer, Gildea, and Kingsbury 2005: 75). In addition to verb-specific numbered roles, PropBank also defines other more general roles that can be applied to any verb. Finally, its link to VerbNet roles is provided.

It should be emphasized that resources such as PropBank are intended for natural language processing. Therefore, this level of description might be useful and necessary at the same time but, since our objective is to enhance a resource useful for text generation, this level of specificity is unnecessary.

3.2.4 ADESSE

ADESSE (*Base de datos de verbos, Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español*) is an online database[109] of syntactic and semantic information for verbs in Spanish, currently being developed at the University of Vigo (Spain). It is an extended version of the Spanish syntactic database BDS (*Base de datos sintácticos del español actual*), which contains the grammatical features of verbs and verb-arguments of a corpus of general Spanish referred to as *Arthus*[110] (***AR**chivo de **T**extos **H**ispánicos de la **U**niversidad de **S**antiago de Compostela*) of 1.5 million words and 159,000 clauses. In this sense, ADESSE has inherited all the syntactic information provided by the BDS and complemented it with semantic information.

The semantic classification in ADESSE is a conceptually-based hierarchical classification. Verbs are organized in verb classes and subclasses of various levels. Each

[109] <http://adesse.uvigo.es/> [02/03/2012].
[110] <http://adesse.uvigo.es/data/corpus.php> [02/03/2012].

class and subclass contains verbs with their different senses. According to Vaamonde, González, and García (2010: 1907), verb classes are based on the notion of frame in FrameNet:

> We think that each individual verb evokes a conceptual frame, that is, a complex conceptual representation which includes some basic participants in a scene […]. The main goal of ADESSE verb classification is to represent generalizations over these types of conceptual frames evoked by individual verbs.

In this sense, the first level is composed of six general classes, namely, MENTAL, RELATIONAL, MATERIAL, VERBAL, EXISTENTIAL, and MODULATION. Each of these general classes is divided into different subclasses (e.g. within the MATERIAL class, SPACE, CHANGE, OTHER FACTS, and BEHAVIOR are distinguished). In addition, most of them are further subdivided in more subclasses, which produce a third level (e.g. OTHER FACTS has the subclasses of MATERIAL process, METEOROLOGY, EMISSION, CONTACT, CONTROL and USE). Finally, a fourth level is established when further semantic subdivisions can be made within subclasses (e.g. CONTACT has the subclasses of SIMPLE CONTACT, IMPACT, PRESSION, FRICTION and AFFECTIVE CONTACT). These classes can be compared to the lexical domains provided by the Lexical Grammar Model (§3.1.2.1). Table 52 (following page) shows a list of some of the semantic classes defined by ADESSE, focusing on the MATERIAL class[111].

In addition, according to García, Costas, and Martínez (2005: 374), ADESSE is based on premises of Cognitive Grammar (§3.1.3) and Construction Grammar (§3.1.3.3), and more specifically, on the idea of semantic compatibility between verb meaning and constructional meaning. ADESSE contains about 3,450 different verb lemmas, which correspond to more than 4,000 verb entries with their different senses.

[111] The complete semantic classification can be found in: <http://adesse.uvigo.es/data/clases.php>. [03/03/2012].

1. MENTAL
2. RELATIONAL
3. MATERIAL
   - SPACE
   - CHANGE
   - OTHER FACTS
     - METEOROLOGY
     - EMISSION
     - CONTACT
       - SIMPLE CONTACT
       - IMPACT
       - PRESSURE
       - FRICTION
       - AFFECTIVE CONTACT
     - CONTROL
     - USE
     - SUBSTITUTION
     - ACTIVITY
     - COMPETITION
     - BEHAVIOR
4. VERBAL
5. EXISTENTIAL
6. MODULATION

Table 52. Semantic classification in ADESSE. Extract displaying the MATERIAL class

We describe the ADESSE entry for 'sacudir' [shake, hit], which is one of the most recurrent verbs in our corpus of Spanish texts concerning the EXTREME EVENT.

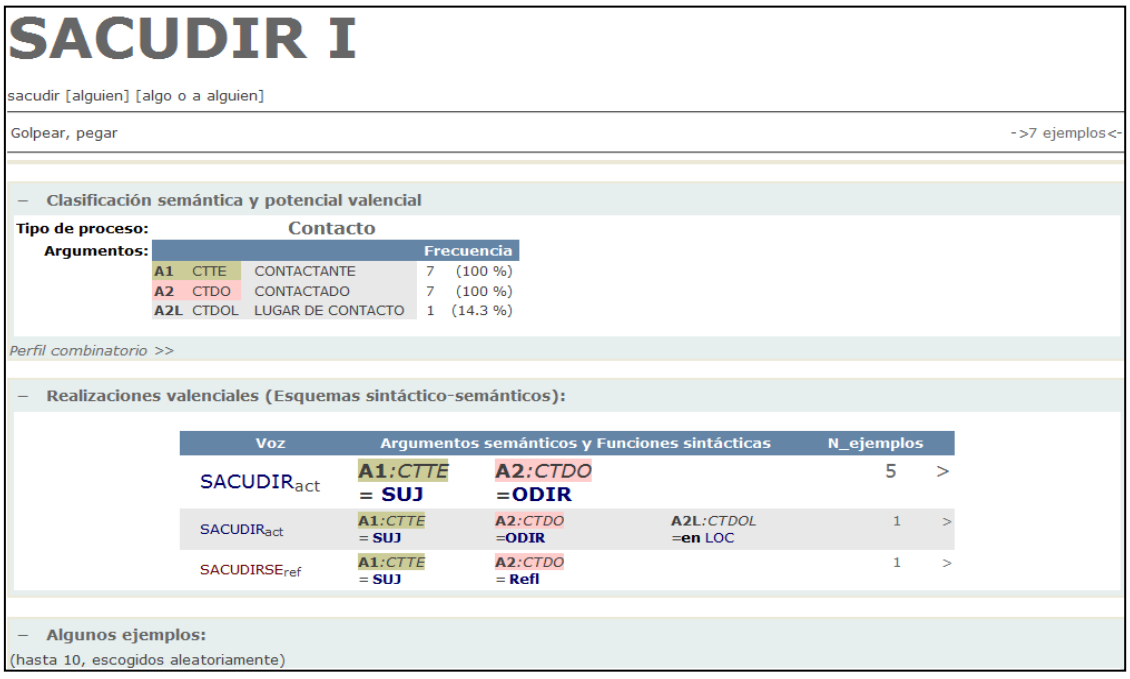


Figure 59. Extract from the entry for 'sacudir' in ADESSE

As shown in Figure 59, each entry in ADESSE is divided into three sections: (i) *Clasificación semántica y potencial valencial* [semantic classification and valency potential]; (ii) *realizaciones valenciales, esquemas sintáctico-semánticos* [valency realizations and syntactic-semantic schemas]; (iii) *algunos ejemplos* [some usage examples]. The valency potential of a verb refers to the set of potential arguments which can be selected by the verb, whereas the valency realizations are the set of arguments actually expressed by each syntactic form (Vaamonde, González, and García 2010: 1905).

Within the first subdivision (*clasificación semántica y potencial valencial*), the class or subclass to which the verb entry belongs is specified. As shown, 'sacudir' belongs to the CONTACT subclass described in Table 52. When CONTACT is clicked on, ADESSE displays another window with a complete description of the CONTACT subclass (see Figure 60, below) by means of the following: (i) the definition of the verb class; (ii) the typical semantic roles in the class, along with its description, a usage example from the corpus in which the semantic role is realized, and its frequency in the corpus; (iii) the most frequently verbs activated in the class; (iv) the most frequently constructions realized in the class. As shall be seen, in the verb entry proposed in this thesis, we also first specify the lexical domain from the LGM to which the verb belongs in order to provide a general semantic characterisation of it, along with the frame activated and a definition of the frame.

As shown, in the description of the first subdivision of 'sacudir' (Figure 60), the class or subclass of the verb is specified as well as the possible semantic roles associated with its arguments. In this respect, the semantic roles associated with the verb 'sacudir' are those of *contactante* [IMPACTOR], *contactado* [IMPACTEE], and *lugar de contacto* [PLACE OF CONTACT]. As can be observed, both the IMPACTOR and IMPACTEE are selected in 100% of the constructions, whereas the PLACE OF CONTACT is only present in 14.3% of them.

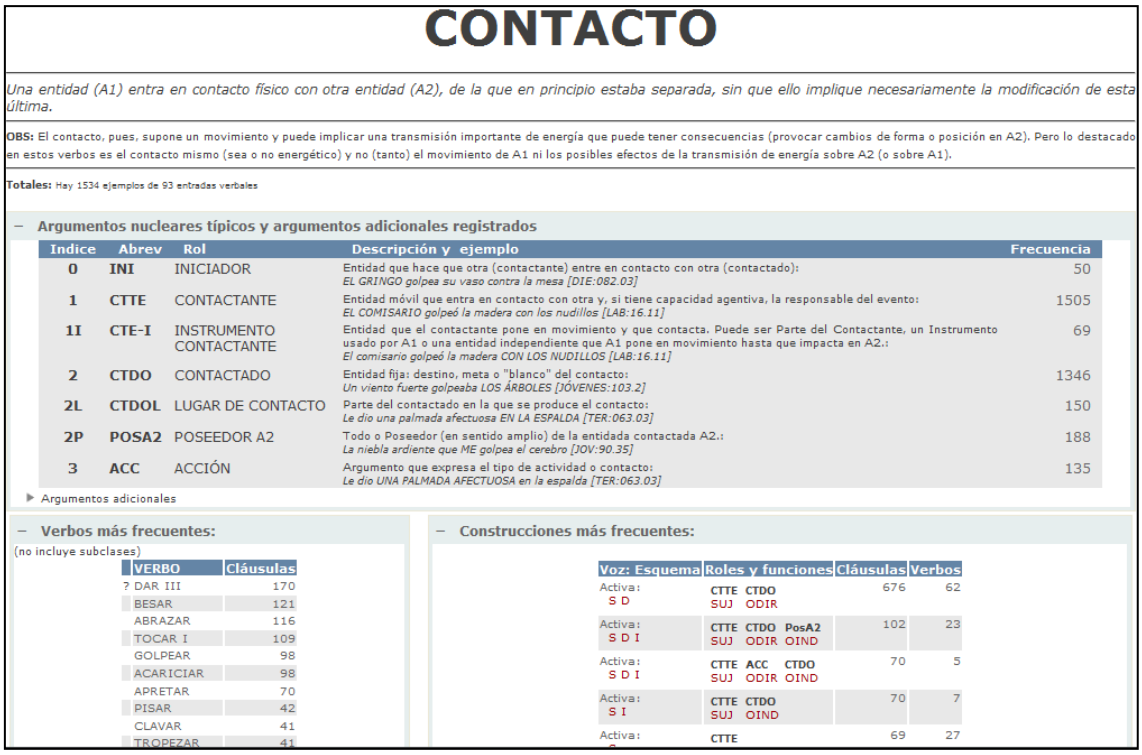


Figure 60. Extract of the CONTACT verb subclass in ADESSE

Table 53 provides the inventory of semantic roles recognized by ADESSE for each semantic class[112]:

Semantic classes	A0	A1	A2	A3	A4	A5
Mental		EXPERIMENTER	PHENOMENON			
Sensation		EXPERIMENTER	STIMULUS			
Volition		EXPERIMENTER	STIMULUS			
Perception	INITIATOR	PERCEIVER	PERCEIVED	PERCEIVED-2		
Cognition		COGNIZER	CONTENT	CONTENT-2		
Knowledge	INITIATOR	COGNIZER	CONTENT	CONTENT-2		
Believe	INITIATOR	COGNIZER	CONTENT	CONTENT-2		
Election		ELECTOR	ELECTED	OPTION		
Attribution	ASSIGNATOR	ENTITY	ATTRIBUTE			
Relation	ASSIGNATOR	ENTITY	ENTITY-2			
Property	ASSIGNATOR	ENTITY	ATTRIBUTE			
Measure	MEASURER	ENTITY	MEASURE			
Appearance		ENTITY	APPEARANCE	PERCEIVER		
Designation	DESIGNATOR	DESIGNATED	DESIGNATION			
Possessed		POSSESSOR	POSSESSED			
Acquisition		END-POSSESSOR-	POSSESSED	INITIAL-POSSESSOR		
Transference	DONOR	END-POSSESSOR	POSSESSED			
Space	INITIATOR	MOVING ENTITY	ORIGIN	DIRECTION	PATH	
Movement	INITIATOR	MOVING ENTITY	ORIGIN	DIRECTION	PATH	
Location	INITIATOR	MOVING ENTITY	LOCATION			

[112]The subset of roles was translated from <http://adesse.uvigo.es/Docu/Roles> [22/03/2012].

Position	INITIATOR	MOVING ENTITY				
Orientation	INITIATOR	MOVING ENTITY	ORIENTATION			
Manner-motion	INITIATOR	MOVING ENTITY				
Union	INITIATOR	ENTITY	ENTITY-2			
Change	AGENT	PATIENT				
Creation	CREATOR	CREATION				
Modification	AGENT	AFFECTED				
Body care	AGENT	AFFECTED				
Destruction	DESTROYER	DESTROYED				
Meteorology		METEOR				
Emission		EMITTER	EMISSION			
Contact	INITIATOR	IMPACTOR	IMPACTEE	ACTION		
Control		CONTROLLING ENTITY	CONTROLLED ENTITY	POSSESSOR A2		
Use		USER	IMPLEMENT	PURPOSE		
Consumption	CONSUMER	CONSUMED				
Substitution	INITIATOR	SUBSTITUTE	SUBSTITUTED			
Activity		AGENT	ACTIVITY			
Competition		COMPETITOR	ANTAGONIST	GOAL		
Behavior		AGENT	MANNER			
Physiology	INITIATOR	AGENT	PRODUCT			
Ingestion		INGESTOR	INGESTION			
Social relationships		AGENT	CO-AGENT			
Communica-tion		COMMUNICATOR	MESSAGE	RECIPIENT	TOPIC	CO-DE
Request		EMITTER	REQUEST	RECIPIENT		
Assessing		ASSESSOR	ASSESSED	TOPIC	ASSESS MENT	
Existence	INITIATOR	EXISTENCE				
Phase	INITIATOR	ENTITY/PROCESS	CO-OCCURRING			
Time	INITIATOR	EXISTENT	DURATION	EXISTENT-2		
Life	AGENT	PATIENT				
Causation	CAUSATOR	AFFECTED	ACTION			
Induction	INDUCTOR	AFFECTED	ACTION			
Allowance	ALLOWER	ALLOWED	ACTION			
Obligation	OBLIGATOR	OBLIGATED	ACTION			
Disposition		AGENT	ACTION			
Acceptance		ACCEPTOR	ACCEPTED			
Verbs of support	INITIATOR	AGENT	BASE	REFERENCE		

Table 53. Inventory of semantic roles in ADESSE (own translation)

Along with the roles specified in Table 53, ADESSE considers certain roles not to be class-bound, but as belonging to various classes, such as the following: BENEFICIARY, LOCATION, ATTRIBUTE, INITIATOR, POSSESSOR, MEDIUM, FEATURE, REFERENCE, ROLE, PURPOSE, ORIGIN, OBJECT, FINAL STATE, MANNER, COMPANY, CAUSE, ISSUE, DOMAIN, and NON-DEFINED.

Within the first subsection of *clasificación semántica y potencial valencial* in Figure 59, when users click on *perfil combinatorio* [combinatorial profile], they can access a complete characterisation of the types of arguments associated with each

semantic role as shown in Figure 61. This argument description takes into account the following features: (i) syntactic function (i.e. subject, direct object, indirect object, etc.); (ii) clitics (i.e. if accompanied by a morpheme that has the syntactic characteristics of a word, but shows evidence of being phonologically bound to another word); (iii) if headed by a preposition; (iv) syntactic category (i.e. noun phrase, prepositional phrase, adverbial phrase, adjectival phrase, etc.); (v) semantic type, i.e. animate, concrete, abstract and propositional; (vi) frequent lexical realizations.

Perfil combinatorio (resumen cuantitativo de propiedades de argumentos)

SACUDIR I - Golpear, pegar

ejemplos anotados

Argumento: (Frec. explícito)	A1 (Contactante) 7 (100%)	A2 (Contactado) 7 (100%)	A2L (Lugar de contacto) 1 (14%)
Función sintáctica:	**SUJ** 7	**ODIR** 6 Refl 1	LOC 1
Clíticos objeto:		Ac/Dat 3 Dat 1	
Preposiciones:	N/A o nulo 4	N/A o nulo 1 *a* 1	*en* 1
Categoría sintáctica:	**FN** 4	FN 2	FN 1
Tipo semántico:	Animado 3 Abstracto 2 Concreto 2	**Animado** 5 N/A o nulo 1 Concreto 1	Concreto 1
Realizaciones léxicas frecuentes: (sólo no animados)	*pregunta* 1 *golpe* 1 *música* 1 *arrebato* 1	*jardín* 1 N/A o nulo 1	***médula*** 1

Figure 61. Combinatorial profile of 'sacudir' in ADESSE

As an example, Figure 61 shows that the arguments associated with the IMPACTOR (*contactante*) semantic role can be an *animate*, *abstract* or *concrete* entity, which normally have the syntactic function of *subject* and belong to the grammatical category of *noun phrase*. Typical lexical realizations linked to the semantic role of IMPACTORS are 'pregunta' [question], 'golpe' [knock], 'música' [music], and 'arrebato' [fit (in its sense of short period)].

The second subdivision in an entry is entitled *realizaciones valenciales, esquemas sintáctico-semánticos* [valency realizations and syntactic-semantic schemas] (Figure 59, above), and includes the various constructions in which the verb can participate. For each construction, the entry specifies the semantic roles and the syntactic functions associated with each semantic role, depending on its voice (passive, active, etc.), along with the number of examples in the corpus of the construction. In this sense, it can be seen that 'sacudir' typically appears in three diferent constructions (see Table 54):

(1) IMPACTOR + IMPACTEE (subject) + (direct object)	(2) IMPACTOR + IMPACTEE + LOC. (subject) + (direct object) + (pp)	(3) IMPACTOR + IMPACTEE (reflex.) (subject) + (reflexive)

Table 54. Constructions in which 'sacudir' can appear

Finally, a verb entry within its third subdivision (Figure 59 above) offers usage examples from the corpus. Figure 62 shows the examples for 'sacudir':

– **Algunos ejemplos:**
(hasta 10, escogidos aleatoriamente)

Verbo	Texto	Referencia	
SACUDIR-I	me expulsan de la cátedra; me bato en duelo con algunos amantes de Dorothy; *me* ***sacuden****; se ríen; bebo.. Y en mis ratos libres sigo* intentando averiguar --sin éxito, desde luego-- cómo era el	[CIN:092.20]	+info
SACUDIR-I	vez más. Sueña. Abre los ojos. Por el sueño ha *pasado una bandada de sombras. ¡****Sacúde****te! ¡Golpéate!* Así. Ahora avanza, un paso, otro paso, más	[MIR:113.28]	+info
SACUDIR-I	y casi violenta. La música terriblemente hermosa, *que desgarraba todavía más la noche, me* ***sacudió*** y me abofeteó. Fue entonces cuando salí	[CAR:148.16]	+info
SACUDIR-I	a que -conforme al símil anteriormente trazado-- su otra mitad, *el ofuscado cónyuge,* ***sacuda*** *al presunto soñador a* bofetadas e intercepte con brutalidad la ingeniosa	[PAI:092.11]	+info
SACUDIR-I	Ni mi mayo francés ni los constantes arrebatos de estos chicos que *veo en la universidad me* ***sacuden*** *como antes en la médula.* Me veo lejana, distante, un poco ajena. He debido	[JOV:128.06]	+info
SACUDIR-I	, compañeros de militancia y de su propia y desdichada esposa le ametrallan de modo *implacable,* ***sacudiéndo****le como un punchingball: el* pueblo vietnamita ha asumido victoriosamente su	[PAI:144.06]	+info
SACUDIR-I	angustias de la savia para seguir subiendo tronco arriba. *Pero más le duelen los golpes que* ***sacuden*** *el jardín* como paletadas de sepulturero; hachazos cuya torpeza	[SON:113.06]	+info

Figure 62. Extract of the usage examples for 'sacudir' in ADESSE

As can be deduced from the description in this section, the main objective in ADESSE is to describe the most general meaning of each verb and compare it with the prototypical cases of each class. However, there are verbs that are likely to belong to different semantic classes. Based on this presupposition, our verb description also establishes at the first level a specification of the class or lexical domain to which the verb belongs, based on the LGM (§3.1.2.1). However, the same verb can belong to different domains or subdomains, depending on their construal and on the notion of muldimensionality of concepts in terminology. As shall be seen, in our research only the

sense(s) of the verb associated with the notion of natural disaster were taken into account. However, when various relevant senses were encountered, assignment to a particular class and frame within a class was performed in terms of the differences of the semantic roles activated by each specific sense.

A very valuable aspect of ADESSE is that it offers a fine-grained description of the arguments associated with each semantic role, which has been really useful for our verb entry proposal. The idea of providing the constructions in which the verb can participate by means of its semantic roles has also been one of the inspirations for our verb description.

3.2.5 SenSem

The project SenSem[113] *(Sentence Semantics: Creación de una Base de Datos de Semántica Oracional)* was created by the GRIAL[114] (*Grup de Recerca Interuniversitari en Aplicacions Lingüístiques*) research group in Catalonia (Spain). It led to the creation of a corpus of syntactically and semantically annotated sentences in Spanish and to the development of a verbal lexicon derived from the annotated data of the corpus. From the corpus of 13 million words, about 30,000 sentences were randomly selected (about 120 for each of the 250 most frequent verbs in Spanish), 25,000 belonging to journalistic texts and around 5,000 to literary texts (Fernández and Vázquez 2012: 158).

Each sentence was annotated at various levels (Vázquez and Fernández 2011: 171): (i) at the lexical level (the sense of the verb); (ii) at the morphological level (morphosyntactic category); (iii) at the syntactic level (functions and syntactic categories); (iv) at the semantic level. The semantic annotation was carried out at the level of word (*Aktionsart* and the sense of the verb), at a syntagmatic level (semantic roles of the arguments and the verbal syntagmatic aspect), and at the sentence level (constructions, modality, and aspectuality).

This annotation was the basis for the creation of a lexical database of verbs in Spanish. The SenSem corpus was also used, at a later stage to implement the Catalonian counterpart of the verbal database. This was achieved by translating the already annotated Spanish sentences into Catalonian. However, the Catalonian repertoire is more limited since the translation is only of the newspaper texts (which are 80% of the total number of sentences).

[113] <http://grial.uab.es/fproj.php?id=10> [18/03/2012].
[114] <http://grial.uab.es/> [18/03/2012].

The methodology followed for the annotation of the Catalonian sentences was exactly the same as for the Spanish ones. Therefore, the lexical database of verbs developed by SenSem can currently be searched either for Spanish or Catalonian[115]. Since Catalonian is not a language represented in EcoLexicon, the rest of this section focuses on the description of the verbs in Spanish in the database.

As previously mentioned, the verbal lexicon of SenSem contains the 250 most frequent verbs for Spanish. In this sense, the verbs belonging to the EXTREME EVENT, such as 'golpear', 'sacudir', 'temblar' [to hit, to strike, or to shake] are not present in the database. Therefore, the example given of the information provided in SenSem verb entry is exemplified by 'sufrir' [to suffer]. Although 'sufrir' is not a specialized verb, it can serve as a support verb for constructions involving contact or impact, such as 'sufrió un impacto', 'sufrió un golpe'.

As shown in Figure 63, each verb is described in terms of its various senses. In this regard, for each sense, apart from the definition, the semantic roles associated with the arguments in that particular sense, its aspectual class (i.e. event, process, process/event, state) and frequency in the corpus of annotated sentences is provided.

SenSem distinguishes twenty semantic roles, some of which (e.g. AGENT, EXPERIENCER, INITIATOR, THEME and TIME) are subdivided into more specific ones. The complete set of roles is the following[116]:

- AGENT (agent all, agent plural, agent destination, agent experiencer, agent goal, agent source, agent theme, agent cause)
- QUANTITY
- CAUSE
- CIRCUMSTANCE
- COMPANY
- QUALITY
- DESTINATION
- EXPERIENCER (experiencer all, experiencer, experiencer plural)
- PURPOSE
- INITIATOR (initiator all, initiator, initiator plural)
- INSTRUMENT

[115] <http://grial.uab.es/sensem/lexico/main> [18/03/2012].
[116] The semantic roles in SenSem were extracted from the search menu within the database interface: <http://grial.uab.es/sensem/lexico> [18/03/2012].

- LOCATION
- MANNER
- MEDIUM
- ORIGIN
- RECIPIENT
- ROUTE
- SUBSTITUTE
- THEME (theme all, theme, theme initial-stage, theme result-stage, theme deplaced-object, theme plural, affected theme, affected-theme beneficiary, affected-theme created, affected-theme destroyed, affected-theme plural, affected-theme victim)
- TIME (time all, temporal end, temporal beginning, temporal location)

Regarding 'sufrir', the third sense ('recibir un ataque, golpe, o cambio dramatico') [to receive an attack, hit or dramatic change] is most relevant to EXTREME EVENT in its sense of natural disaster, whose associated semantic roles are DESTINATION and THEME, and which belongs to the aspectual class of *event*. It is represented in 44 sentence examples out of the total of 120 activated by the verb 'sufrir'.

sufrir

Sentido	Definición	Roles	Clase aspectual	Frecuencia
1	Padecer algún dolor físico o moral.	Experimentador, Origen	proceso	55/120
2	Ser capaz de soportar.	Experimentador, Tema	proceso	21/120
3	Recibir un ataque, golpe, o cambio drástico.	Destino, Tema	evento	44/120

Figure 63. Display of the various senses of a verb in SenSem. The example of 'sufrir' [suffer]

When users click on the definition of each sense, the system displays a new window with a complete description of the sense in question in terms of its definition, semantic roles, aspectual class (which were already described in the general entry verb), the number of the entry in WordNet with which it is associated, synonyms (if any), its frequency within the journalistic texts and within the literary texts, its Catalonian translation, and the various argument structures and adjunct constituents where the sense is realized (see Figure 64 below):

sufrir 3

Definición:	Recibir un ataque, golpe, o cambio drástico.
Roles semánticos:	Destino, Tema
Clase aspectual:	evento
Wordnet:	01204493v
Sinónimos:	
Frec. periodístico:	42/100
Frec. literario:	2/20
Trad. catalán:	sofrir_3 patir_2
Estructuras argumentales:	• [41] SN V SN • [1] SP V SN • [1] SN V SADV • [1] SN V
Constituyentes adjuntos:	• [10] SADV • [10] SP • [1] SN

Figure 64. SenSem information for each verb sense. Third sense of 'sufrir' [suffer]

The various argument structure and adjunct constituent schemas provide access to the set of annotated sentences that express each particular schema. As such, for example, the argument structure "SN V SN" is present in 41 examples, which can be displayed when clicking on the specific construction. Each construction is further subdivided into two large group types: (1) those whose logical subject coincides with the syntactic subject; (2) those whose subject is not expressed because of generalization or because it is not known]. As pointed out by Fernández and Vázquez (2012: 163), choosing a correct label to name these two groups was far from easy:

> Por coherencia en la anotación, se optó por distinguir básicamente entre dos grandes tipos de construcciones: aquellas en que el sujeto lógico coincidía con el sintáctico y aquellas en las que no quedaba expresado, bien por generalización o desconocimiento. Elegir las etiquetas para denominar estos fenómenos ha sido complicado. Por el momento, se optado por denominar estos fenómenos como casos de topicalización y destopicalización del sujeto lógico, respectivamente.

In addition, as shown in Figure 65, the construction "SN V SN" concerning the topicalization of the logical subject is further subdivided into seven subconstructions: (i) "SN V PR-Rel", for constructions in which the second NP is headed by a relative

pronoun; (ii) "(elidido) V PR-Rel", referring to examples whose first NP is not expressed and whose second NP is headed by a relative pronoun; (iii) the construction "SN V SN"; (iv) "(elidido) V SN", when the first NP is not expressed; (v) "PR Rel V SN", in which the first NP is headed by a relative pronoun; (vi) "SN V SPron", whose second NP is realized by a pronominal phrase not being a relative cause; (vii) "Nom. Propio V SN", referred to those sentences whose first NP is realized by a proper noun.

Esquemas de realización de "SN V SN" [41]

sufrir_3 - Recibir un ataque, golpe, o cambio drástico.

- Topicalización del sujeto lógico:
  1 - SN V PR-Rel [6]
  2 - (Elidido) V PR-Rel [1]
  3 - SN V SN [11]
  4 - (Elidido) V SN [14]
  5 - PR-Rel V SN [4]
  6 - SN V SPron [1]
  7 - Nom.Propio V SN [4]
- Destopicalización del sujeto lógico:

Topicalización del sujeto lógico

1

Categorías:	Sintagma nominal (nombre común)		Sintagma pronominal (pronombre relativo)
Funciones sintácticas:	Sujeto	V	Objeto directo
Roles semánticos:	Destino		Tema

- **Orden neutro de los argumentos**
  - **Sin adjuntos:** No hay ejemplos
  - **Con adjuntos:** No hay ejemplos
- **Alteración de orden en los argumentos**
  - **Sin adjuntos:**
    1. Escribo para manifestar mi indignación por la manipulación que **han sufrido** varios cientos de niños que acudieron con ilusión a ver a la selección española de fútbol el pasado día 27 en Barcelona. (id: 27557) (más info.)
    2. Quienes asistimos regularmente a la Filmoteca de la Generalitat de Catalunya hemos constatado el alto grado de degradación: sillas rotas, muelles que se te clavan, mal funcionamiento del proyector (con cortes constantes que te desconcentran) y el alto grado de envejecimiento que **sufre** el cine Aquitania, sede de la Filmoteca. (id: 27598) (más info.)
    3. Quiero manifestar la degradación que **sufre** este barrio. (id: 27608) (más info.)
    4. Como vecino del barrio de Sants veo con preocupación la continua degradación que **está sufriendo** el entorno del centro cívico Les Cotxeres. (id: 27651) (más info.)
  - **Con adjuntos:**
    1. Este es el tercer incendio que **sufre** la empresa en menos de un año. (id: 27576) (más info.)
    2. El informe también hace alusión a la capacidad del comercio barcelonés para adecuarse a las transformaciones que **sufre** el sector en la actualidad. (id: 27607) (más info.)

Figure 65. Extract of the SenSem information displayed for the "SN V SN" argument structure concerning the third sense of 'sufrir' [suffer]

More specifically, the example displayed in Figure 65 belongs to the first subconstruction "SN V PR Rel". As shown, for each subconstruction, the syntactic function fulfilled by each member of the construction is provided, along with the semantic roles associated with each element (i.e. the first NP corresponds to the subject and is assigned the role of DESTINATION, whereas the relative clause is the direct object and is assigned the role of THEME). Then, the six corpus examples for this particular construction are displayed.

SenSem also confirms that there can be no predicate-argument representation without taking into consideration both syntax and semantics, and accounting for the arguments, the roles associated with them, syntactic and grammatical categories and

usage examples. However, the information offered in the various subdivisions of constructions in SenSem is excessive for our purposes.

3.3 Summary

This chapter describes the linguistic theories that can be applied to the study of phraseology in Terminology, and more specifically, those whose premises were particularly useful for our study. In addition, a description of the most significant lexical resources with verb information is included. The aim of this description is to describe the theoretical foundations of our verb analysis in order to integrate Terminology theories (§1) and Linguistic theories. It also justifies the design of our verb template, which is based on the positive and negative aspects of these verb-argument resources, along with the positive and negative aspects of the lexicographic and terminographic resources analyzed in Chapter 2.

The verbs under study in this thesis are first of all classified in lexical domains as proposed by the LGM. It is worth mentioning that the Spanish lexical resource ADESSE also classifies verbs in general classes, as well as WordNet, which classifies verbs in 15 base types, and VerbNet, which uses a total of 274 classes. Subsequently, our verbs are grouped in frames, in a similar way to FrameNet, but our frames are, nevertheless, more comparable to subdomains within the LGM.

In order to categorize and specify the various frames activated within our subfield, an argument analysis is undertaken. To that end, linguistic realizations of argument were identified and were described taking into account the following: (i) their semantic and thematic roles; (ii) their semantic category; (iii) their morphological or syntactic description.

As is well known, semantic roles generally express the set of properties that a verb entails for a given argument. In other words, they provide the semantic relations between noun phrases with respect to the verbs with which they co-occur. As previously mentioned throughout this chapter, almost every linguistic theory or meaning-based resource establishes its own set of semantic roles. However, there is considerable disagreement as to their number, nature, or function and even the name to refer to them. Surprisingly, one of the most important meaning-based resources in our days, WordNet, does not encode semantic roles, which is one of its main limitations.

As shown in this chapter, the functional and functional-cognitive linguistic theories do express the underlying link between predicates and arguments, but they do it in different ways. Succinctly put, RRG establishes semantic roles at two different levels: *thematic roles* (related to the type of predicate that they occur with, e.g. PERCEIVER, STIMULUS, COGNIZER, CONTENT), and *macroroles* (generalized semantic functions, e.g. ACTOR and UNDERGOER). In his Cognitive Grammar, Langacker introduces the notion of *archetypical role,* as a pre-linguistic conceptualization. At one level, the roles of SETTING, LOCATION, and PARTICIPANT are distinguished, and more specifically, within PARTICIPANT roles, those of AGENT, PATIENT, INSTRUMENT, MOVER, EXPERIENCER, and ZERO are included. Similarly to the RRG macroroles of ACTOR and UNDERGOER, Cogntive Grammar distinguishes between *agent orientation alignment*, on the one hand, and *theme orientation alignment*, on the other hand. In turn, Goldberg in her Construction Grammar distinguishes between *participant roles*, which are roles associated with a particular verb (e.g. HANDER, HANDEE AND HANDED) and *argument roles*, which are more general roles supplied by the construction (e.g. AGENT, CAUSE > RECIPIENT, EXPERIENCER > INSTRUMENT > PATIENT, THEME > LOCATION, SOURCE, GOAL).

Regarding the meaning-based resources described in this chapter, FrameNet designates roles in terms of frame-specific situational roles, referred to as *frame elements* (e.g. IMPACTEE, IMPACTOR, IMPACTORS) rather than general semantic roles. The fact that FrameNet roles are frame-specific means that the same role does not appear in different frames. Therefore, the number of frame elements identified by FrameNet is open-ended and unconstrained. VerbNet identifies a closed set of 23 *thematic roles* (e.g. AGENT, ATTRIBUTE, BENEFICIARY). In turn, PropBank establishes more than 1,400 roles. In this sense, PropBank's roles can be compared to the open-ended frame element roles depicted by FrameNet. In contrast, the Spanish resource ADESSE has a reduced set of roles for each of their semantic classes (e.g. INITIATOR, MOVING ENTITY, ORIGIN, DIRECTION, PATH, for the class of MOVEMENT). Along with the roles bound to each specific class, ADESSE regards certain roles not to be class-bound, but as belonging to various classes, such as the general roles of BENEFICIARY, LOCATION, ATTRIBUTE, INITIATOR, POSSESSOR, MEDIUM, FEATURE, REFERENCE, ROLE, CAUSE, PURPOSE, FINAL STATE, and MANNER. The other Spanish meaning base resource, SenSem, distinguishes twenty semantic roles (e.g. AGENT, EXPERIENCER, INITIATOR, THEME and TIME), some of which can be further subdivided into more specific ones.

In line with the words of Ureña, Faber, and Buendía (in press), every predicate argument analysis should be based on a previously specified closed set of semantic roles, depending on the objectives being pursued:

> The main issue is how much is too much and how little is too little. If the inventory of roles is open-ended, then the roles have no general descriptive adequacy. If the inventory is too restrictive, then it is not sufficient to describe the complexity of meaning. Evidently, there should be an explicitly defined inventory of semantic roles for a subset of reality. This inventory should be well defined, though not overly restricted.

The set of semantic roles in our study largely coincides with the most general thematic relations provided by Role and Reference Grammar and the argument roles as proposed by Goldberg (§3.1.3.3). Others were taken for the inventory proposed by VerbNet, ADESSE and Sensem. As previously mentioned, in our opinion, the FEs as proposed by FrameNet, although informative, have the disadvantages of being open-ended and unconstrained. It is thus less systematic since the tagging of FEs relies entirely on the intuition of the analyzer. Consequently, sometimes the same role label is shared in different frames but corresponds to a different role. As such, our inventory of roles is more restrictive than the FEs in FrameNet or the rolesets in PropBank. Table 55 is a non-exhaustive list of our inventory of semantic roles for the particular subfield of the EXTREME EVENT. A complete description of each semantic role is, however, found in 5.4.2.1.2.

Semantic roles
AGENT
DESTINATION
EXPERIENCER
FREQUENCY
GEOGRAPHICAL LOCATION
MANNER
NATURAL FORCE
ORIGIN
PATH
PATIENT
RESULT
SITUATION/EXPERIENCE
THEME
TIME

Table 55. Set of semantic roles in our verb analysis

Along with the semantic roles associated with each argument, an additional categorization in terms of macroroles, as proposed by RRG, is also provided. In other words, each argument is assigned a semantic role and when possible, a macrorole (i.e. ACTOR or UNDERGOER) as well.

Apart from role characterization, we have also annotated each argument with what we have referred to as *semantic tagging* or *labeling* (§5.4.2.1.1), which is a more specific characterization of its meaning, and which points out to a specific category. For example, in the sentence 'The tropical storm evolved into a hurricane', the first argument (i.e. 'tropical storm') is associated with the semantic role of THEME, the macrorole of UNDERGOER, but at the same time it is given the semantic label of NATURAL DISASTER. This semantic labeling process is explained in 5.4.2.1.1. Finally, the morphosyntactic structure associated with each argument has been specified in our analysis.

This analysis shows the constructions in which the verb can participate. Similarly to VerbNet that links semantic roles to their syntactic category, we link the linguistic realizations of specialized concepts to their conceptual category, associate semantic roles with macroroles, and provide the morphosyntactic descriptions of terms. As such, for instance the construction instantiated in 'The tropical storm evolved into a hurricane' is the following:

1st argument		2nd argument
THEME		RESULT
UNDERGOER	evolve	_____
NATURAL DISASTER		NATURAL DISASTER
Noun phrase		Prepositional phrase (into)

Table 56. Construction activated in the sentence 'The tropical storm evolved into a hurricane'

Finally, usage examples of use are included in our templates since this is a component that is included in most of the resources described.

4. MATERIALS AND METHODS

> […] Compilers make the best corpus they can in the circumstances, and their proper stance is to be detailed and honest about the contents. From their description of the corpus, the research community can judge how far to trust their results, and future users of the same corpus can estimate its reliability for their purposes (Sinclair 2005: para. 15)

4.1 Object of study: NATURAL HAZARDS or the EXTREME EVENT

The concepts of ENVIRONMENT and ECOLOGY are very often used interchangeably; though the scope of ENVIRONMENT is much broader than that of ECOLOGY. In this regard, the European Environment Information and Observation Network[117] (EIONET), defines ENVIRONMENT as follows[118]:

> A concept which includes all aspects of the surroundings of humanity, affecting individuals and social groupings. The European Union has defined the environment as "the combination of elements whose complex interrelationships make up the settings, the surroundings and the conditions of life of the individual and of society, as they are or as they are felt". The environment thus includes the built environment, the natural environment and all natural resources, including air, land and water. It also includes the surroundings of the workplace.

In contrast, ECOLOGY is regarded as a subdomain of ENVIRONMENT, and is defined as follows[119]: "The study of the interrelationships between living organisms and their environment". It is worth mentioning that EOINET is a partnership network of the European Environment Agency[120] (EEA). It is in charge of the collection and organization of data and the development and dissemination of the information concerning Europe's environment. One of the tools developed by the EEA with the support of EOINET is the *GEneral Multilingual Environmental Thesaurus*[121] (GEMET).

This international focus on the environment is relatively recent. It stems from the rising concern about the severe environmental degradation due to industrialization and certain consumer habits, as a result of industrial and technological growth that began in

[117] <http://www.eionet.europa.eu/> [12/11/2011].
[118] <http://www.eionet.europa.eu/gemet/concept/2944> [12/11/2011].
[119] <http://www.eionet.europa.eu/gemet/concept/2470> [12/11/2011].
[120] <http://www.eea.europa.eu/> [12/11/2011].
[121] <http://www.eionet.europa.eu/gemet/> [12/11/2011].

the 20th century. As such, the ENVIRONMENT is a relatively new conceptual domain, which is multidimensional and dynamic with an enormous social impact (Alonso and DeCesaris 2007: 71)**.**

Because of its relatively recent appearance, environmental terminology has not been standardized and is characterized by a significant degree of variation. In addition, its multidimensional nature signifies that many of the terms also originally come from other domains. Since environmental terminology is dynamic, the terms evolve and become polysemic. Finally, its social impact often generates a determinologization process that projects environmental terminology into general language. This is only natural because the environment is everyone's concern, and even a lay person may become interested at some point in certain aspects of the domain (Alonso and DeCesaris 2007: 71). For example, in a geographic area where hurricanes are frequent, (e.g. subtropical zones such as Florida), the inhabitants are familiar with the characteristics of this type of tropical cyclone. They also know how the intensity of the storm is mentioned and are familiar with the Saffir-Simpson hurricane scale, which goes from 1 to 5. In contrast, the inhabitants of Spain, where earthquakes occasionally occur, are more familiar with terms such as 'epicentro' ['epicenter'], 'hipocentro' ['hypocenter'], 'Escala de Richter' ['Richter scale']. These terms can be considered as part of the general language lexicon.

As is well known, the Earth is currently experiencing an increasing number of disasters due to both natural hazards and human-induced accidents, and therefore, there is a rising interest and concern for the environmental domain. However, to date, there have been few terminological studies specifically centered on natural hazards. Within this context, we decided to focus our analysis on this subdomain, which we refer to as the EXTREME EVENT frame, following the conception of events in Frame-based Terminology (§1.4.3).

As previously mentioned in 1.4.3, in Frame-based Terminology, the configuration of specialized domains is based on the events that generally take place in them. For the domain of the environment, the Environmental Event (EE) (see Figure 5 in §1.4.3) is a representation of the most generic conceptual level and underlies the organization of all concepts in the domain. The EE is composed of several subevents, one of which is the EXTREME EVENT in its sense of natural disaster.

We began working on the design of EXTREME EVENT by giving an informal description of the type of situation activated by the event. This was done by recurring to

"armchair linguistics" (Fillmore et al. 2003: 299) as well as the use of specialized knowledge resources. In other words, intuition, dictionaries, thesauri and specialized texts were initially used to design the EXTREME EVENT frame.

In this regard, we first looked up the concept of NATURAL DISASTER in EuroVoc[122], the multilingual thesaurus of the European Union. Table 57 displays the screenshot of the organization of this concept. As shown, NATURAL DISASTERS in EuroVoc include AGRICULTURAL DISASTERS, BAD WEATHER, CYCLONES, DROUGHTS, EARTHQUAKES, EROSIONS, FLOODS and VOLCANIC ERUPTIONS. BAD WEATHER is the generic concept for COLD WAVE, FROST, HAIL, STORM and THUNDERSTORM, and CYCLONE is the generic concept for TYPHOON, HURRICANE and TORNADO.

Natural disaster
UF *natural catastrophe*
52 ENVIRONMENT MT 5216 deterioration of the environment BT1 degradation of the environment NT1 agricultural disaster NT1 bad weather NT1 cyclone NT1 drought NT1 earthquake NT1 erosion NT1 flood NT1 volcanic eruption

Table 57. Extract of EuroVoc displaying the NATURAL DISASTER types[123]

Along with EuroVoc, the organization of NATURAL DISASTERS provided by *ScienceDaily*[124], one of the Internet's most popular science news websites, was also consulted. As shown in Table 58, the classification of NATURAL DISASTERS in *ScienceDaily* includes EARTHQUAKES, STORMS, LANDSLIDES, NEAR-EARTH OBJECT IMPACTS, SEVERE WEATHER CONDITIONS (e.g. EL NIÑO and LA NIÑA, FLOODS, HURRICANES and CYCLONES, STORMS and TORNADOS), SNOW and AVALANCHES, TSUNAMIS, and VOLCANOES.

[122]<http://eurovoc.europa.eu/drupal/> [08/01/2012].
[123]<http://eurovoc.europa.eu/drupal/?q=requestandconcepturi=http://eurovoc.europa.eu/413andtermuri=http://eurovoc.europa.eu/218606andlanguage=enandview=ptandifacelang=es> [08/01/2012].
[124] <http://www.sciencedaily.com/> [14/01/2012].

Natural Disasters
- Earthquakes
- Geomagnetic Storms
- Landslides
- Natural Disasters
- Near-Earth Object Impacts
- Severe Weather
  - El Niño and La Niña
  - Floods
  - Hurricanes and Cyclones
  - Storms
  - Tornadoes
- Snow and Avalanches
- Tsunamis
- Volcanoes

Table 58. Extract of *ScienceDaily* website index

Finally, the typification of hazards provided by the European Environment Agency (EEA 2010: 19) (Table 59) also contributed to the design of our EXTREME EVENT especially given the fact that the source of the EEA terminology is the United Nations Office for Disaster Risk Reduction (UNISDR).

Although the terminological unit 'natural disaster' is used worldwide, it is not entirely correct since disasters only happen because they are perceived as such by humans. Nature has nothing to do with the classification of this type of event (EEA 2010: 18). This is the reason why the EEA prefers to refer to these phenomena as 'natural hazards' (EEA 2010: 18). In contrast, a DISASTER is defined by the (UNISDR 2009: 9) as follows:

> A serious disruption of the functioning of a community or a society causing widespread human, material, economic or environmental losses that exceed the ability of the affected community or society to cope using its own resources.

Therefore, a DISASTER can often be described as the result of a HAZARD, which is conceived as follows (UNISDR 2009: 17):

> A dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage.

According to the UNISDR (2009: 21), a NATURAL HAZARD is a subset of all hazards and is defined as follows (UNISDR 2009: 21):

> Natural process or phenomenon that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage.

When humans are affected, the term 'disaster' is used. Therefore, it is the NATURAL HAZARD that causes the human DISASTER. However, the term 'disaster' is often used in a less strict sense for events that cause great damage, destruction, and human suffering (EEA 2010: 18). It can also be the case that a NATURAL HAZARD does not have a significant effect on ecosystems and could thus be considered simply to be a NATURAL DISTURBANCE (EEA 2010: 20).

As shown in Table 59, according to the EEA, a NATURAL HAZARD can be hydrometeorological, geophysical, or technological, depending on whether its origin is (i) atmospheric, hydrological or oceanographic (HYDROMETEOROLOGICAL HAZARDS); (ii) geological or physical (GEOPHYSICAL HAZARD); (iii) technological or industrial conditions (TECHNOLOGICAL HAZARDS).

Hazard	**Hazard type**
Storms	Hydrometeorological
Extreme event temperatures	Hydrometeorological
Forest fires	Hydrometeorological
Water scarcity and droughts	Hydrometeorological
Floods	Hydrometeorological
Snow avalanches	Geophysical
Landslides (incl. debris flow)	Geophysical
Earthquakes/ volcanoes	Geophysical
Oil spills	Technological
Industrial accidents	Technological
Toxic spills	Technological

Table 59. Typology of hazards. Adapted from EEA (2010: 19)

The focus of this thesis is on both HYDROMETEOROLOGICAL and GEOPHYSICAL HAZARDS (following the EEA terminology). Artificial processes, such as TECHNOLOGICAL HAZARDS (e.g. oil spills, industrial accidents, toxic spills), are not included within the scope of our study.

Nevertheless, we decided not to distinguish between HYDROMETEOROLOGICAL and GEOPHYSICAL HAZARDS. Even though the origin of certain processes is very clear (e.g. EARTHQUAKES are GEOLOGICAL PROCESSES), for many hazards, the boundaries between hydrometeorological and geophysical are extremely fuzzy because both phenomena are closely related. For example, this is the case of TSUNAMI (UNISDR 2009: 17-18):

> Tsunamis are difficult to categorize; although they are triggered by undersea earthquakes and other geological events, they are essentially an oceanic process that is manifested as a coastal water-related hazard.

Even the UNISDR and the EEA are not consistent in categorizing a HAZARD as hydrometeorological or geophysical. For example, Table 59 shows that the EEA classifies AVALANCHES as GEOPHYSICAL HAZARDS (as reflected in their terminology) but they also include AVALANCHES within the definition of HYDROMETEOROLOGICAL HAZARDS (EEA 2010: 18):

> Hydrometeorological hazards include tropical cyclones (also known as typhoons and hurricanes), thunderstorms, hailstorms, tornados, blizzards, heavy snowfall, avalanches, coastal storm surges, floods including flash floods, drought, heatwaves and cold spells.

Since it is true that both kinds of origins are closely linked, our study is on NATURAL HAZARDS without further subclassifications. Based on the premises of Frame-based Terminology, these phenomena are the components of a subevent within the general Environmental Event, which is known as the EXTREME EVENT. The EXTREME EVENT includes VOLCANOES, LANDSLIDES, FLOODS, AVALANCHES, TROPICAL CYCLONES (including TYPHOON and HURRICANES), TORNADOES, DROUGHTS, EARTHQUAKES, FIRES, and TSUNAMIS.

4.2 Design, compilation and description of the corpus

This section describes the English and Spanish corpora compiled for our verb argument analysis. After providing an overview of the concept of CORPUS (§4.2.1.1) and types of corpora (§4.2.1.2), we explain the methods and criteria for corpus compilation (§4.2.2). As part of our study, we designed a protocol for the evaluation of online resources that assured the reliability and quality of the texts of our corpus (§4.2.2.1.2). Section 4.2.3 describes the characteristics of our corpus and the software applications used (§4.3).

4.2.1 Basic notions of corpus

4.2.1.1 The concept of corpus

In general language, a corpus is simply a group of texts. As such, the *Diccionario de la Real Academia Española* (DRAE) (DRAE 2001) defines a corpus as "un conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que puedan servir de base a una investigación" [a set of data or texts (scientific, literary, etc.) as large and as well organized as possible which can be used as the basis for a research study].

However, the notion of corpus in Corpus Linguistics is somewhat more complex. There are many authors who have provided their own vision of a corpus. Generally speaking, not every collection of texts can be referred to as *corpus*. In line with this, Atkins, Clear and Ostler (1992: 1) classify text collections into four types: (i) archive; (ii) electronic text library, (iii) corpus; (iv) subcorpus.

An *archive* is a repertory of texts in electronic form in which the texts do not reflect any type of relation or coordination. For example, the *Oxford Text Archive*[125] is one of the largest text compilation centers aimed at research and teaching, including more than 2,000 titles and 450 independent collections of both oral and written texts in more than 25 languages. An *electronic text library* (ETL) is a collection of texts in electronic and standard format, which is not based on linguistic criteria. A *corpus* is "a subset of ETL built according to explicit criteria for a specific purpose" (Atkins, Clear and Ostler 1992: 1). From this definition, it can be inferred that a corpus is a text collection which in contrast to the other two types, has been compiled based on linguistic criteria to exemplify the behavior of one or more languages. Finally, a

[125] <http://ota.ox.ac.uk/> [02/07/02001].

subcorpus is a portion of a corpus, which can be either a static component of a larger corpus or a dynamic selection made while the corpus is consulted online.

In our opinion, the definition of *subcorpus* provided by Atkins, Clear and Ostler does not specify whether a subcorpus should contain the same number of textual genres as the larger corpus (i.e. a small-scale version) or on the contrary, whether it can be any portion of the larger corpus. In this regard, Sinclair states that a corpus can be divided into subcorpora and that both a corpus and a subcorpus can be further subdivided into *components.* He defines a subcorpus as follows: "[…] having all the properties of a corpus but happens to be part of a larger corpus" (Sinclair 1995: 17). According to this definition, a *subcorpus* should contain all the properties of a larger corpus in order to be representative of the larger corpus of which it is a subdivision. In contrast, a *component* is an example of a particular type of language and is selected "according to a set of linguistic criteria that serve to characterize its linguistic homogeneity" (Sinclair 1995: 17). A component differs from a subcorpus in that it is not intended to be representative of the corpus to which it belongs, and as such, it does not have to be an adequate sample of a language.

Francis (1982:7) was one of the first linguists to provide a definition of corpus:

> A collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis.

Sinclair (1991: 171) initially defined a corpus as shown below:

> A collection o naturally-occurring language text, chosen to characterize a state or variety of a language.

In turn, Engwall (1992: 167) underlined that for him a corpus should be the following:

> A closed set of texts in machine-readable form established for general or specific purposes by previously defined criteria.

Some years later, in the EAGLES *(Expert Advisory Group on Language Engineering Standards)* report, EAGLES (1996a: 4) enhanced his definition of corpus:

> A collection of pieces of a language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.

Subsequently, Sinclair added *electronic* to his definition (John Sinclair 2005):

> A collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.

Zanettin, Bernardini, and Stewart (2003: 1) may fall a bit short in defining a corpus simply as "principled collections of texts in electronic format". McEnery and Wilson (1996: 23) coincide in the fact that a corpus is a collection of real texts selected in order to be representative of a specific language or language variety. The only difference is that instead of *electronic*, they refer to the texts as *machine-readable* (McEnery and Wilson 1996: 21):

> In principle, any collection of more than one text can be called a corpus […]. But the term "corpus" when used in the context of modern linguistics tends most frequently to have more specific connotations than this simple definition provides for. These may be considered under four main headings: sampling and representativeness, finite size, machine-readable form, a standard reference.

Bowker and Pearson's definition of corpus includes the following (2002: 19):

> A corpus is a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria.

In addition, they compare the notions of *printed text* and *corpus* and maintain that these terms cannot be used interchangeably. Table 60 summarizes the main differences between a corpus and a written text. Items 1-6 highlight the advantages of a corpus as compared to a written text, and item 7 is its only shortcoming (Bowker and Pearson 2002: 9-21):

	Corpus	**Printed text**
1	An electronic corpus does not have to be read entirely. Corpus analysis tools help to find specific sections.	A collection of printed texts has to be read from beginning to end.
2	A corpus is compiled and consulted more quickly, thanks to the web and corpus analysis tools.	Printed texts are compiled and consulted less quickly. Users must go to the library and spend hours photocopying.
3	The size of an electronic corpus is larger.	The size of a collection of printed material is smaller.
4	A corpus of electronic texts can be updated automatically.	Printed material is more difficult to update since it is on paper.
5	Frequency information is obtained automatically..	Frequency information is obtained manually.
6	Learners find working with electronic resources more enjoyable and exciting.	Learners find working with printed material more tedious and time-consuming.
7	Corpus texts must be carefully assessed for reliability and quality since anyone can publish on Internet.	Printed texts are carefully evaluated and revised before they are published.

Table 60. Differences between a printed text and a corpus

Table 61 displays a summary of all the definitions of corpus provided in this section. Our definition of corpus is based on their shared features.

	A collection of		Format		Criteria	Representative
	Texts	**Pieces of language**	**Electronic**	**Machine readable**		
Francis (1982)	✓					✓
Atkins, Clear and Ostler (1992)	✓		✓		✓	
Engwall (1992)	✓			✓	✓	
Bowker and Pearson (2002)	✓		✓		✓	
Zanettin, Bernardini and Stewart (2003)	✓		✓			
Sinclair (2005)		✓	✓		✓	✓
McEnery and Wilson (1996)	✓			✓		✓

Table 61. Definitions of corpus: shared features

As can be observed, Sinclair's definition of corpus (2005) is the most complete. Consequently, in order to be able to refer to a group of texts as a *corpus*, the following three conditions are necessary:

(i) A corpus must be a collection of texts or pieces of text.

(ii) The texts should be in electronic format.

(iii) Texts or pieces of texts should be selected according to specific criteria in order to be representative of the language.

The first item in the definition of a corpus refers to a collection of texts or pieces of a text. By including complete texts, worries about differences between the parts of a text disappear, and it is always possible to randomly select smaller samples from a larger text collection (Sinclair 1991: 19). However, as Sinclair (2005: section 12) himself suggests, "some corpora still use sample methods rather than gather complete texts or transcripts of complete speech events". He refers to these samples as *pieces of language*. Consequently, as shall be seen in 4.2.3, our corpus includes complete texts as well as extracts of texts.

In addition, the texts must be in electronic format so that they can be easily processed with corpus analysis tools (see section 4.3 on the corpus analysis tools used in this research). As shall be seen, the majority of texts in our corpus come directly from the web. However, in our corpus there are also a large number of books that have been scanned or which have been entered as plain text files (.txt).

Finally, in order to be representative of the language, the selection of candidate texts for the corpus should be based on a set of specific criteria. Bowker and Pearson (2002: 10) point out that one cannot start downloading texts from the web at random and refer to that collection as *corpus:*

> […] a corpus is not simple a random collection of texts, which means that you cannot just start downloading texts haphazardly from the Web and then call your collection a 'corpus'.

Biber (1993: 243) defines representativeness as "the extent to which a sample includes the full range of variability in a population", and subsequently identifies a series of external and internal criteria for the compilation of a corpus. This distinction between internal and external criteria was also adopted by authors such as Atkins, Clear and Ostler (1992), Nakamura (1993), and Bowker (1996), *inter alia*.

External criteria are predominantly extralinguistic (e.g. participants, communicative function, situational context, gender, modality, origin and finality of the texts). In contrast, internal criteria refer to the linguistic patterns of texts. Generally speaking, internal criteria are more objective than external criteria since they are based on the texts themselves and depend entirely on the linguistic data included in the texts. According to the majority of authors, the compilation of a representative corpus should be based on both external and internal criteria:

> Both external and internal criteria are therefore to be considered in the classification of texts. A typology of texts for inclusion in corpora cannot be based entirely on external or internal criteria (EAGLES 1996b).

> A corpus selected entirely on internal criteria would yield no information about the relation between language and its context of situation. A corpus selected entirely on external criteria would be liable to miss significant variation among texts since its categories are not motivated by textual (but by contextual factors) (Atkins, Clear and Ostler 1992: 5).

There are other authors, nevertheless, who state that texts should be only selected according to external criteria. This is the case of Sinclair (2005). In our opinion, this is justified by the fact that the initial selection of texts is almost completely based on external criteria, whereas internal criteria become important in a subsequent phase when linguistic aspects are evaluated with corpus analysis tools.

4.2.1.2 Types of corpus

The lack of consensus regarding the concept of corpus can be extended to corpus typology. As such, there is still considerable disagreement as to the types of corpus and how they should be classified. In the words of Bowker and Pearson (2002: 11):

> There are almost as many different types of corpora as there are types of investigations. Language is so diverse and dynamic that it would be hard to imagine a single corpus that could be used as a representative sample of all language.

Authors that propose standardized corpus typologies include the EAGLES (1996b), Teubert (1996), Austermühl (2001), Corpas (2001b), Bowker and Pearson (2002), and Laviosa (2002), *inter alia*. Based on the EAGLES reports and Corpas (2001b) and Laviosa (2002), this section gives a schematic classification of corpus

typology to situate and describe our corpus with the following parameters: (i) way of transmission; (ii) languages; (iii) quantity of text; (iv) terminological specificity; (v) time period; and (vi) metadata or annotation.

Way of transmission

The way of transmission involves two kinds of discourse mode, namely, oral and written. Consequently, there are two kinds of corpus, i.e. an oral corpus and a written corpus.

An oral corpus includes informal and spontaneous conversations (EAGLES 1996b). However, as underlined in the EAGLES report, boundaries are not always easy to establish since not everyone understands the concepts of spontaneity and informality in the same way. In their own words, they question "[…] how impromptu is impromptu, how informal is informal" (EAGLES 1996b). In contrast, there are other authors who use the label of *oral corpus* to refer to any kind of language in which the speakers behave orally. As such, texts written to be spoken are considered to be an oral corpus for these authors. In our opinion, however, an oral corpus is a corpus composed of recordings or the transcription of the spoken language. Therefore, a written text read out loud will not be regarded as an oral corpus since it does not contain the specific characteristics of an oral corpus.

A written corpus, as its name implies, is a corpus fully formed by written texts. Since the two corpora compiled for our research purposes are formed entirely of written texts, there is no need to further describe oral corpora. Consequently, the typology proposed is for written corpora.

Languages

Regarding the number of languages, a corpus can be classified as monolingual or multilingual, and a multilingual corpus can be further subdivided into parallel and comparable. As its name suggests, a *monolingual corpus* is a corpus that is only composed of written texts in one language. Monolingual corpora are very useful for translators in that they offer information about lexical units in the target language which can help translators to adopt semantic shifts and collocations imitating native speakers (Zanettin, Bernardini and Stewart 2003: 6).

A *multilingual corpus* is a corpus of written texts in two or more languages. As previously mentioned, a multilingual corpus can be further subdivided into a parallel

corpus or comparable corpus. Generally speaking, a *parallel corpus* is a corpus that contains texts in the source language and their translations into a target language. According to Bowker and Pearson (2002: 12), a parallel corpus can be *bilingual* if it contains texts and their translations into one language, and *multilingual* if it contains the original source texts and their translations into two or more languages. According to Teubert (1996: 245), a parallel corpus corresponds to one of the following: (i) original texts written in a language A and their translations into languages B, C, etc.; (ii) the same number of original texts written in languages A and B and their translations; (iii) only texts translated into languages A, B and C, where the original texts are written in language Z.

A parallel corpus fulfils a basic role in translation studies since it can provide students with evidence of how translators overcome difficulties during translation practice (Baker 1995: 231). As such, by observing strategies that other translators have adopted, students can start to develop their own strategies of translation. The most well-known parallel corpus comes from official bilingual organisms, where the majority of documents must be published in all the official languages of the community (e.g. the Canadian Parliament). In this regard, one of the largest parallel corpus is *The Open Parallel Corpus (OPUS)*[126], composed of a growing collection of translated texts from the web in more than 60 languages.

A *comparable corpus* is a corpus that contains texts of the same genre and deals with the same topic in two or more languages, but that are not translations of each other (Bowker and Pearson 2002: 93). The label *comparable* indicates that the texts in the various languages have been selected because they have common features, such as topic, text genre, time when they were written, or level of specificity (Peters, Picchi and Biagini 1996: 69). In a similar way as occurred with a parallel corpus, a further distinction is made between *comparable bilingual corpus* (when including original texts in two languages), and between *multilingual comparable corpus* to refer to a corpus which includes original texts in three or more languages.

As shall be seen in 4.2.3, our corpus is in English and Spanish. In other words, it is composed of two subcorpora in English and Spanish, which include original texts of the same type and length, belonging to the same domain (i.e. NATURAL HAZARDS or the EXTREME EVENT), but which are not necessarily translations of each other.

[126] <http://opus.lingfil.uu.se/> [02/07/2011].

Quantity of text

Regarding the quantity of text in a corpus, there should be a distinction between *full text corpus* and *sample corpus*. According to Laviosa (2002), a *full text corpus* includes complete texts with the objective of offering a representative sample of the language. Needless to say, when the corpus is composed of complete texts, worries about the differences between the various parts of a text disappear and it is always possible to select at random smaller samples from a large collection of text (Sinclair 1991: 19). Corpas (2001b) refers to this type of corpus as *corpus textual* [textual corpus].

Another type of full text corpora is a *monitor corpus*, which is a corpus composed of complete texts without any length restrictions and which updates constantly (Laviosa 2002: 34). Sinclair (1991: 103). This type of corpus emerged thanks to unlimited storage capacity and improved electronic processing of computers (EAGLES 1996b). Bowker and Pearson (2002: 12-13) refer to this kind of corpus as *open corpus*. Most authors regard a monitor corpus as an independent type of corpus. In our opinion, nevertheless, this type of corpus is a subtype of *complete text corpora* since it includes complete texts with the peculiarity that it is being continuously updated with new documents.

A *sample corpus* (Sinclair 1991; Laviosa 2002: 34), also referred to as *corpus léxico* [lexical corpus] (Corpas 2001b: 157-159) or *corpus de fragmentos textuales* [corpus of text fragments] (Pérez 2002), includes extracts of texts, usually of the same length. However, today almost all corpora include complete texts. In fact, the practice of including samples from texts has now become a relic of the past linked to the limitations in the storing capacity of computers. Examples of corpus formed by sample texts rather than complete texts are the Brown Corpus[127] and the Lancaster-Oslo/Bergen Corpus of British English (the LOB corpus)[128].

As previously mentioned, our corpus includes both complete texts as well as text extracts, such as book chapters or sections. In other words, our corpus includes books on the environment as well as book sections on natural disasters.

[127] <http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/index.html> [12/11/2011].
[128] <http://khnt.hit.uib.no/icame/manuals/lob/index.htm> [12/11/2011].

Specialization of the language

The level of specialization of a language is a further criterion of corpus classification as reflected in the following types: (i) general corpus; (ii) (general) reference corpus; (iii) special corpus.

A *general corpus* is a corpus that focuses on language for general purposes (LGP), i.e. "language used by ordinary people in everyday situations" (Bowker and Pearson 2002: 12). A *(general) reference corpus* is a general corpus that is representative of a given language and which can therefore be used to make general assumptions about a language (Bowker and Pearson 2002: 12). Accordingly, it is usually composed of millions and millions of words from both written and oral texts and a wide variety of text types. It is important to emphasize here that a reference corpus serves as a basis for the elaboration of grammars, dictionaries, thesauri, and other reference works. Some of the most well known general reference corpora are the *British National Corpus* (BNC)[129] for English, and the *Corpus de Referencia del Español Actual* (CREA)[130] for Spanish.

In contrast, a *special corpus*, also referred to as *specialized corpus* or *special corpus purpose* (Bowker and Pearson 2002), can be defined as follows:

> Those which do not contribute to a description of the ordinary language, either because they contain a high proportion of unusual features or their origins are not reliable as records of people behaving normally (Sinclair 1995: 23).

In other words, a special corpus is constrained to a particular aspect of language, i.e. the LSP of a particular field, a specific text type, a particular language variety, or to the language used by members of a specific group, such as teenagers (Bowker and Pearson 2002: 12). Although the terms *specialized corpus* and *special corpus* are generally regarded as synonymous, the term *special corpus* is preferred. Initially, Sinclair (1987: xvi) used the notion of *specialized corpus* to refer to a set of small corpora designed with various purposes in mind. However, in later publications, *specialized corpus* was not included in his corpus typology (Sinclair 1995). This seems to indicate that *specialized corpus* was included in *special corpus.*

[129] < http://www.natcorp.ox.ac.uk/> [12/11/2011].
[130] < http://corpus.rae.es/creanet.html> [12/11/2011].

The two corpora compiled for this research are special corpora belonging to the environmental domain with texts that focus on the subdomain of the EXTREME EVENT (§4.1). However, as previously mentioned, the social impact of the environment and the fact that extreme events have high media impact means that at some point, specialized environmental terms become a part of general language.

Time period

Based on the time period of the texts, a corpus can be diachronic or synchronic. A diachronic corpus includes texts from different time periods compiled with the objective of studying the evolution of a language over a period of time (e.g. the Brown Corpus[131] or the LOB corpus[132]).

In contrast, a *synchronic corpus*, also referred to as *periodical corpus or chronological corpus* (Corpas 2001b: 158), is a corpus that contains texts produced during a limited time period (e.g. the Collins Cobuild Corpus[133]). As shall be seen in 4.2.3, the corpus compiled for our research is synchronic since it includes texts from 1996 to 2012.

Codification

Codification refers to whether the texts in the corpus are kept in their original form, i.e. in the same way as when downloaded from the web, or if they have been tagged or enriched with some kind of linguistic or metatextual information. Depending on whether there has been some sort of codification process, a corpus is a *plain corpus* or an *annotated corpus.*

As its name suggests, a *plain corpus*, also referred to as a *non-annotated corpus* (Corpas 2001b: 159) or *unannotated corpus* (McEnery and Wilson 2004: 73), is a corpus formed by *plain texts.* Sinclair (2004: 48) describes plain text as follows:

> Imagine that you had a long thin reel of paper to write on rather a rectangular sheet-like a reel of sticky tape but made of paper. You have in front of you a piece of writing that you want to record onto this reel of paper- just a paragraph. How would you do it? I expect that you would ignore line ends, remove hyphens that marked words split at line-ends, and

[131] <http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/index.html> [12/07/2011].
[132] <http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM> [12/07/2011].
[133] < http://www.mycobuild.com/about-collins-corpus.aspx> [12/07/2011].

> otherwise produce a continuous stream of letters, numbers and punctuation marks in the same sequence as the original. That is plain text, and it consists of an alphanumeric stream.

Therefore, a *plain corpus* is a corpus formed by plain texts, i.e. documents that are in text-only format or ASCII, which have not been enriched with any kind of metalinguistic information. As is well-known, most corpus analysis tools (e.g. WordSmith Tools, Sketch Engine) and terminology software programs (e.g. TermoStat) only accept documents in text-only or txt format. As such, when texts are in pdf, Word or any other format, they have to be converted into txt format before loading the texts and running the analysis. In contrast, an *annotated* or *tagged corpus* (Austermühl 2001) is a corpus in which the texts have been linguistic or metatextually tagged, either manually or automatically (Corpas 2001b: 159).

Although plain corpora have been and still are extensively used for language analysis, that the usefulness of a corpus can be considerably enhanced when the texts are tagged. According to Leech (1997: 4) "The fact is that to extract information from a corpus, we often have to begin by building information in". That is exactly what annotation involves, enriching corpus texts with information so that implicit characteristics of textual elements will become explicit.

Succinctly put, annotation can be defined as "[…] the practice of adding interpretative linguistic information to a corpus" (Leech 2004). Initially, Leech (1997) considered the following types of annotation: orthographic, phonetic, prosodic, grammatical, syntactic, semantic, discourse and pragmatic/stylistic. However, in a later publication (Leech 2004), he established the following categories: grammatical, phonetic, semantic, pragmatic, discourse, stylistic and lexical. Accordingly, based on Leech's proposals, six types of annotation can be distinguished:

- **Part of speech tagging (POS**) identifies the different parts of a sentence, such as nouns, verbs or articles.
- **Lexical annotation or lemmatization** consists of adding the lemma identification of each word of a text. In English, it might be considered redundant, but in languages such as Spanish or German, it can be very useful for information extraction.
- **Syntactic annotation** consists of adding syntactic information to a corpus through the incorporation in the text of syntactic structure indicators.

- **Semantic annotation** consists of adding information about the semantic categories of words.
- **Discourse annotation** involves the addition of information about anaphoric links in a text.
- **Pragmatic annotation** consists of adding information regarding the kinds of speech act that occur in a spoken dialogue.

Figure 68 depicts the different types of annotation. As shown, part of speech and lexical annotation are the most basic types of annotation and are at the base of the pyramid. In contrast, pragmatic annotation refers to the production context of the speech act, and is thus at the top.

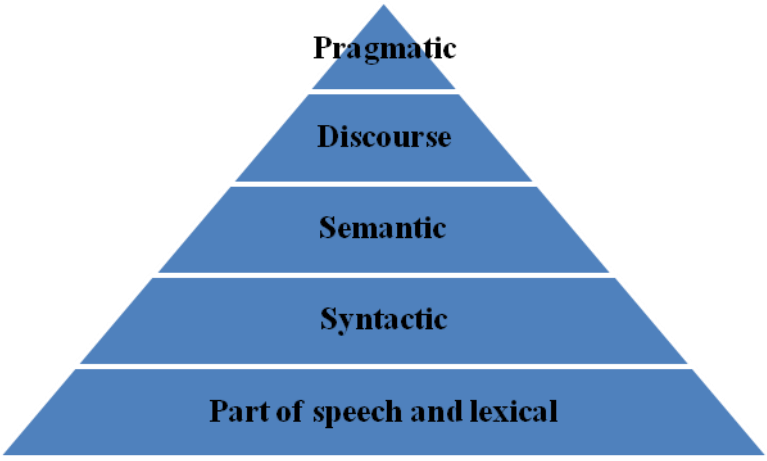


Figure 66. Levels of linguistic annotation

Annotation has a wide range of applications. For example, it makes extraction and recovery of information easier; it allows the re-usability of corpora; and it is multi-functional (Leech 1997, 2004).

In this study, no linguistic annotations were used since we focused more on the methodology for describing verbal collocates. However, sentences were annotated by using different colors to identify the arguments in consonance with FrameNet. In other words, each argument was labeled with a kind of semantic label, which corresponds to a semantic category previously defined, (e.g. 'debris of a tornado' was tagged with the semantic label of MATERIAL) with the view of grouping all verbs with arguments belonging to the same categories into the same frame. The underlying idea is that verbs with the same types of semantic argument patterns presumably activate the same kind

of semantic information. A more detailed description concerning argument analysis can be found in 5.4.2.1.1.

Summary

Table 62 summarizes the typology of corpus described in this section:

	Classification	**Type of corpus**
1.	Way of transmission	Written corpus
		Oral corpus
2.	Languages	Monolingual corpus
		Multilingual corpus -Parallel corpus -Comparable corpus
3.	Quantity of text	Full text corpora -Monitor corpus
		Sample corpus
4.	Specificity	General corpus -Reference corpus
		Special(ized) corpus
5.	Time	Diachronic corpus
		Synchronic corpus
6.	Codification	Plain corpus
		Annotated corpus

Table 62. Corpus classification summary

The corpus compiled for this research is a bilingual corpus composed of two written subcorpora, one in English and one in Spanish. It is a comparable corpus since its texts deal with the same topic (i.e. the subdomain of extreme event or natural hazards within the general domain of the environment), but which are not translations of each other. In this regard, it is a special corpus on the subdomain of natural hazards. It contains both full-text documents and samples of texts. However, our corpus can be referred to as a *full text corpora.* Even though it contains book chapters, these sections have a clear beginning and end and are of considerable extension. Regarding time period, our corpus is a *synchronic corpus* since it contains texts from 1996 to 2012. It is

also a plain corpus because it does not have any type of linguistic annotation though certain tags were added to enhance the retrieval and categorization of information.

4.2.2 Methods and criteria for corpus compilation

The compilation of a corpus is an important factor that greatly influences the results obtained in any research study. Many years ago, corpus compilation used to be an arduous process that required many hours spent in libraries. However, this is no longer true because, thanks to the Internet, hundreds of texts can be compiled in a few minutes.

The Internet has brought a new way of organizing and obtaining information. It provides the possibility of accessing any type of information at any time and at any place (Jiménez and Ortiz-Repiso 2007: 18). The amount of data circulating on the Internet on any given day is greater than all the information available in the 19th century (Austermühl 2001: 7). Because of the vast amount of information offered, the Internet constitutes *a fabulous linguists' playground* (Kilgarriff and Grefenstette 2003: 333). However, it is important to emphasize here that it has led to an *information overload*. In the words of Fuertes-Olivera and Niño-Amo (2011: 170-171):

> The biggest problem for a real information society is […] the fact that data cannot be found, or found in such great quantity that information stress or information death result both of which usually force potential users to abandon the search before finding the results.

English continues to dominate the web with 45% of the total number of web pages. Other European languages with a significant percentage of webpages are German (5.9%), French (4.41%), Spanish (3.8%), Italian (2.66%), and Portuguese (1.39%)[134]. Nonetheless, the growing number of Internet users who speak other languages is now in the process of changing the situation[135].

The main question that arises is whether the web should be regarded as a corpus itself. In this sense, there are two approaches to the web (De Schryver 2002): (i) *web for corpus (WfC)*, in which the web is used as a source of texts in digital format for the subsequent implementation of an offline corpus; (ii) *web as corpus (WaC)*, which uses

[134] These results correspond to the investigation carried out by the Union Latina (Latin Union). <http://dtil.unilat.org/LI/2007/es/resultados_es.htm> [05/07/2011].
[135] Internet World Stats. Usage and Population Statistics. <http://www.internetworldstats.com/stats7.htm> [05/07/2011].

the web directly as a corpus (Kilgarriff and Grefenstette 2003; Fletcher 2004, 2007; Baroni and Bernardini 2006).

Accordingly, we conducted a study with a view to comparing and evaluating these two approaches in the context of a scientific and technical translation course at university level (López and Buendía 2011). We asked two groups of students in the Translation and Interpreting Degree Program at the University of Granada to perform a technical translation assignment. One of the groups used the WaC approach, whereas the other group used traditional WfC methods. Our objective was to find out whether the Web as Corpus approach was able to compensate for the lack of subject field knowledge of one of the student groups. We wished to see if the quality of the translations of these students was more or less similar to that of the other group that had previously translated texts in the subject field. The results obtained showed that these two methods are complementary and that students should decide for one or the other, depending on their needs, (i.e. translation assignment, novelty of the translation, directionality and specificity of the translation, time allotted, or level of analysis required).

In line with this, our research analysis integrated both WfC and WaC methodologies. In other words, on the one hand, a corpus of texts in the subdomain of extreme event was used (i.e. a traditional *Web for Corpus* approach). On the other hand, various examples, contexts of use, and phrases were directly retrieved from the web (i.e. following a *Web as Corpus* approach) when the information derived from our corpus was not sufficient to satisfy our needs, when no examples were found or when we needed to offer a wider range of contexts.

4.2.2.1 The Web for Corpus approach

As previously mentioned, the Web for Corpus (WfC) is the approach that has been traditionally used to compile texts in digital format for the subsequent implementation of offline corpora. Authors such as Sinclair (2005) are clearly in favor of this traditional approach. Although Sinclair admits Internet's usefulness for linguists, he underlines the fact that the World Wide Web is not a corpus because it has not been defined from a linguistic perspective.

The WfC approach involves manually searching the web for valuable information. Users thus enter a list of keywords in a search engine (i.e. Google), which leads them to

other websites. Then, they select texts to download and process in a corpus analysis program, such as Wordsmith Tools (§4.3.2).

Although corpus design is evidently dependent on the research objectives being pursued, the quality and reliability of the texts is an important issue. The fact that Internet is currently the main source of texts for corpus compilation means that corpus quality is directly related to the quality of websites. For this reason, the selection of texts is crucial for the development of a representative corpus, which will be the source of reliable data. In the words of Austermühl (2001: 52):

> Finding data on the World Wide Web is no problem at all. But finding reliable information is rather a difficult task. And finding the information you really need can be very time-consuming and often frustrating.

In this regard, the following section discusses the main parameters that determine the quality of websites, based on which we propose our own set of evaluation parameters.

4.2.2.1.1 Evaluation of online resources

Since the quality of texts that make up a corpus greatly depends on the quality of the websites from which they are taken, it is first necessary to specify the differences between the terms *webpage, website*, and *homepage*, which are frequently used indiscriminately to refer to online documents. A *webpage* is every single document displayed on the screen. As such, the *website* of a company or organization can be made up of various *webpages.* The entrance gate to a *website* is usually called a *homepage.* A *homepage* is the starting point for navigation through a website (Austermühl 2001: 48). A *website* can thus be defined as a group of *webpages* with the same *homepage* that share internal hyperlinks and which constitute a kind of documental unity (Jiménez and Ortiz-Repiso 2007: 33).

As is well known, evaluation is intimately associated with quality since the ultimate objective of any evaluation is to establish the maximum level of quality. The term *quality* is often used to denote *good quality* or *high quality*. In relation to the information available via the Internet, *quality* refers to sources which are accurate and reliable (Cooke 1999: 14). In this sense, the ISO 8402-94 defines *quality* as "the set of

characteristics of an entity that give that entity the ability to satisfy expressed and implicit needs."

There is much written about the parameters that determine the quality of digital resources, but as yet there is no consensus of opinion. According to Sinclair (2005), any selection must be based on a certain set of criteria, and the first step in corpus building is the determination of the criteria upon which the texts of the corpus will be selected. Cooke (1999) establishes an inventory of ten parameters: purpose of a source, coverage, authority and reputation, accuracy, currency and maintenance of a source, accessibility of a source, presentation and arrangement of information, how easy a source is to use, and the overall quality of a source. Both Auer (1999) and Alexander and Tate (1999) underline the following five parameters: authority, coverage, objectivity, accuracy, and currency. Codina (2000) names six parameters: authority, content, accessibility, ergonomics, luminosity[136] and visibility, whereas the inventory of Jiménez and Ortiz-Repiso (2007) includes search and recovery, authority, content, resource administration and design. Seghiri (2006) provides set of criteria for the parameters of authority, content and design and ergonomics.

In this regard, Gordon-Murname (1999) examines the evaluation policies of twelve Web evaluation services and highlights the lack of consensus among review services. She concludes that the only parameter in which all of them coincide is *content*[137].

[136] *Luminosity* is a term coined by Codina (2000) and refers to the number of links a webpage has to other webpages.

[137] The twelve web evaluation services being analyzed are: *CyberStacks, Best Information on the Net (BIOTN), Librarian's Index to the Internet, Scout Report, Argus Cleringhouse, Blue Web'N, Dow Jones Business Directory, Finding Business Research on the Internet, The PH Directory of Online Business Information 1998, Lycos Top and Magellan/MackKinley SelectSurf.*

Criterion	Evaluators that include criterion on their list
Content	12
Design/Presentation/Format	11
Update Frequency	8
Audience/Community Needs	7
Currency/Timeliness	7
Rating System	7
Authority	5
Availability/ Speed	5
Value/Usefulness	5
Accessibility/Searchability	4
Scope	4
Cost	3

Table 63. Criteria proposed by the twelve evaluation services

4.2.2.1.2 Protocol for the Evaluation of On-Line Resources

If the number of websites consulted had been small, we could have analyzed them on the basis of the 509 criteria proposed by Wilkinson, Bennett and Oliver (1997). However, in large research projects, like ours, we require quality results that can be obtained rapidly. Thus, following Alexander and Tate (1999), Cooke (1999), Codina (2000), Seghiri (2006), Jiménez and Ortiz-Repiso (2007), our evaluation protocol is based on three parameters: (i) *authority*, which refers to the reputation and expertise of the authors; (ii) *content,* which includes coverage, accuracy, objectivity, currency, and audience; (iii) *design*, composed of navigational aids, accessibility and presentation and management.

Authority

The main difficulties inherent in the assessment of the quality of a particular piece of information on the web are its decontextualized nature and the lack of standardization affecting the presentation of information (Jiménez and Ortiz-Repiso 2007: 149). Therefore, the identification of the authors and their professional status and reputation are crucial in order to guarantee the scientific nature of the information. In the words of Alexander and Tate (1999: 11):

> Authority is the extent to which material is the creation of a person or organization that is recognized as having definitive knowledge of a given subject area.

An assessment of the authority of a document is based upon a range of factors, but primarily the knowledge and expertise of those responsible for producing it. Thus, a source is authoritative if it is written by an expert or is produced by an institution with recognized knowledge in the field. As Austermühl (2001: 64-65) states that the validity of an online source depends on the credibility of its author. Thus, an anonymous or "authorless" resource should be given little credibility. Sometimes, when a particular webpage is found with a search engine, it may be necessary to go back in the URL in order to obtain information about the author. Another way of checking the authority of a person or institution, when information is not provided in the webpage, is to look for this person's background on a search engine such as Google.

The address of a site also gives clues about the credibility of the authority. The following URL is an example of how a website is organized: http://lexicon.ugr.es/[138]. The last part "es" is referred to as *top-level domain (TLD)* and the other parts are referred to as *sub-domains*. The TLD "es" stands for "Spain"; the letters "ugr" stand for "University of Granada"; the word "lexicon" is the name of the research group; and the first part of the URL –http://– designates the communications protocol to be used to transfer the information requested. As is well known, 'http' is the standard protocol for transferring HTML documents on the Internet (Austermühl 2001: 46). Country domains normally use the codes assigned by the International Standardization Organization (ISO). The codes usually consist of the first two letters of the country's original name[139].

Top-level domains are intended to help categorize Internet resources. This is the reason why they allow us to accept or reject a page even before opening it. They provide us with essential information that may determine the reliability of a page. According to (Austermühl 2001: 47), the most common top-level domains are the following:

- **.com**: for commercial institutions (the most widely used TLD);
- .**mil**: for websites belonging to the US military;

[138] It corresponds to the website of our research group Lexicon.
[139] The Donald Bren School of Information and Computer Sciences of the University of California offers a list of all country codes. Available online: <http://ftp.ics.uci.edu/pub/websoft/wwwstat/country-codes.txt> [05/07/2011].

- **.net:** for companies or organizations that act as network providers or have to do with the administration of networks;
- **.edu**: originally for all educational entities in the United States (universities, colleges, schools, educational service organizations, etc.), but now restricted to four-year colleges and universities. Schools and two-year colleges are registered in the country domain, i.e. "us";
- **.gov**: for agencies of the U.S. Federal Government such as the State Department, the Senate, the White House or the Library of Congress;
- .**org**: used by organizations of various kinds, mostly international entities such as the United Nations;
- **.int:** for organizations established by international agreements.

In line with this, special care should be taken with personal home pages, whose typical TLD are .name, .members, .users, people, ~, %. Most of them do not offer more information than "this is my cat" or "this is where I live" (Cooke 1999: 10). Sometimes individuals create and develop a personal site within a particular job or university, but after a few years they may move on without removing the information. When users attempt to follow the links to search for information, they only encounter either *file not found* or an outdated page.

As shall be seen in 4.2.3, the online texts retrieved from the Web that conform our corpus belong to either the TLD .com (e.g. *National Geographic* or the *ScienceDaily* website), to that of .net (e.g the specialized journal *Natural Hazards*) or .org (e.g. the specialized journal *Natural Hazards Review*). No personal homepages were included.

Along with the address of the website, other indicators of an authoritative site may be if the site provides a way (i.e. email, address, telephone) to contact the organization, company, or person responsible for the contents of the site; and if it includes a feedback section in order to submit suggestions, complaints, or requests to the author of the website.

There are people who support the idea that the more visited a page is the more reliable it is. We did not take the visibility of a page into account in terms of evaluation because a site's popularity does not necessarily correspond to its quality. A clear example of this is the large number of web pages with sexual content. Although they are among the most visited on the web, this has little to do with their quality.

Finally, it is also important to mention copyright laws related to the use of electronic information and, more specifically, to information available via the Internet. The fact that authors provide a statement of the copyright ownership of materials and details of how materials should be cited in a publication or attributed to an author, as well as details of who should be contacted when copyright permission is required can also be an index of quality (Cooke 1999: 70).

Content

Within the parameter *content*, the notions of *coverage*, *accuracy*, *objectivity*, *currency*, and *audience* are taken into account. The *coverage* of a website refers to whether it contains a sufficient quantity of valid information regarding the topic.

Accuracy is in the words of Alexander and Tate (1999: 11) "[…] the extent to which information is reliable and free from errors." As such, it becomes necessary to analyze whether the texts on the website are free of grammatical, spelling, and typographical errors; and if there are references to other sources of information. However, as Sinclair (2005) recognizes, there are always errors in any corpus because what a user might perceive as perfectly accurate in a corpus is often just the opposite. In fact, perfect accuracy is likely to be systematically inaccurate. He points out that in a corpus of about a hundred million words, 99% accuracy means that there are more than a million errors.

Objectivity refers to presenting the information without distortion by personal feelings of other biases. Although it is very difficult to measure the objectivity of a piece of information, the answer to questions such as the following, might be useful to reflect on the possible objectivity or subjectivity of a website (Alexander and Tate 1999: 13): (i) if the point of view of the individual or organization responsible for providing the information is evident; (ii) if the page is free of advertising and if not, to what extent an advertiser might influence the information contents; (iii) in case the site has nonprofit or corporate sponsors, if they are clearly listed and there are links to them in order to learn more about them.

As proposed by Cooke (1999: 63), the *currency* of a source refers to "how up-to-date it is." The criteria we follow to determine the currency of a website is to check the date when the material was first created and the date of the last website update.

Finally, as is well known, the audience is "the target reader, the person for whom the author is writing" (Pearson 1998: 61). As Cooke (1999: 55) points out, the level of

detail or depth of coverage relates to the intended audience of a source. This is the reason why users will need to evaluate whether a source provides sufficient information, and whether the information is pitched at an appropriate level for their needs. For our purposes, we follow the typology established by Pearson (1998: 35-39) who distinguishes between *expert to expert communication*, *expert to initiates, relative expert to the uninitiated* and *teacher-pupil communication*. As previously described in 1.3.2, we merge these communication types into three communicative settings, namely, specialized communication, semi-specialized communication, and non-expert/lay communication. Therefore, our text selection was done with these three types of users in mind.

As shall be further described in 4.2.3, regarding specialized communication, articles from specialized online journals, such as *Natural Hazards*[140] and *Natural Hazards Review*[141] were extracted; for semi-specialized communication, textbooks on the domain of natural hazards were considered, i.e. the monographs *Natural Hazards* (Bryant 2005), *Natural Hazards and Environmental Change* (McGuire, Mason and Kilburn 2002), *At risk* (Wisner et al. 2004), and reports of the European Union, such as *The Terminology on Disaster Risk Reduction*, compiled by the United Nations International Strategy for Disaster Reduction (UNISDR 2009) and the report entitled *Mapping the impacts of natural hazards*, authored by the European Environmental Agency (EEA 2010). Regarding non-expert communication, the articles collected were from magazines, such as *National Geographic*[142] and *ScienceDaily*[143].

Design

As previously mentioned, design is becoming more and more important for determining the reliability of web contents. It is evident that the presentation and arrangement of information on the screen can greatly influence the ease of assimilating knowledge (Cooke 1999: 72). Within the *design* parameter, *navigational aids*, *accessibility,* and *presentation and management* are considered.

In the words of Alexander and Tate (1999: 50), *navigational aids* are "elements that help a user locate information at a website, and allow the user to easily move from

[140] <http://www.springer.com/earth+sciences+and+geography/natural+hazards/journal/11069> [18/06/2012].
[141] <http://ascelibrary.org/nho/> [18/06/2012].
[142] <http://www.nationalgeographic.com/> [18/06/2012].
[143] <http://www.sciencedaily.com/> [18/06/2012].

page to page within the site." The most important navigational aids that should be taken into account are site maps or indexes, hypertext links, the title, and if the website has an internal search engine and a help section.

As such, it becomes necessary to check whether the website has a site map or index on the home page or on a page directly linked to the home page. A *site map* is a display, often graphical, of the major components of a website, whereas an *index* is a listing, often alphabetical, of the major components of a website (Alexander and Tate 1999: 52). Both site maps and indexes provide a quick overview of the pages contained within the entire site, and therefore, can be an important tool in determining the coverage of the site and help to navigate through it.

Regarding hypertext links, it is interesting to check whether there are links on the website to enhance navigation, especially if there is a link to the home page, if there are links to the page one level up in the hierarchy (for sites arranged in a hierarchy), and if shortcuts are available. In reference to the title, it is helpful when it is short and only for that the site. It is also useful if the browser title clearly indicates what site the page is from. This can be done by using a logo if it clearly describes the contents of the page.

Interest in *accessibility* has increased considerably in recent years. In fact, it has become a key concept for any website and any lexicographical or terminographical resource since it allows users to access quickly and easily the data they are looking for. As such, when additional software is required in a website, "[…] it should be easily accessible (there should be a link to a relevant site for downloading the software), and instructions should be available for downloading and using the software" (Cooke 1999: 67). Cost is obviously a consideration in the use of electronic sources. Some services charge for accessing an electronic version of paper-based materials, and assessors might need to consider the relative value of paying to access the electronic version. Restrictions on access, such as registration, passwords, or proof of eligibility or membership of an organization may also condition the speed of accessing a website. Another major consideration is whether the site is stable or frequently changes. If the site has moved to another address, forwarding information should be provided.

The criterion *presentation and management* refers to whether the source is logically and clearly presented and arranged, and also to the quality and usefulness of the graphics or moving images in the text. In addition, should the website have advertisements, it is necessary to verify whether they are appropriately used or if they distract the user from the information or the main purpose of the page.

Summary

On the basis of the protocol described in the previous section, Table 64 shows our template for the evaluation of Internet texts that were selected for our corpus. This template is not a questionnaire that evaluates the quality of websites with a score[144]. Rather its aim is to heighten user awareness of the potential reliability or lack of reliability of the contents of a website before starting to download any text from the Web.

As shown in Table 64, every feature within each parameter is accompanied by either a check mark "√" or a question mark "?". The check mark means that the feature is a possible indicator of quality, whereas the question mark is a warning sign. In other words, it does not necessarily point out bad quality, but rather questions the possible reliability of the site.

PARAMETERS OF EVALUATION FOR WEBSITES	
AUTHORITY	
Author	√
Reputation and expertise	√
URL (TLD) – Personal: *~, %, users, .members, people, .name.* – General, including country domains: *.com, .mil, .net, .edu, .gov, .org, .int.*	 ? √
Way to contact the organization, company or person responsible for the site – email, address, telephone –	√
Feedback for the author – suggestions, complaints, requests –	√
Copyright	√
CONTENT	
Coverage:	
Valid and considerable information regarding the topic	√
Accuracy:	
Free of grammatical, spelling and typographical errors	√
References to other sources of information	√
Objectivity:	
Evident point of view of the author or organization responsible for the site	?

[144] See Seghiri (2006: 85-96) for an evaluation template with a scoring system.

If advertising, influence of the advertiser on the informational contents	?
Nonprofit or corporate sponsors clearly listed	√
Existence of links to the sites of a nonprofit or corporate sponsor in order to learn more about them	√
Currency	
Date when the material was first created _____	√
Date of the last update _____	√
Audience	
Specialized communication Semi-specialized communication Non expert/ lay communication	√
DESIGN	
Navigational aids	
Site map or index on the home page	√
Hypertext links: - Link to the home page - For sites arranged hierarchically, link to the page one level up in the hierarchy - Shortcuts available	√
Internal search engine	√
Help section	√
Browser title: - Clearly indicates source site - Clearly describes the contents of the page - Is short and unique for the site	√
Accessibility	
Fast and easy to access	√
In case additional software is required, easily accessible for downloading	√
Links work well	√
Restrictions on access (registration, passwords, etc.)	?
Cost of accessing	?
Site stability If it moves, forwarding information provided	√
Presentation and management	
Logically and clearly presented and arranged	√

Graphics or moving images add value to the text	√
Advertisement: – Appropriately used – Distract the user	 √ ?

Table 64. Template for the evaluation of online resources

4.2.2.2 The Web as Corpus approach

The *Web as Corpus (WaC)* refers to a new approach that as its name implies, regards the Web directly as a corpus. According to the classification of Bernardini, Baroni and Evert (2006), there are three ways of approaching the WaC from a linguistic perspective: (i) the Web as a corpus surrogate; (ii) the Web as a Corpus supermarket; (iii) the mega-corpus or mini Web.

As previously mentioned, we used both the WaC as well as the WfC approach. As pointed out by Bosque (2004), the compiler of REDES (§2.2.3), even in a huge corpus of millions of words, it is not possible to retrieve all the possible combinations and representative examples for a given lexical unit. Therefore, the context examples provided for the verb templates proposed in this thesis stem both from our offline corpus and directly from the web, when certain word combinations were not encountered or were not very illustrative in our offline corpus.

The Web as a Corpus surrogate

This first approach to the Web as Corpus regards the web itself as a huge corpus. Systems that implement this approach generally have an interface in which the search words are entered. Results are then displayed as concordances, in the same way as if a corpus had been entered in a corpus analysis tool on the user's computer (e.g. Wordsmith Tools, see section 4.3.2), but with the difference that the corpus is online.

These systems are rather different from conventional search engines, such as Google, in that they pre-process the questions before sending them to the search engines and then post-process the results and present them in such a way as to facilitate linguistic analysis. Some of the most widely known are *WebCorp*[145] (Kehoe and Renouf 2002), *KWiCFinder*[146] (Fletcher 2001), *Linguistic's Search Engine* (Elkiss and Resnik

[145] <http://www.webcorp.org.uk/> [05/07/2011].
[146] <http://www.kwicfinder.com/KWiCFinder.html> [05/07/2011].

2004), *WebCorpus*[147] (Fletcher 2007), *Corpeus*[148] (Leturia et al. 2007), or *Exemplar*[149]. Exemplar has been particularly useful in our research in that it searches over 1,900 journals and about 4,000 books from the Springer collection. However, these systems of pre-post processing have certain limitations, which coincide to a great extent with the limitations of search engines.

Firstly, the quantity of web text searched is limited by time constraints, and thus the recall can be poor. Since search engines offer a limited number of results for a particular query, these systems cannot retrieve more results than the search engines because they depend on them. As a result, WaC systems normally offer fewer results since they have to filter the results that do not satisfy the user's search query. Additionally, if information is unavailable on the search engine, it is almost impossible for these web corpus systems to provide it.

Secondly, the percentage of potentially relevant web texts is limited by the search criteria of search engines. Systems such as WebCorp do not have any control over Google rankings. When making a query, the system should ideally offer a random sample of reliable webpages. However, search engines return a list of pages according to specific criteria, such as popularity or geographical proximity, something that is less interesting for linguists. Therefore, when the same query is entered in the same search engine, the results will be different, depending on, for example, whether the query is made in the United Kingdom or the United States (Hundt, Nesselhauf and Biewer 2007: 2-3). Regarding popularity, Fletcher (2012) states that search engine hits are very different from corpus frequencies, and that "most widespread does not necessarily mean 'preferred' in linguistic terms".

Thirdly, search engines are inherently fragile. The information on Internet updates so rapidly that experiments can never be replicated. Fletcher (2007: 37) talks about the volatility of the web, and states that "not only do hit counts vary widely due to non-linguistic factors, but the same query on the same search site can return different sets of SERPs[150], not only from different places at different times, but even during a single user session". Ntoulas, Cho and Olston (2004) also studied the dynamicity and volatility of the web, based on the analysis of 154 webpages. The results of their analysis concluded that new webpages appear at a rate of 8% a week. However, *new* does not necessarily

[147] <http://webascorpus.org/searchwac.html> [05/07/2011].
[148] <http://www.corpeus.org> [05/07/2011].
[149] < http://www.springerexemplar.com/> [05/07/2011].
[150] SERPs stands for *search engine report pages*.

mean *additional* or *novel*. The study concluded that the total number and size of webpages remained relatively constant since *old* pages disappear, though only 5% of *new* pages have new content.

The Web as a corpus supermarket

The web can also be perceived as a *supermarket* where a corpus can be selected and acquired. Internet users go to the web to search for texts on a search engine. This approach, which is often adopted by translators, has much in common with traditional corpus compilation methods (Web for Corpus). However, in the case of the Web as a corpus supermarket, the selecting and downloading of texts can be automatized.

There are valuable tools for translation that permit users to quickly and automatically compile corpora from the web. For example, the BootCat (*bootstrapping corpora and terms from the web*) toolkit provides translators with knowledge of the terminology of a given specialized domain (Marco Baroni and Bernardini 2004). WebBootCat (Baroni et al. 2006) is a version of the BootCat tools. It is a web service to aid translators by quickly producing corpora for specialist areas in any language from the web. The application does not have to be downloaded, but can be easily accessed with the corpus analysis tool Sketch Engine.

Sketch Engine[151] (SkE) is a corpus query system incorporating word sketches, grammatical relations, and a distributional thesaurus. Word sketches are one-page automatic, corpus-based summaries of a word's grammatical and collocational behavior (Kilgarriff et al. 2004). These sketches are one of the most valuable aspects provided by Sketch Engine. As can be seen in Figure 69, a Sketch Engine account offers the user:

- Pre-loaded corpora (60 million - 2 billion words) in a wide range of languages (i.e. English, French, German, Japanese, Russian, Italian and Spanish, and for other languages such as Arabic, Chinese, Dutch, Croatian, Greek, Hebrew, Hindi, Persian, Polish, Portuguese, Romanian, Serbian, Slovenian, Swedish and Vietnamese).
- Access to WebBootCaT. As previously mentioned, this program allows users to compile a corpus of thousands of tokens in a few minutes from the 'seed terms' entered. Additionally, it permits users to do the following: (i) download the corpus to their computer; (ii) add new documents to their corpus from the web

[151] < http://www.sketchengine.co.uk/> [12/01/2011].

or from the hard disk; (iii) extract keywords of the domain; (iv) view the different texts in plain format or vertical format (i.e., annotated morphologically and by lemmas); (v) open the corpus with a lexical analysis program provided by Sketch Engine in order to work with it and do things like generate concordances, wordlist, frequency lists, collocations, and word sketches.

– A CorpusBuilder, which permits users to upload and set up their own corpora from the hard drive, and work with them from a linguistic perspective.

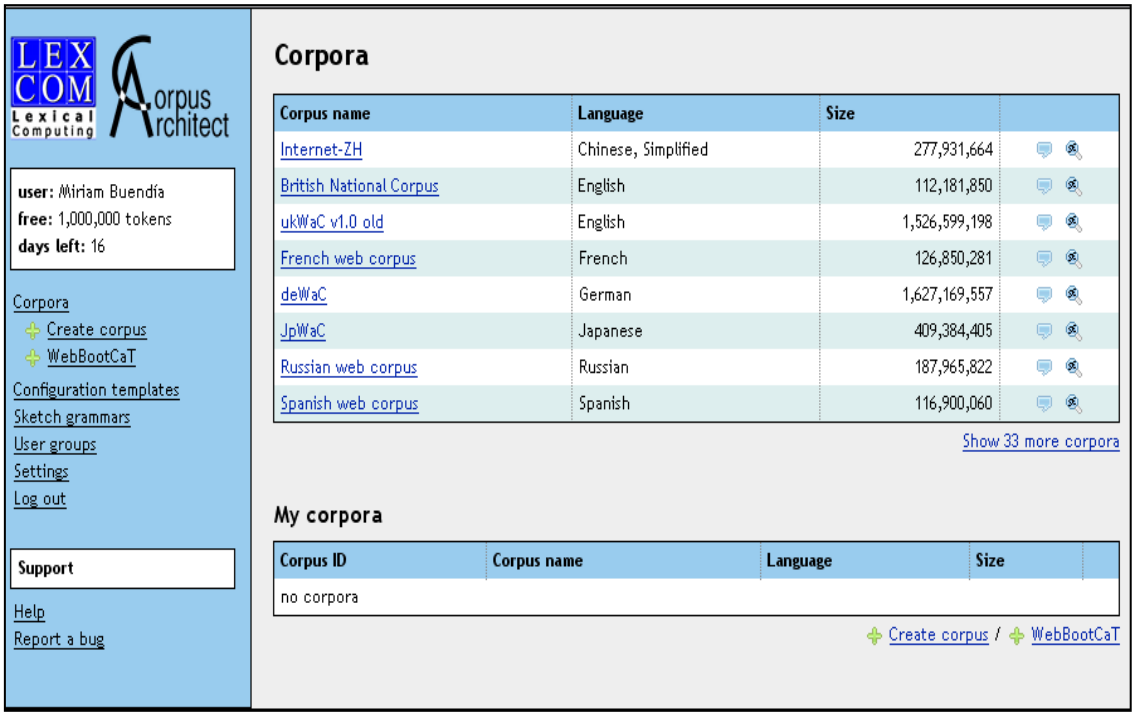


Figure 67. Interface of Sketch Engine showing its main facilities

As previously mentioned, word sketches constitute the principal novelty of Sketch Engine. Figure 70 displays an example of the word sketches obtained for the term 'earthquake' in our English corpus:

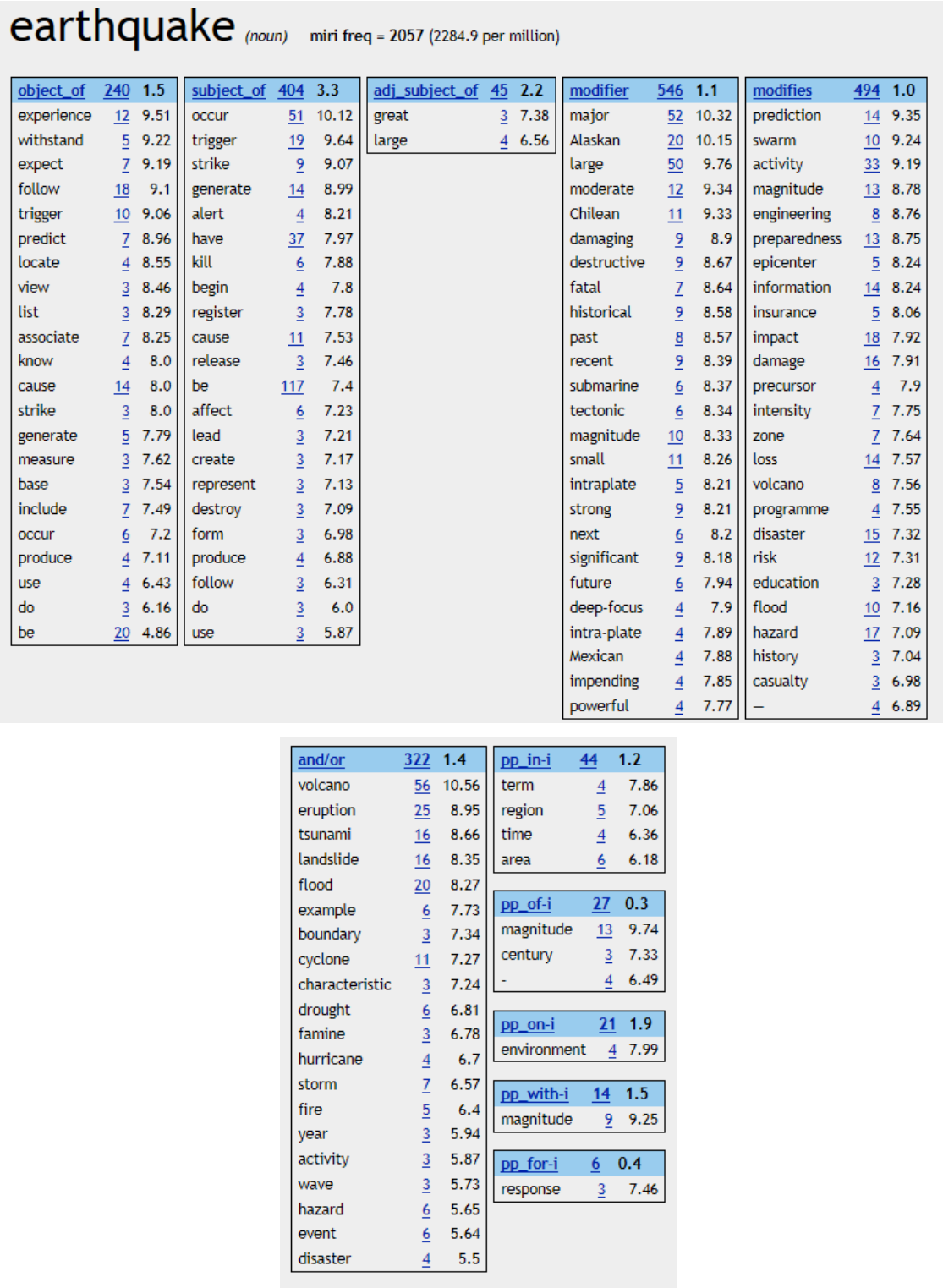


Figure 68. Word sketches obtained for 'earthquake'

As shown in Figure 70, each word sketch results page is headed by the frequency of the term in the corpus (i.e. 'earthquake' appears a total of 2057 times in our corpus in English). Subsequently the various terms that collocate with 'earthquake' are listed and classified into the following categories:

- *object_of,* which lists the verbs with which *earthquake* fulfils the role of object (e.g. ***experience*** *an earthquake*);
- *subject_of,* which displays all the verbs with which *earthquake* acts as a subject (e.g. *earthquakes* ***occur***);
- *adj_subject of,* which lists adjectives that post-modify *earthquake* (e.g. *earthquake* ***greater*** *than*);
- *modifier,* for adjectives that pre-modify *earthquake* , (e.g. ***major*** *earthquake*);
- *modifies,* for nouns modified by *earthquake* (e.g. *earthquake* ***prediction***);
- *and/or,* displays terms that behave semantically and syntactically in the same way as *earthquake* (e.g. ***volcano***);
- *pp-in-i,* for prepositional phrases with *in* that post-modify *earthquake* (e.g. *earthquakes* ***in*** *terms of loss*);
- *pp-of-i,* for prepositional phrases with *of* that post-modify *earthquake* (e.g. *earthquakes* ***of*** *magnitude 9.0*);
- *pp-on-i,* for prepositional phrases with *on* that post-modify *earthquake* (e.g. *earthquakes* ***on*** *natural environment*);
- *pp-with-i,* for prepositional phrases with *with* that post-modify *earthquake* (e.g. *earthquakes* ***with*** *magnitude*)
- pp-for-i, for prepositional phrases with *for* that post-modify (e.g. *earthquakes* ***for*** *response*).

As in the WfC approach, the advantages of the Web as corpus supermarket is that results can be replicable and that users can control corpus contents and search methods. A further advantage is that it supports linguistically sophisticated searches (Bernardini 2008). However, in the same way as the Web as a corpus surrogate, its shortcomings stem from the fact that it continues to rely on search engines and the variety of corpus contents and the noise encountered.

In this regard, Castagnoli (2006) conducted a study to analyze the advantages and limitations of the use of BootCat in a course in Terminology and LSP and concluded that that the benefit of using automatically assembled corpora is in direct relation to the user's familiarity with the specialized domain, and his/her ability to critically evaluate texts/terms retrieved (Castagnoli 2006: 171).

Given the limitations of both approaches to the Web as corpus, it became necessary to develop a search engine for linguists (Lüdeling, Evert and Baroni 2007). This is the approach followed by those who regard the Web as a *mega corpus*.

The mega-corpus or mini Web

In order to overcome the disadvantages of other approaches, a group of linguists created a new object, namely, a mini-web or mega corpus adapted to linguistic research. This new search engine for linguists targets users that wish to study aspects of language through the Web, and also those users who wish to investigate aspects of the Web through language (Bernardini, Baroni and Evert 2006: 14).

The method involves compiling a corpus directly from the web without having to trust a search engine to automatically download documents. If it were possible to access the corpus obtained from the web by means of an interface offering sophisticated search options (linguistic annotation, metadata, *inter alia*), this would be a real *search engine for linguists* (Volk 2002; Kilgarriff 2003; Fletcher 2004, 2007). Various research groups are currently working on the full implementation of this type of system (e.g. Webcorp project, GlossaNet[152] project, and the Wacky[153] project from the University of Bolonia-Forli).

A short description of the Wacky project is offered as an example of what this *mega corpus* or *mini Web* approach involves. The main objectives of Wacky (*Web as corpus kool ynitiative*) are to compile huge corpora (more than two billion words) extracted from the web in several languages, and to provide tools to process and exploit them. The corpora compiled by the Wacky team are not gathered at random, but rather follow a selection process that assures the representativeness and reliability of the texts. As described by Bernardini (2008), this process involves the following steps: (i) submission of random word combinations to Google and subsequent generation of a list of URLs (seeding); (ii) crawling performed using the Heritrix[154] crawler (which stops after ten days of continuous running); (iii) cleaning based on contents, i.e. code removal (HTML and javascript) and boilerplate stripping[155]; (iv) language filtering; (v) near-duplicate detection, i.e. "documents with substantial overlapping portions" (Baroni et al.

[152] <http://glossa.fltr.ucl.ac.be> [05/07/2011].
[153] <http://wacky.sslmit.unibo.it> [05/07/2011].
[154] <http://crawler.archive.org/> [05/07/2011].
[155] Following Fletcher (2004), the Wacky team uses the notion of "boilerplate" to refer to all those parts of web documents which tend to be the same in many webpages (for instance disclaimers, headers, footers, navigation bars, etc.) (Baroni et al. 2009: 215).

2009: 216); (vi) tokenization, POS-tagging and lemmatization; (vii) indexing and querying.

Thanks to the Wacky initiative, there are currently certain corpora already available: *deWaC* (for German), *itWaC* (for Italian), *ukWaC* (for English), and *frWaC* (for French), which can be easily accessed with Sketch Engine. They are currently working on a Spanish corpus.

For example, in the compilation of the *ukWac* corpus, the seeding involved a set of 1,000 pairs constructed by randomly combining mid-frequency content words from the BNC, from the oral section of the BNC, and from a vocabulary list for foreign learners of English. Although a maximum of ten seed URLs were retrieved for each random seed pair query, only one (randomly selected) URL was kept. Then, the filtered URLs were sent to a crawler, which limited its search to the .uk TLD. Subsequently, the processing took place, where only files between 5 and 200kb were stored; perfect duplicates were discarded; and code, boilerplate, files with unconnected text, near-duplicates, and pornographic pages were removed. The process generated an English corpus of about 2 billion words. Table 65 summarizes the size data within the process and the characteristics of the corpus compiled (adapted from Baroni et al. 2009: 218):

ukWac
▪ 2,000 seed word pairs ▪ 6,528 seed URLs ▪ 351 GB raw crawl size ▪ 19 GB after document filtering ▪ 5.69 M of documents after filtering ▪ 12 GB after near-duplicate cleaning ▪ 2.69 M of documents after near-duplicate cleaning ▪ 30 GB size with annotation ▪ 1,914,150,197 tokens ▪ 3,798,106 types

Table 65. Size data for ukWac

4.2.3 Characteristics of the corpus

As previously mentioned, our corpus is composed of two comparable subcorpora, one in English and one in Spanish, specialized in the subdomain of NATURAL HAZARDS or

the EXTREME EVENT. The English subcorpus includes 45 texts (five of which are books), which comes to a total of 779,995 tokens and 26,285 types, whereas the Spanish subcorpus contains 44 texts, which represents 449,416 tokens and 31,230 types. Although the initial objective was to compile texts with a similar number of words for each subcorpus with a view to obtaining a more balanced bilingual corpus, in the end, the final size of the corpus was conditioned by the availability of texts, based on the selection criteria. This was particularly true of the Spanish subcorpus since most of the textbooks in the domain were written in English. Consequently, it was fairly easy to retrieve textbooks on natural hazards in English in machine-readable format, but it was considerably more difficult to find textbooks in Spanish in this same format. Therefore, the textbooks in Spanish had to be scanned and converted to txt format with an OCR. They also had to be manually revised. For this reason, it was not possible to have two subcorpora of the same size, and this also explains why the Spanish corpus is smaller than the English one.

Table 64 shows the general characteristics of the two subcorpora compiled. The type/token ratio (TTR) and the standardized type/token ration (STTR), obtained with WordSmith Tools (§4.3.2), express the relation between the total number of tokens and the total number of different types found in a text. As such, they provide valuable information concerning the lexical diversity of the texts. Of the two numbers, the most interesting and reliable for comparing texts of different size is the STTR. Since the TTR is sensitive to the size of the textual sample, its value is relative. In other words, the probability of retrieving new words diminishes as the size of the texts increases. Therefore, the number of tokens upon which the TTR is calculated is usually standardized. The quotient of the tokens is calculated at regular intervals of 1,000 words with the objective of showing the mean value of the results. As stated in the help section of WordSmith[156]:

> The standardised type/token ratio (STTR) is computed every n words as Wordlist goes through each text file. By default, n = 1,000. In other words the ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of your text or corpus. A running average is computed, which means that you get an average type/token ratio based on consecutive 1,000-word chunks of text. (Texts with less than 1,000 words (or whatever n is set to) will get a standardised type/token ratio of 0.).

[156] <http://www.lexical.net/downloads/version5/HTML/index. html> [12/01/2012].

As such, in the subcorpus in English, there are on average 40.61 different words for every text sequence of 1,000 tokens, whereas for the subcorpus in Spanish the average is slightly higher (41.28). These data provide evidence that even though the English subcorpus is larger than the Spanish, the two corpora are comparable in lexical diversity (see Table 66).

	English corpus	**Spanish Corpus**
Number of tokens	779, 995	449,416
Number of types	26,285	31,230
Type/token ratio	3.71	7.24
Standardised TTR	40.61	41.28
Number of texts	45	44
Medium	written	
Topic/domain	natural hazards/extreme event	
Publication date	from 1996 to 2012	
Language	English	Spanish
Source of text	**specialized journals**	**conference proceedings**
	-Natural Hazards *-Natural Hazards Review*	*-Conference Proceedings of International Conference on Environmental Geology and Land-Use Planning* *-Riesgos naturales y antrópicos en geomorfología : libro de actas de los simposios desarrollados durante la VIII Reunión Nacional de Geomorfología (Toledo, 2004)*
	monographs	**monographs**
	-McGuire et al. (2002) -Wisner et al. (2004) -Bryant (2005) -UNISDR (2009) -EEA (2010)	-Olcina (2006a) -Olcina (2006b) -Keller and Blodget (2007) -UNISDR (2009)
	scientific magazines	**scientific magazines**
	-National Geographic *-ScienceDaily*	*-National Geographic* *-Muy Interesante*

Table 66. General characteristics of the two subcorpora compiled

Chronologically speaking, texts are relatively recent, ranging from 1996 to 2012. They can thus be regarded as a representative sample of current language on the domain. The texts were classified into the three user types identified in §1.3.2, namely, specialized, semi-specialized, and lay. Regarding specialized communication, ten articles from the specialized journal, *Natural Hazards*[157] and ten from the *Natural Hazards Review*[158] were extracted for English. The Spanish corpus included ten articles from the conference proceedings of the 6th Spanish Congress and International Conference on Environmental Geology and Land-Use Planning, entitled *Riesgos Naturales, ordenación del territorio y medio ambiente* (Chacón and Irigaray 1996), and ten from the proceedings *Riesgos naturales y antrópicos en geomorfología : libro de actas de los simposios desarrollados durante la VIII Reunión Nacional de Geomorfología, celebrada en Toledo, 22-25 de septiembre de 2004* (Benito and Díez 2004), in Spanish. For semi-specialized communication, text books on the domain of natural hazards were considered. These included the books, *Natural Hazards* (Bryant 2005), *Natural Hazards and Environmental Change* (McGuire, Mason and Kilburn 2002), *At risk* (Wisner et al. 2004), and reports of the European Union, such as *The Terminology on Disaster Risk Reduction*, compiled by The United Nations International Strategy for Disaster Reduction (UNISDR 2009) and the report, *Mapping the impacts of natural hazards*, developed by the European Environmental Agency (EEA 2010). Semi-specialized texts in the Spanish corpus were *Terminología sobre reducción del riesgo de desastres*, the Spanish version of the UNISDR 2009 report, the books *Riesgos Naturales (vol.1)* (Olcina 2006a), *Riesgos Naturales (vol.2)* (Olcina 2006b), and *Riesgos Naturales: procesos de la tierra como riesgos, desastres y catastrophes* (Keller and Blodget 2007). Popular science texts in the English corpus included ten articles from *National Geographic*[159] and ten from Science*Daily*[160]. Non-specialized texts in the Spanish corpus were articles from the popular science magazine, *Muy Interesante* and the Spanish version of *National Geographic*. Tables 67 and 68 list the text sources of each subcorpora as well as types of user groups:

[157] <http://www.springer.com/earth+sciences+and+geography/natural+hazards/journal/11069> [18/06/2012].
[158] <http://ascelibrary.org/nho/> [18/06/2012].
[159] <http://www.nationalgeographic.com/> [18/06/2012].
[160] <http://www.sciencedaily.com/> [18/06/2012].

	ENGLISH CORPUS												
	Semi-specialized texts						**Non-specialized texts**			**Specialized texts**			**All**
Texts. Sources	Total	McGuire et al. (2002)	Wisner et al. (2004)	Bryant (2005)	UNISDR (2009)	EEA (2010)	Total	Science Daily	National Geographic	Total	Natural Hazards	Natural Hazards Review	
Number of texts	5	1	1	1	1	1	20	10	10	20	10	10	
Tokens	508,185	60,171	171,521	191,553	5, 642	79,298	86, 698	49,595	37, 103	185,112	93,086	92,026	779,995
Types	20,505	5,403	10,764	11,865	1,362	6,368	7,995	6,047	4,710	11,155	8,104	6,540	26,285
Types/ token ratio	4.38	9.72	6.75	6.74	24.30	8.96	9. 66	12. 86	13. 18	7.05	9.63	8.83	3.71
Standardized TTR	42.25	40.08	44.88	42.09	42.96	38.56	41.93	42. 25	41.51	35.45	37.57	33.31	40.61

Table 67. English corpus characteristics

	SPANISH CORPUS											
	Semi-specialized texts					**Non-specialized texts**			**Specialized texts**			**All**
Texts. Sources	Total	Olcina (2006a)	Olcina (2006b)	Keller and Blodget (2007)	UNISDR (2009)	Total	Muy Interesante	National Geographic	Total	Chacón and Irigaray (1996)	Benito and Díez (2004)	
Number of texts	4	1	1	1	1	20	10	10	20	10	10	
Tokens	319,834	52,693	57,048	201,078	9,015	71,666	48,514	23,151	57,917	34,244	23,673	449,416
Types	27,720	7,066	8,225	20,851	1,614	8,175	6,742	3,562	6,453	4,312	3,664	31,230
Types/ token ratio	9.01	13.89	15.18	10.76	18.51	11.91	14.50	16.08	11.78	13.27	16.46	7.24
Standardized TTR	42.18	41.56	42.26	42.54	37.32	39.14	40.13	36.74	39.45	39.32	39.31	41.28

Table 68. Spanish corpus characteristics

4.3 Computer software

This section describes the software tools used to process and extract the data for this research: TermoStat (§4.3.1), and Wordsmith Tools (§4.3.2). More concretely, we specify which modules of each tool were used and for what purposes. However, this section only provides a short introduction to the software programs. The actual analysis is described in Chapter 5.

4.3.1 TermoStat

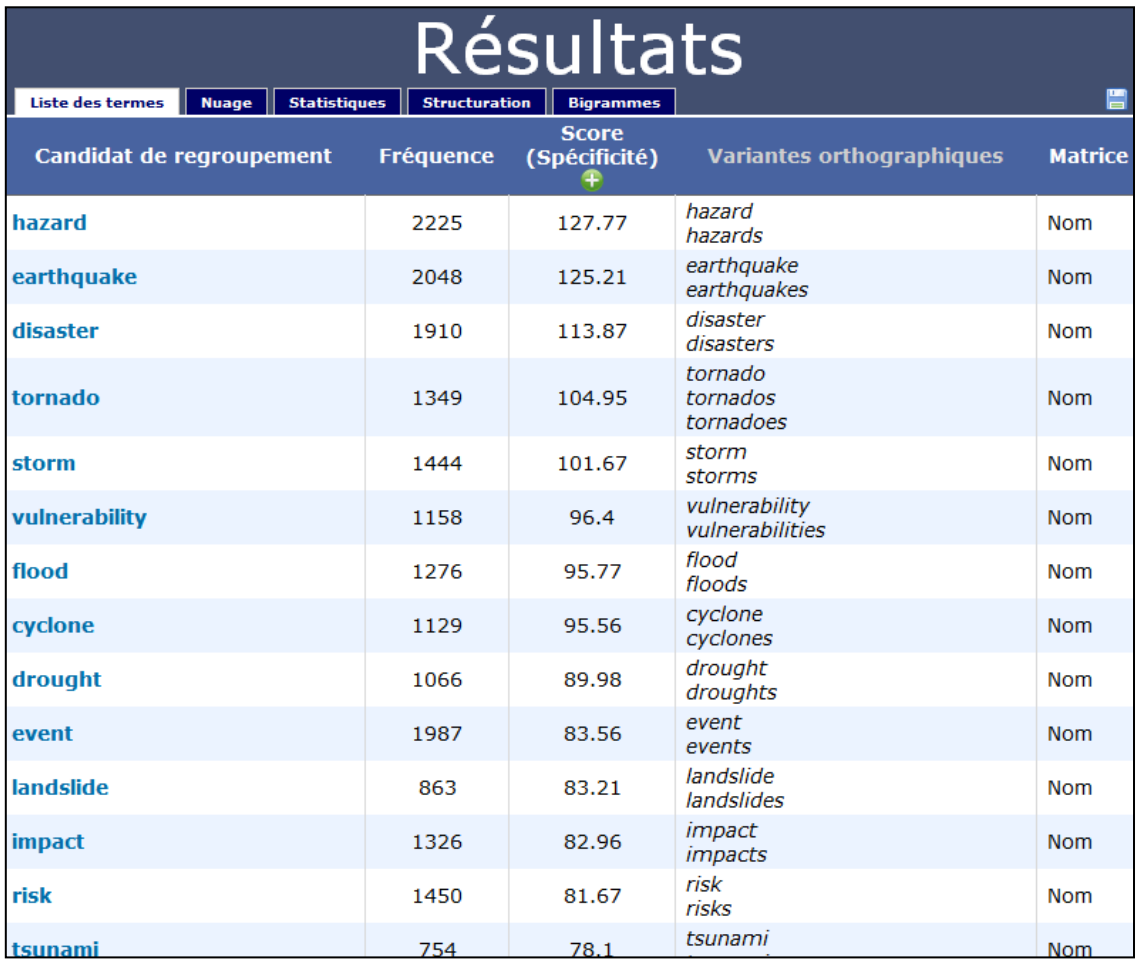
TermoStat is a term extraction tool developed by Drouin (Drouin 2003) at the Observatoire de linguistique Sens-Texte (OLST) at the University of Montréal. It is based on a term extraction technique for domain-specific corpora that leverages information from a general reference corpus. In other words, terms are statistically identified in such a way as to compare frequencies in a domain-specific corpus and in a general reference corpus:

> [The members of corpus specific vocabulary] are those lexical items whose frequency differs significantly from what is considered to be "normal" based on a comparison with previous observations (Drouin 2003: 100).

TermoStat extracts terms in English, French, Spanish and Portuguese, and thus has a reference corpus for each of these languages. The reference corpus for English is composed of about 8 million words with about 456,000 different forms. Half of the corpus is formed by texts published between March and May 1989, extracted from a *The Gazette*, a Montreal newspaper, whereas the other half of the corpus comes from the BNC. The Spanish reference corpus has 30 million words and about 527,000 different forms, corresponding to texts from the European Parliament.

The TermoStat system receives an input text corpus in .txt format and then generates a list of candidate terms (CTs) extracted from the corpus. CTs can be both monolexical and polylexical units, and the extraction can be restricted to a specific grammatical category (adjective, adverb, noun or verb) and/or to monolexical or polylexical units. Each CT is given a specificity score which is the product of the comparison of the candidate's frequency in the input text corpus and in the general reference corpus. A high specificity score indicates that the CT is particularly frequent in the input text corpus and is therefore specific to it. It is thus offered as a possible CT.

Figure 71 shows an example of the results interface of TermoStat. It corresponds to the term analysis performed in our English corpus, constraining it to monolexical noun units. As shown, results can be accessed by means of the following sections: *listes des termes* [term list], *nuage* [cloud], *statistiques* [statistics], *structuration* [structuration] and *bigrammes* [bigrams].



Résultats

Liste des termes | Nuage | Statistiques | Structuration | Bigrammes

Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
hazard	2225	127.77	*hazard hazards*	Nom
earthquake	2048	125.21	*earthquake earthquakes*	Nom
disaster	1910	113.87	*disaster disasters*	Nom
tornado	1349	104.95	*tornado tornados tornadoes*	Nom
storm	1444	101.67	*storm storms*	Nom
vulnerability	1158	96.4	*vulnerability vulnerabilities*	Nom
flood	1276	95.77	*flood floods*	Nom
cyclone	1129	95.56	*cyclone cyclones*	Nom
drought	1066	89.98	*drought droughts*	Nom
event	1987	83.56	*event events*	Nom
landslide	863	83.21	*landslide landslides*	Nom
impact	1326	82.96	*impact impacts*	Nom
risk	1450	81.67	*risk risks*	Nom
tsunami	754	78.1	*tsunami*	Nom

Figure 69. Results interface of TermoStat

As shown in Figure 71, the *term list* section displays the CTs (first column of the left) in their lemmatized form, followed by their frequency in the analyzed text; the score specificity assigned to each unit; the orthographic variants (i.e. non-lemmatized forms) as well as spelling variants; and the patterns, i.e. the grammatical categories of the CTs. For example, as can be seen, some of the most recurrent units in our English corpus are 'hazard' (with a frequency of 2225 occurrences); 'earthquake' (2048 occurrences); 'disaster' (1910 occurrences); 'tornado', 'flood', 'cyclone', 'drought',

'landslide', 'impact', 'tsunami', which are indeed very significant terms associated with the extreme event domain[161].

The *nuage* [cloud] option provides the 100 CTs with the highest specificity scores, varying in front size according to the score. The cloud displayed for the terminological extraction generated by our English corpus is shown in §5.2. The *Stats* section lists the number of CT of each grammatical category. The *Structuration* tab offers the CTs along with their frequency, and finally, the *bigrams* tab displays the verb+noun units that frequently appear in the input text. We did not make use of the bigrams option since the system only provided a few results for each analysis which is not representative and not particularly helpful. Nevertheless, in the future, it should be possible to perform the bigrams option, which would be indeed of tremendous help for our purposes.

As shall be seen in chapter 5, we have made use of TermoStat in two ways:

- As a starting point to obtain a general overview of the terminology of our corpus and help us in the design of the EXTREME EVENT. To that end, the term extraction process was restricted to monolexical nouns and verbs. This analysis along with the wordlists obtained with WordSmith Tools (§4.3.2) were used to design the conceptual structure of the EXTREME EVENT frame (§5.2);
- To select the specialized verbs activated within the EXTREME EVENT subdomain and consequently, the verbs analyzed in this research (§5.3). To this end, a second analysis was launched, constraining it only to verbal units.

Despite the fact that the results provided by TermoStat are quite accurate, it should not be forgotten that they are obtained automatically based on formal cues. Therefore, human evaluation is always necessary to determine whether the terms extracted are indeed terms, whose meaning is relevant to the specific domain of study. The evaluation process and the selection of verbs are explained in 5.3 and 5.4.

4.3.2 Wordsmith Tools

Wordsmith Tools[162] is a corpus analysis tool developed by Mike Scott at the University of Liverpool. Its first version was released in 1996, and its 5th version is currently on the

[161] The complete description of the most frequent terminological units in our corpus is found in section 5.2.
[162] <http://www.lexically.net/wordsmith> [12/12/2011].

market (Scott 2008). In fact, the beta version of WordSmith Tools 6 is already available from the website. The software includes three modules: *Wordlist*, *Concord*, and *Keywords.*

Wordlist displays the frequency of each word in the corpus. Lists can be arranged either by frequency or alphabetically. *Wordlist* also provides statistical details about the texts under study. In our analysis, the statistical option was used to describe our corpus (number of tokens, types, type/token ratio) (see Table 69, in §5.2). The frequency option was then used to continue finding the SKU in the EXTREME EVENT frame. As Bowker and Pearson (2002: 119) point out:

> The most important consideration of all is that word lists enable you to consider investigating patterns that you might not otherwise have been observed.

The list with the 50 most frequently units in the corpus can be found in §5.2 (Table 69), along with the tag cloud provided by TermoStat, which displays the 100 most frequent words in the corpus by means of assigning a different size to each one depending on its frequency (see Figure 72 in §5.2).

The *Concordance* function generates concordances, i.e. a set of examples of a given word or phrase. In other words, Wordsmith displays concordances in the form of a KWIC (key-word-in-context) concordance line, which shows all occurrences of a given term in contexts of a fixed length within the corpus. Wordsmith allows users to arrange the list according to their needs. In other words, it permits the edition, manipulation, and reorganization of the information so that meaningful patterns can be more easily identified.

In our analysis, the *Concordance* option of Wordsmith Tools was used in two ways. First, text concordances regarding the terms denoting the ten event-specific concepts identified in Chapter 1 (i.e. VOLCANO; LANDSLIDE; FLOOD; AVALANCHES; TROPICAL CYCLONE, including TYPHOON and HURRICANE; TORNADO, DROUGHT, EARTHQUAKE, FIRE, and TSUNAMI) were analyzed in order to extract and validate their conceptual relations. To that end, we followed the methodology described in 1.4.3.1, which focused both on the analysis of the definitions of each concept along with the analysis of the KWIC of the term(s) associated with each concept. This analysis along with the data extracted from the Wordlist option allowed us to complete the design of the EXTREME EVENT (Figure 73 in §5.2).

The *Concordance* option was also used in a subsequent phase (§5.4) to analyze the behavior of the selected verbs for this study in texts. This analysis was the main focus of study.

Finally, the *Keywords* option in Wordsmith Tools was used to identify the key words in one or more texts, i.e. those words whose frequency is unusually high in comparison with some norm. As such, it could have been used to identify the most representative verbs within our corpus. However, we preferred to use TermoStat since this application provided results of a better quality than those of *Keywords.* As such, the *Keywords* module of WordSmith was not used in our analysis.

4.4. Conclusion

The fact that the Earth is suffering from an increasing number of disasters, both natural and human-induced, has made the environment an interesting area of study for a wide range of people all over the world. However, as previously mentioned, there have been few terminological studies focused on natural hazards. The growing interest in this domain and the lack of studies in this field were the main reasons why we decided to analyze it, and more specifically, the EXTREME EVENT in consonance with the conception of events in Frame-based Terminology (§1.4.3).

This chapter first specifies the boundaries of concepts within the scope of the EXTREME EVENT. The first step in the design of the EXTREME EVENT frame was an informal description of the type of situation activated by this event. In this regard, we first looked up the concept of NATURAL DISASTER in EuroVoc, the multilingual thesaurus of the European Union. Along with EuroVoc, the organization of NATURAL DISASTERS provided by the website *ScienceDaily* was also taken into account as well as the typology of hazards provided by the European Environment Agency. Consequently, it was concluded that the concepts activated by the EXTREME EVENT were the following: VOLCANO, LANDSLIDE, FLOOD, AVALANCHE, TROPICAL CYCLONE (including TYPHOON and HURRICANES), TORNADO, DROUGHT, EARTHQUAKE, FIRE, and TSUNAMI.

Subsequently, this chapter explained the design, compilation and description of the corpus used to analyze the data. To that end, a brief revision of the concept of corpus was provided (§4.2.1.1) as well as a succinct description of the typology of corpora (§4.2.1.2).

In addition, we described the methods and criteria for corpus compilation (§4.2.2) since corpus compilation determines, to a great extent, the results of our research analysis. In this regard, we designed a protocol for the evaluation of online resources which assures the reliability and quality of the texts in our corpus (§4.2.2.1.2), and which can be applied to any type of corpus. The two methods of corpus analysis were also described: (i) *web for corpus (WfC)*, in which the web is used as a source of texts in digital format for the subsequent implementation of an offline corpus; (ii) *web as corpus (WaC)*, which uses the web directly as a corpus. As previously mentioned, our research analysis integrates both methodologies. In other words, firstly, two subcorpora of texts (one in English and one in Spanish) in the subdomain of EXTREME EVENT were compiled (i.e. a traditional *Web for Corpus* approach. However, certain examples, contexts of use, and phrases were directly retrieved from the web (i.e. following a *Web as Corpus* approach). This was the case when the information derived from our compiled corpus was not sufficient to satisfy our needs, when no examples were found, or when we wished to offer a wider range of contexts.

Finally, section 4.3 describes the software application used, namely, the term extractor, TermoStat, and the corpus analysis tool, WordSmithTools. TermoStat was mainly used to select the specialized verbs activated within the EXTREME EVENT and, thus, the verbs to be analyzed in this thesis. Regarding WordSmithTools, the *Concordance* option was used to analyze the behavior of the verbs selected for this study.

5. Results and discussion

5.1. Introduction

In consonance with Frame-based Terminology (§1.4.3), the research in this thesis was carried out by using an integrated bottom-up and top-down approach. Accordingly, the conceptualization of the extreme event is based on the factorization of dictionary definitions for the most salient concepts (top-down approach) as well as the extraction and analysis of verbs and argument patterns in corpus texts (bottom-up approach). Since the analysis was corpus-based, candidate verbs were first extracted, after which a final inventory was selected, analyzed, and organized in frames (bottom-up approach). However during the process, the corpus analysis was complemented with definitional information extracted from specialized dictionaries and other reference material (top-down approach). Our approach to frames differs from that used in FrameNet because, as previously mentioned (§3.1.3.2), FrameNet uses a top-down approach to describe frames, i.e. the frame is first identified and described, and then the set of lexical units within the frame are specified.

5.2 Description of the EXTREME EVENT: conceptual organization

Within the field of the environment, this analysis focuses on the subfield of NATURAL HAZARDS, as instantiated by the EXTREME EVENT, according to the conception of events in Frame-based Terminology. As previously described in 4.1, the EXTREME EVENT includes the following basic concepts: VOLCANO, LANDSLIDE, FLOOD, AVALANCHE, TROPICAL CYCLONE (including TYPHOON and HURRICANE), TORNADO, DROUGHT, EARTHQUAKE, FIRE, and TSUNAMI. In this section, a complete description of the conceptual organization underlying the EXTREME SUBEVENT is provided following the methodology described in 1.4.3.1 for the configuration of concepts in EcoLexicon.

As shown in 1.4.3.1, the methodology used to design EcoLexicon is based on specifying the conceptual system domain by means of an integrated top-down and bottom-up approach. As such, as illustrated by the concept of TROPICAL CYCLONE for each of the ten main concepts associated with the EXTREME EVENT, various dictionaries were used. The definitions were then compared so as to craft a well-structured definition that would encode the relations and attributes of the concepts in question. In a parallel

way, corpus analysis was also used to extract conceptual information regarding semantic relations. To that end, a set of 20 lines of concordances were analyzed for each of the ten aforementioned concepts based on the steps described in 1.4.3.1.

Subsequently, a word list for English and another one for Spanish were obtained from the corpus, thanks to the WordList option of WordSmith Tools (Table 69). The word lists obtained for each language were very useful to find more lexical units associated with the EXTREME EVENT.

Function words, such as prepositions, articles, conjunctions, etc. were not considered. These words were eliminated by means of a stoplist of words that we did not wish to include in our analysis.[163] Table 69 displays the wordlist obtained from our English subcorpus. It displays the 50 most frequent units. The number on the right of each lexical unit expresses the number of occurrences of the word in the subcorpus (e.g. 'earthquake' appears a total of 2210 times):

The list shows that some of the most frequent lexical units are indeed the terms which designate the event-specific concepts, such as 'earthquake', 'drought', 'flood', 'tsunami', 'tornado', 'cyclone', 'fire'. Other frequent words are 'damage', 'disaster', 'change', 'loss', 'risk', 'impact', which are some of the results/consequences of NATURAL HAZARDS. Also on the list are adjectives, such as 'natural', 'tropical', 'high', 'large', and 'major', which point to some of the attributes of these phenomena. The presence of 'climate', 'pressure', 'wind', 'water' reflect that NATURAL HAZARDS are tightly linked to these weather conditions;. The frequent occurrence of 'scale' indicates how their intensity is measured. The presence of 'time' and 'years' are evidence of the cyclic nature of these phenomena and the fact that they have a duration. In addition, other words, such as 'surface', 'place', 'sea', 'area', 'land', point to the close link between NATURAL HAZARDS and the place where they generally happen. This is also confirmed by the high percentage of lexical units denoting cardinal points, such as 'north', which show that these hazards take place in specific regions. Other words, such as 'people' or 'population', reflect that NATURAL HAZARDS negatively affect people and are threats to human life and property.

[163] The stoplist used for the subcorpus in English was obtained from <http://www.unine.ch/info/clef/> and the stop list for the Spanish subcorpus was obtained from the website of *Corpus del Español*, <http://www.corpusdelespanol.org/>.

1	HAZARDS	2717	**26**	ECONOMIC	807
2	EARTHQUAKE	2210	**27**	TSUNAMI	796
3	EVENT	2034	**21**	ERUPTION	754
4	PEOPLE	1843	**29**	AVALANCHE	735
5	WATER	1682	**30**	FIRE	727
6	AREA	1660	**31**	FATALITY	725
7	STORM	1658	**32**	MAJOR	724
8	RISK	1625	**33**	FAMINE	719
9	FLOOD	1582	**34**	RAINFALL	714
10	TORNADO	1469	**35**	SLOPE	702
11	CYCLONE	1266	**36**	VOLCANO	701
12	DISASTER	1261	**37**	MAGNITUDE	700
13	VULNERABILITY	1194	**38**	NUMBER	698
14	NATURAL	1142	**39**	SCALE	696
15	TROPICAL	1029	**40**	WIND	673
16	DAMAGE	1028	**41**	CLIMATE	661
17	TIME	1018	**42**	PRESSURE	659
18	DROUGHT	939	**43**	LOSS	655
19	LARGE	900	**44**	FLOW	650
20	YEARS	900	**45**	NORTH	646
21	SEA	893	**46**	LAND	636
22	LANDSLIDE	835	**47**	POPULATION	636
23	CHANGE	831	**48**	SURFACE	625
24	IMPACT	810	**49**	SOCIAL	617
25	HURRICANE	810	**50**	TEMPERATURE	615

Table 69. Fifty most frequent words in the English corpus obtained with *Wordlist*

The tag cloud offered by TermoStat (see Figure 72) was also useful in the design of the conceptual structure of the EXTREME EVENT since it provided a general overview of the event by displaying the 100 terms with the highest scores of the corpus. As shown in Figure 72, some of the most frequent terms in our English subcorpus are ‘avalanche’, ‘cyclone’, ‘damage’, ‘disaster’, ‘drought’, ‘earthquake’, ‘eruption’, ‘evacuation’, ‘event’, ‘famine’, ‘fatality’, ‘fire’, ‘flood’, ‘flooding’, ‘flow’, ‘hazard’, ‘hurricane’, ‘impact’, ‘landslide’, ‘loss’, ‘rainfall’, ‘risk’, ‘storm’, ‘tornado’, ‘tsunami’, ‘volcano’, ‘vulnerability’, ‘water’, ‘wave’, and ‘wind’:

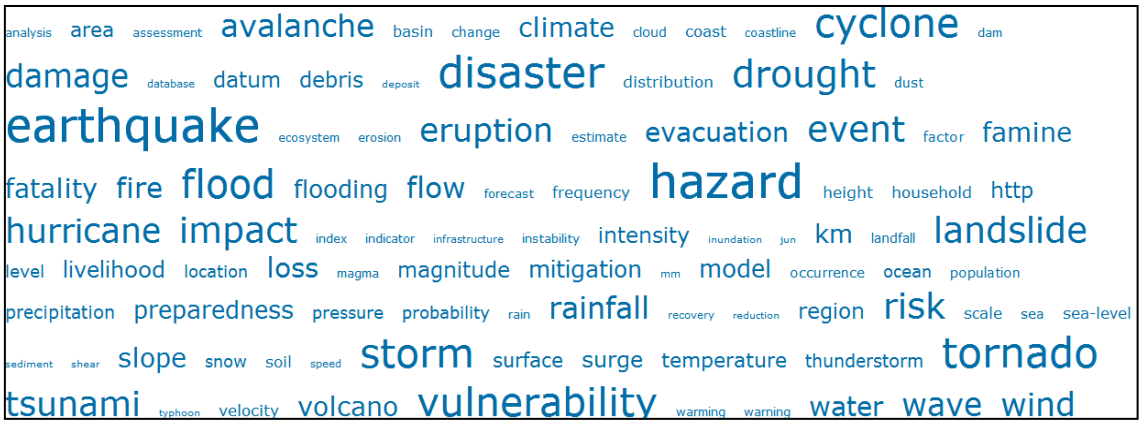


Figure 70. TermoStat tag cloud of the 100 most frequent specific terms of the English subcorpus

This preliminary analysis was based on the Wordlist option of WordSmith Tools and the tag cloud option of TermoStat, along with the definitions and concordances obtained for the ten main concepts of the EXTREME EVENT. The methodology used in Frame-based Terminology allowed us to specify the conceptual structure of the EXTREME EVENT, as displayed in Figure 73:

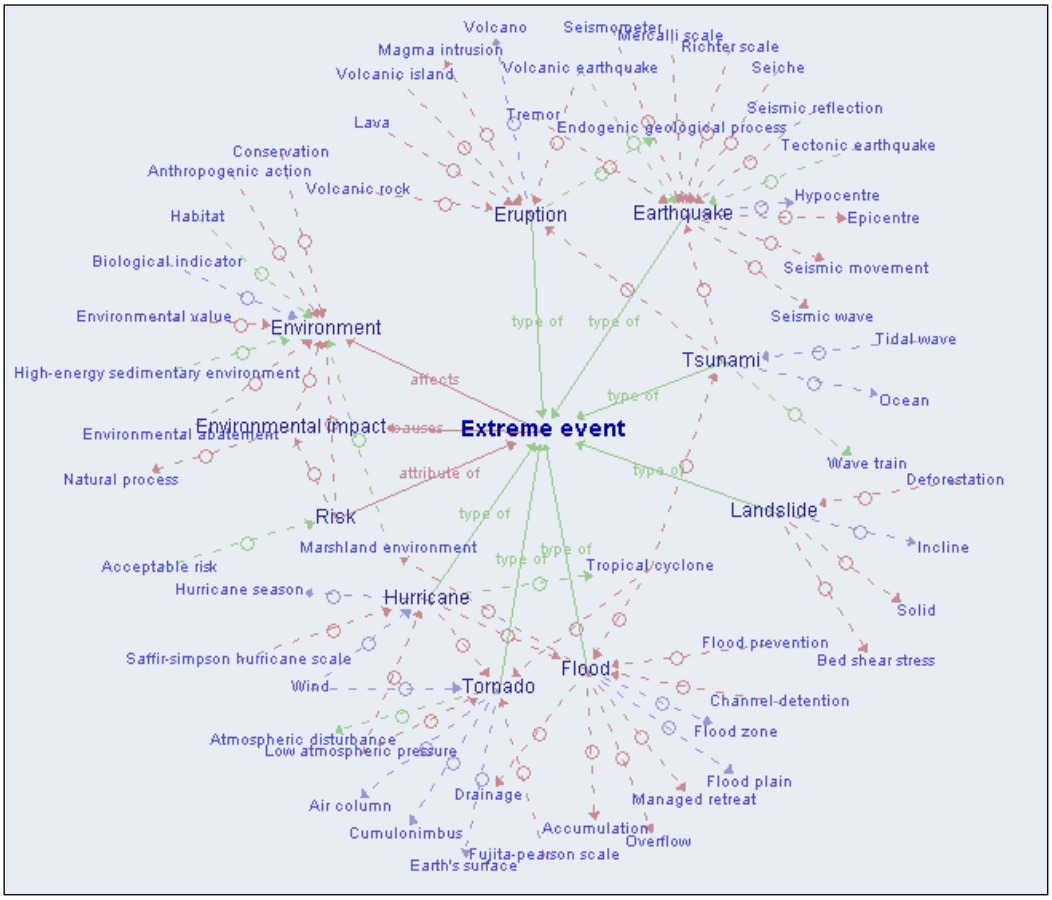


Figure 71. Conceptual structure of the EXTREME EVENT

5.3 Extraction of the candidate verbs

The extraction of candidate verbs was performed with TermoStat (§4.3.1). For that purpose, a second analysis was performed with this software application[164] in which the search was restricted to verbs. The first search was done for the English subcorpus and the second for the Spanish subcorpus. Figure 74 displays a screenshot of the English verbs that were extracted:

Résultats

Liste des termes | Nuage | Statistiques | Structuration | Bigrammes

Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
occur	900	50.8	occur occurs occurred occurring	Verbe
cause	850	40.71	cause causes causing	Verbe
trigger	256	38.07	trigger triggers triggered triggering	Verbe
flood	225	35.34	flood floods flooded flooding	Verbe
affect	544	33.19	affect affects affected affecting	Verbe
increase	810	32.74	increase increases increasing	Verbe
associate	400	30.18	associate associated associating	Verbe
evacuate	155	29.22	evacuate evacuates evacuating	Verbe

Figure 72. Extract of TermoStat displaying the most frequent verbs in the English subcorpus

As shown in Figure 74, the term list section displays the candidate terms [*candidat de regroupement*] (first column of the left) in their lemmatized form, followed by their frequency in the text; the score specificity assigned to each unit; the orthographic variants (i.e. non-lemmatized forms) as well as the spelling variants. For example, 'occur', 'cause', 'trigger', 'flood', 'affect', 'increase', 'associate', 'evacuate' were among the most frequent verbs in the English corpus.

The analysis of the English corpus produced a total of 352 verbs and that of the Spanish corpus, 323 verbs. However, some of the candidate verbs generated by TermoStat had to be discarded since they actually corresponded to other parts of speech (e.g. 'crater', 'cluster', 'firefight', 'map'). The other candidate terms suggested by

[164] The objective of the first analysis performed with TermoStat was to obtain an overview of the terms in our corpus as shown in the tag cloud (Figure 72).

TermoStat were initially accepted even though some of them were not specifically linked to the domain of the environment and to the subdomain of natural hazards (e.g. 'feel', 'study', 'analyze', 'observe', 'detect'). All these verbs were then classified in lexical domains as proposed by the LGM, based on their definitions, as shall be seen in 5.4.1. This provided the most prominent lexical domains activated within the EXTREME EVENT, which would be the focus of our study.

5.4 Verb analysis

5.4.1 Top down analyis

As previously mentioned, the candidate verbs offered by TermoStat were classified in lexical domains as proposed by the LGM (§3.1.2.1). The only lexical domain that was not activated in the EXTREME EVENT was that of SENSE PERCEPTION. Each lexical domain was further subdivided into subdomains. As such, each verb given by TermoStat was placed in one of the lexical domains, or in the case of polysemous verbs, they could even belong to various domains or subdomains. Our membership criteria were established according to the most prototypical meaning of the verb, within the context of the EXTREME EVENT. However, for certain polysemous verbs, there were various senses relevant to our field of study. When this occurred, the various senses were retained and each sense was assigned a number, based on their meaning derived from corpus analysis.

For example, 'burst' has three different senses within the EXTREME EVENT frame: (i) 'burst1', meaning *to move forcefully (esp. of wind events),* within MOVEMENT; (ii) 'burst2', meaning *to begin to exist (esp. of volcanic events),* within EXISTENCE; (iii) 'burst3', meaning *to cause to come apart, (esp. of construction entities)* within ACTION.

The classification of verbs in domains and subdomains was based on the definition of the verbs. Definitions were specified by consulting different dictionaries and other reference material, according to the Lexical Grammar Model and Frame-based Terminoloy (§1.4.3.1). Since each subdomain focuses on a particular area of meaning, this means that all verbs within a subdomain can be defined either directly or indirectly in terms of the same hyperonym. As such, each verb has a *genus* which corresponds to the superordinate verb in the hierarchy. The means that the more specific the hierarchy becomes, the more specialized the information will be. Table 70 shows the

classification of all the lexical domains and subdomains along with the verbs or verb senses in each subdomain. See Appendix 1 for the definitions of all the verbs, along with their classification in domains and subdomains.

As shown in Table 70, after specifying the lexical domain, each subdomain appears in boldface type and italics, followed by their lexicalization in English and Spanish. For example, as can be seen, within the EXTREME EVENT, the verbs belonging to EXISTENCE, are further classified into the subdomains of *to begin to exist*¸ *to begin to exist from sth else*, *to begin to exist becoming sth else*, *to cause to exist/happen*, *to prevent sth from existing/happening*, *to be likely to happen*, *to exist in time*, *to continue to exist (of people)*, *to continue to exist (of natural disaster)*, *to cease to exist*, *to cease to exist in the perception of others*, *to cause sb to cease to exist*, and *to cause to cease to exist (of fire disaster)*.

TO BE [EXISTENCE]
to begin to exist
start2, break out, originate, develop, evolve, blow up, form, burst2, erupt1, explode2
empezar, despertar, comenzar, originarse, desarrollarse, formarse
to begin to exist from sth else
originate from, develop from, evolve from, form from
originarse a partir de, desarrollarse (a partir) de, formarse de
to begin to exist becoming sth else
develop into, evolve into
transformarse en, convertirse en, evolucionar a
to cause to exist/happen
cause, start1, form2, generate, produce, spawn, result (from/in), trigger, spread2
causar, generar, producir1, provocar, ocasionar, originar
to prevent sth from existing/happening
prevent, avert
impedir, evitar, prevenir
to be likely to happen (especially sth bad)
threaten1, risk, endanger
amenazar1
to exist in time

happen, take place, occur, recur
producirse, ocurrir, tener lugar, suceder, sobrevenir, pasar
to continue to exist (of natural disaster)
last, persist, extend2
durar, perdurar, persistir, extenderse2
to continue to exist (of people)
survive, recover from
sobrevivir a, recuperarse de
to cease to exist
die, perish, drown1, disappear1
morir, perecer, fallecer, ahogarse, desaparecer1
to cease to exist in the perception of others
disappear2, dissipate, blow out1
desaparecer2, disiparse
to cause sb to cease to exist
kill, drown2
matar
to cause to cease to exist (of fire disaster)
extinguish, douse
apagar, extinguir, sofocar
TO BECOME DIFFERENT [CHANGE]
to become different by going from one state to another
melt, liquefy, condense
condensar, derretir, fundir, solidificar, helar, evaporar
to become larger in size/amount/number/degree
increase, accumulate1, strengthen, build up, intensify, surge, amplify, accelerate, rise
aumentar, amplificar, acrecentar, intensificar, intensar, acentuar, activar1, reforzar, acelerar
to become less in size/amount/degree/importance/intensity
lessen, mitigate, decrease, fall, reduce, damp, absorb1, minimize, log, weaken
disminuir, mitigar, aliviar, paliar, reducir, comprimir, minimizar, angostar, descender, bajar1, decrecer, debilitar
to cause sth to become less in size

erode
erosionar
to cause sth to increase in moisture
wet
humidificar
to cause sth to lessen in moisture
dry
secar, agostar
to cause sth to increase in temperature
warm
calentar
to cause sth to decrease in temperature
cool
enfriar
to cause a change in sth so that it loses its power or control
destabilize
desestabilizar
to cause sth/sb to change for the worse
affect, damage, destroy, devastate, ravage, demolish, wreck, sweep away, burn1, injure, deepen, exacerbate
afectar, castigar, dañar, destruir, destrozar, asolar, arrasar, devastar, demoler, derrumbar, derribar, derruir, quemar, arder1, calcinar, dificultar, empeorar, agravar
to begin to be different in the way that is stated
become, change, vary, range, oscillate, modify
cambiar, alterar, modificar, transformar, variar, oscilar, fluctuar
to separate into two or more pieces
break, share, fracture, breach, ground, crack, rip, rupture
romper, desgarrar, fracturar, rajar, agrietar
to cause sth not to be changed by sth else
resist, withstand
resistir, soportar
TO HAVE/GIVE [POSSESSION]
to cause sth to have sth as a part

include, incorporate, encompass, involve, integrate, aggregate, absorb2
incluir, absorber, incorporar, integrar, agregar
to have sth at one place
trap, entrap
retener
to cause sth/sb to have sth
supply, irrigate
suministrar, abastecer
to cause sth not to have sth anymore
remove, unload, denude
eliminar, retirar
to have a large number of sth
accumulate2
acumular, almacenar, embalsar
TO SAY [SPEECH]
to say that sth will happen
predict, forecast, project, foreshadow
predecir, pronosticar
to say that sth bad will happen
warn, alert, threaten2
avisar, alertar, amenazar2
to say sth to sb
report
informar
TO FEEL/EXPERIENCE [EMOTION]
to feel/experience a particular situation
feel, experience, suffer
sentir, experimentar, sufrir, padecer
TO DO/MAKE [ACTION]
to do sth
perform, carry out, undertake
llevar a cabo, realizar, efectuar, ejecutar
to make sth

create, produce
fabricar1, elaborar, crear, producir2
to stop doing sth
interrupt, disrupt
interrumpir
to come against sth with sudden force
hit, batter, strike, blast2, crash, collide
golpear, impactar, batir, azotar, sacudir, chocar, colisionar
to (cause) to come apart (of construction artifact)
burst, collapse, explode
reventar, explotar, explosionar, estallar, desplomarse
to produce fire
burn2, blaze, smolder/smoulder, flare
arder2
to make an opening in sth with a sharp-edged tool
cut, crop
cortar, picar
to make sth (especially referred to a building, road, or machine)
construct, build, dam, rebuild
fabricar2, construir, alcantarillar
TO USE [MANIPULATION]
to put sth into use
implement
activar2, implementar
to put a special substance on sth
treat
tratar
to regulate a mechanical or scientific process
regulate, control, manage
regular, controlar, monitorizar
to mark sth in order to prepare it for later use
mark, calibrate
marcar, calibrar

TO KNOW/THINK [COGNITION/MENTAL PERCEPTION]
to think carefully about sth
consider, equate, determine, calculate, estimate, underestimate, overestimate, evaluate, assess, calibrate, measure
considerar, identificar, suponer, calcular, evaluar, estimar, averiguar, conocer, examinar, estudiar, analizar, determinar, decretar
TO MOVE (GO/COME) [MOVEMENT]
to move in a certain direction (the wind)
blow
soplar
to move forcefully
surge, sweep1, blast2, burst1
----
to move slowly
drift, slide, creep, shift
deslizar
to move circularly
spin, rotate
girar, orbitar, rotar
to (cause) to move vertically
shake
agitar, temblar, remover
to move in a certain direction
rise, surface, fall, descend, avalanche, converge, transport, penetrate, funnel
ascender, elevar, emerger, transportar, penetrar, adentrar, entrar, cruzar, atravesar, trasladar, trasvasar, caer, bajar2, descender, bajar, volcar, avanzar
to move horizontally in a certain direction over a large space
spread, extend, propagate
extenderse1, propagarse
to cause to move people from one place to another
evacuate, migrate, relocate, resettle, uproot, displace, flee, drown
evacuar, migrar
to cause motion

loft, release, eject, erupt2, blast1, emit, expel, blow out2, spew, spit
echar, expulsar, desprender, liberar, lanzar, despedir, soltar, emitir, emanar, arrojar
to move(especially for liquids)
flow, spill, spill over, overflow
fluir, bombear, derramar
to move sth without changing direction
channel
drenar, avenar, desaguar
to cause sth/sb to change direction
divert, deflect
desviar
to cause sth to stop moving
paralyze, stabilize, block
paralizar, estabilizar, contener, bloquear
TO BECOME AWARE (NOTICE/PERCEIVE) [GENERAL PERCEPTION]
to cause sth to be seen
show
mostrar
to see sth
observe, see, detect, register, record
observar, ver, detectar
to see/hear/taste/smell/touch **[SENSE PERCEPTION]**

TO BE/STAY/PUT [POSITION]
to (cause) to be in a certain place or position
surround, locate, localize, inhabit, populate
situar, localizar, habitar, poblar
to put sth in a certain place
deposit
depositar, verter
to put sth on top of or over sth
cover, engulf, bury
cubrir, sepultar, enterrar, envolver, recubrir

to put water over/in a space
flood, inundate, swamp
inundar, anegar

Table 70. Most frequent verbs in English and Spanish within the EXTREME EVENT classified in terms of LGM domains and subdomains

As is well known, the concepts activated by the EXTREME EVENT (e.g. HURRICANE, TSUNAMI, FIRE) are generally processes which begin to exist, exist over a period of time, and finally cease to exist. During their EXISTENCE, they ACT at a certain POSITION, causing negative CHANGE (i.e. damage) on an affected entity. It is thus not surprising that the most frequent verbs in the EXTREME EVENT frame generally belong to the domains of EXISTENCE, ACTION, MOVEMENT, POSITION, and CHANGE. In fact, verbs of ACTION, MOVEMENT and CHANGE are especially relevant.

Apart from the meaning conveyed by each lexical domain, it is worth mentioning that verbs within each domain also provide pragmatic information (León 2009: 278). In this regard, as Lorente (2000) states, the predominance of a certain group of verbs in a text will determine, to a great extent, the nature of the text and the way contents are treated. León (2009: 278) even goes so far as to say that a close relation between the various lexical domains of the LGM and the verb classification proposed by Lorente (2002, 2007) can be established.

Lorente (2002) proposes a typology of verbs for specialized discourse, which is redefined in Lorente (2007). This classification is composed of four types of verbs: (i) *discursive verbs* [verbos discursivos]; (ii) *connective verbs* [verbos conectores]; (iii) *phraseological verbs* [verbos fraseológicos], and (iv) *terminological verbs* [verbos terminológicos] or [verbos término].

Discursive verbs are tightly linked to the function(s) of the text in which they co-occur (e.g. 'describe', 'tell') ['describir', 'narrar'], to speech acts (e.g. 'say', 'communicate') ['decir', 'comunicar'], to the discourse structure (e.g. 'organize', 'structure') ['organizar', 'estructurar'], and to the discourse purpose of the text (e.g. 'demonstrate', 'convince') ['demostrar', 'convencer']. In general, they do not activate specialized meaning since they are not associated with a certain specialized field, but with the metalanguage used by experts to communicate. Some of these verbs can be hybrid in that they may also have a meaning related to the research methodology followed by experts (e.g. 'deduce', 'analyze', 'classify') ['deducir', 'analizar',

'clasificar']. Lorente maintains that discursive verbs should not be included in terminological resources. In line with Lorente, discursive verbs are not included in our analysis.

Connective verbs usually link two linguistic designations. They are usually copula verbs which express attributive qualities, or relations of equivalence, equality, similarity or dependence (e.g. 'be', 'seem') ['ser', 'parecer'], and are very frequently used in the definitions of concepts. As can be deduced, they do not convey specialized meaning either since their meaning is very similar to the meaning that they have in general language contexts. As such, according to Lorente, they should not be codified in specialized resources either, a premise we have also followed in our proposal.

Phraseological verbs are verbs which express actions, processes and states in specialized texts. Although their meaning does not differ from their meaning in non-specialized texts, it is their co-occurrence with terminological units that makes them specialized. As such, verbs such as 'generate', acquire their specialized meaning when they are included in phrases such as 'generate energy'. Therefore, they can become specialized knowledge units. According to Lorente, they should be included in specialized resources, but only when their meaning differs from their meaning in general language contexts. This is the kind of verbs that we are interested in and the ones that are encoded in the verb entry template presented in this thesis.

Finally, *terminological verbs* are those verbs whose meaning is specifically linked to a certain subject field (e.g. 'to cause to be eutrophic) ['eutrofizar'] linked to the domain of Ecology). These verbs are generally in correlation with nominal or adjectival lexical units (e.g. 'eutrofizar', in Spanish, is linked to the noun derivative 'eutrofización'). According to the author, noun and adjective derivations are the units that should be encoded in specialized resources. In EcoLexicon, these specialized units are encoded as separate concepts in the database. Therefore, such units are not within the scope of our study since they are already represented in EcoLexicon.

In 2007, Lorente redefines her verb typology. The labels coined for each category are thus changed to the following:

(i) *discursive verbs* become ***performative verbs*** [*verbs performatius del discurs*] since all verbs within a specialized text have a discursive function provided by the sender;

(ii) *connective verbs* become ***verbs of logical relations*** [verbs de relació lògica] since even though *connective verbs* are generally regarded as copulas, some predicative

verbs were also included in this category. In this sense, *verbs of logical relations* can be specialized knowledge units and can convey generic logical relations;

(iii) ***phraseological verbs*** keep the same designation although she clarifies that this category includes verbs that appear in collocations and fixed phrases as well as with support verb phrases constructions;

(iv) *terminological verbs* become ***quasi-terminological verbs*** [verbs quasiterms].

Performative verbs are associated with the domains of SPEECH (e.g. 'predict'), MENTAL PERCEPTION ('verify') and GENERIC PERCEPTION ('observe'). Verbs of logical relations are related to the domains of POSITION (e.g. 'engulf'), EXISTENCE (e.g. 'generate'), MANIPULATION (e.g. 'control'), and POSSESSION (e.g. 'unload'). Finally, phraseological verbs and quasi-terminological and, are clearly associated with the domain of action ('build') and change (e.g. 'destroy').

This association is far from random. As Lorente (2007) states in her new proposal, phraseological verbs express actions, processes or states, which are clearly linked to the domains of ACTION or CHANGE. This is the reason why phraseological verbs (in Lorente's terminology) are indeed among the most frequent verbs activated by the EXTREME EVENT.

Verbs can also be used to obtain other types of conceptual information. As such, verbs of MANIPULATION (e.g. 'measure') clearly activate relations of *instrument_of*. In contrast, CHANGE (e.g. 'destroy', 'devastate'), ACTION (e.g. 'hit', 'strike'), MOVEMENT (e.g. 'shake', 'sweep'), and POSITION (e.g. 'flood' 'engulf') activate the relation of *affects*. Therefore, a correspondence of domains and the relations most prototypically activated within the EXTREME EVENT can also be established, as shown in Table 71.

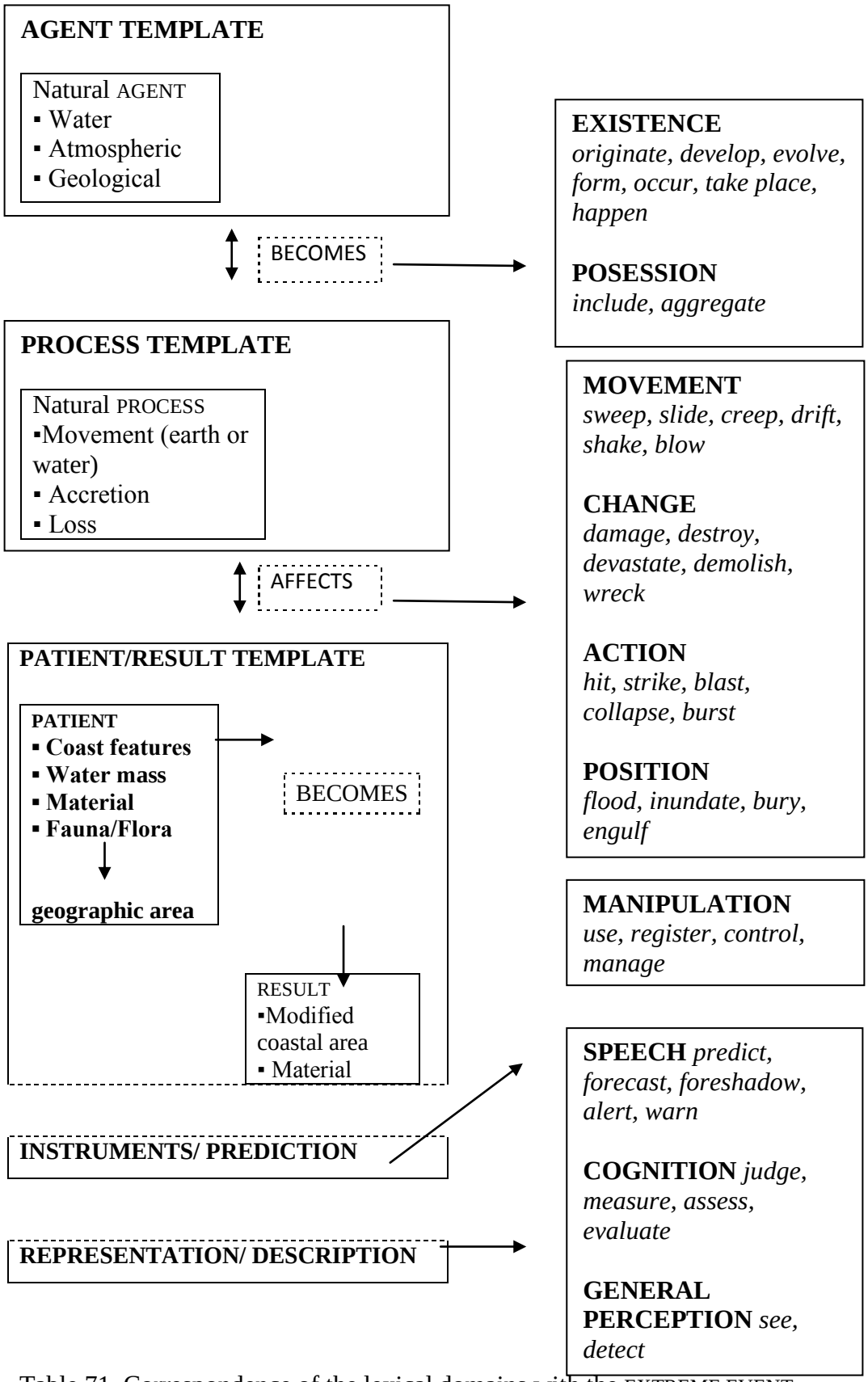


Table 71. Correspondence of the lexical domains with the EXTREME EVENT

5.4.2 Bottom-up analysis: study of concordances for the selected verbs

Once the meaning of the candidate verbs was extracted from dictionaries and verbs were preliminarily classified in lexical domains and subdomains according to their meaning, verb meaning was refined by studying their behavior in texts as reflected in concordances.

Concordance lines were extracted from our corpus and from the web. To this end, we used a Web as Corpus approach. More specifically, context examples were mainly retrieved from WebCorp (http://www.webcorp.org.uk/live/) and Exemplar (http://www.springerexemplar.com/), as well as from search engines, such as Google. For instance, when analysing the instances of 'hit' in the corpus, only one example pointed to its sense of negatively affecting humans in a cognitive/emotional way ('Drought, earthquakes, and other disasters do not hit people in the same way'). Consequently, in order to verify this meaning, 'hit' was searched with Exemplar. The number of hits found for this particular sense confirmed its relevance (e.g. 'Some 106 million people were hit by floods and 60 million by drought, whereas 1.6 million were affected by earthquakes'; 'Every donation, regardless of size, helps to rebuild communities that are hit by natural disasters').

A fixed number of concordance lines was not established for each verb. Rather, we studied all the instances found for each verb in our corpus, and as previously mentioned, went to the web for more instances when needed. Even though all the corpus instances were analyzed, only a selection of the most representative samples were included in the templates, that were the basis for the phraseological templates in EcoLexicon.

5.4.2.1 Argument analysis and description

In this bottom-up analysis, both arguments and predicates were identified and analyzed. The arguments were analyzed in the following steps: (i) identification of the arguments; (ii) identification of the linguistic realizations for the same argument and assignment of semantic labels/tags for the set of linguistic realizations for the same argument (§5.4.2.1.1); (iii) assignment of semantic roles and macroroles to arguments (§5.4.2.1.2); (iv) assignment of morphosyntactic structure (phrase types) to arguments (§5.4.2.1.3).

5.4.2.1.1 Linguistic realizations of arguments and assignment of semantic tags/labels for establishing conceptual categories

Linguistic realizations of an argument are the instantiations of an argument retrieved in texts. When various linguistic realizations refer to the same argument, then they designate the same kind of entity, evoke the same kind of conceptual structure, and have a similar semantic and syntactic behavior. This is the reason why their identification was very useful for the subsequent assignment of labels or tags and the establishment of semantic categories.

Semantic categories are generalizations for a set of terms that are assumed to have a similar semantic and syntactic behavior. In specialized language, verb meaning is more restrictive because of the constraints of specialized subject fields. Consequently, if arguments are classified and structured in a set of conceptual-semantic categories typical of a given domain, along with the semantic roles activated (§5.4.2.1.2), the range of verbs generally associated with a certain category could be predicted within the framework of a specialized event.

However, any typology of semantic categories is still a topic of debate. As previously mentioned, there have been many initiatives in general language, such as WordNet (§3.2.1), Gross's (1994) *classes d'objets*, and meaning-based resources such as ADESSE (§3.2.4) or VerbNet (§3.2.2) or ontologies like SUMO (Niles and Pease 2001). The frame elements proposed by FrameNet (§3.1.3.2) can also be regarded as categories on the basis of which it is possible to make generalizations about arguments. However, the fact that this inventory is open-ended and relies exclusively on the intuition of the annotator means that the roleset in FrameNet lacks descriptive adequacy and is unsuitable for our purposes.

In our study, categories have been preliminary established based on the following: (i) the semantic relations of the concepts expressed by the linguistic realizations in the corpus; (ii) its verification by means of tests inspired on Gross's (1994) *classes d'objets* (§3.1.1.1).

For instance, the definition of concepts, such as EARTHQUAKE or TROPICAL CYCLONE, in EcoLexicon, always start by the genus *natural disaster*. In this regard, NATURAL DISASTER is regarded as a category. A NATURAL DISASTER is defined in EcoLexicon as follows:

> adverse event that affects the environment in a relatively short space of time, and which causes human, material, economic or environmental looses, which exceed those affected to cope with it.

According to a top-down and bottom-up approach, the top-level concept NATURAL DISASTER is characterized by the following properties:

1. A natural disaster *causes* human/economic/material losses.
2. A natural disaster *affects* the environment.
3. A natural disaster *occurs* in a short period of time

In this regard, whenever a concept fulfils these three conditions, it belongs to the category of NATURAL DISASTER. This is verified with corpus information and pattern-based searches. As such, we can see that the referents of the terms 'earthquake', 'typhoon' or 'fire' all cause human, economic and material losses; they affect the environment; and they occur in a short period of time. Furthermore, when they take an active role within argument structure, they all appear with verbs, such as 'hit', 'strike', 'devastate', 'destroy', etc. As an example, Table 74 provides the potential linguistic realizations of the conceptual category NATURAL DISASTER, as well as a usage context for each linguistic realization.

Table 74 shows the main terms that instantiate the category of NATURAL DISASTER. In English, these include 'avalanche', 'drought', 'earthquake', 'quake', 'cyclone', 'hurricane', 'tropical cyclone', 'typhoon', 'tsunami', 'landslide', 'tornado', 'eruption', 'volcano', 'flood', 'flooding', 'fire', and 'natural disaster'*:*

Members of the category NATURAL DISASTER	**Illustrating examples**
avalanche	An **avalanche** struck two remote villages in northeast Afghanistan on Monday
drought	In Europe, over the past 30 years, many countries were hit hard by **droughts** and water scarcity.
earthquake	An **earthquake** had devastated some of the most deprived.
quake	**Quake** destroyed 250 school buildings in Ziarat.
hurricane	The Category 3 **hurricane** made landfall on September 8 on the north coast of eastern Cuba in the province of Holguín near Puerto de Sama, with sustained winds of about 120 mph (193 km/h).
typhoon	The Philippines was hit twice by a **typhoon**.
tropical cyclone	The town hall always offers a refugee whenever a **tropical cyclone** hits land.
cyclone	The **cyclone** crossed Chittagong coast and destroyed the whole area.
tsunami	A hundred people died, half of them drowned, as a **tsunami** hit their villages.
landslide	The **landslide** demolished nearly 300-tin-roofed mud huts in October.
tornado	Two rare **tornadoes** struck the outer edges of New York City at the weekend, leading to power cuts and debris strewn across the streets.
volcano	The **volcano** blasted out with a massive explosion.
eruption	The **eruptions** devastated populated areas on the southern part of the country.
flood	The 1997 **flood** devastated many communities upriver from Winnipeg.
flooding	**Flooding** has hit the regency of Mamuju in West Sulawesi destroying food crops after heavy rain over the past several days.
fire	The **fire** devastated everything in its path.
natural disaster	**Natural disaster** in North Korea destroys thousands of hectare field.

Table 72. English linguistic realizations of the conceptual category NATURAL DISASTER

The complete set of categories for our analysis is displayed in Table 73, along with linguistic realizations that instantiate the category. As shown in Table 73, the categories defined are those of NATURAL DISASTER, ATMOSPHERIC AGENT, WATER

AGENT, ATMOSPHERIC CONDITION, MATERIAL ENTITY, AREA, CONSTRUCTION, ENERGY, HUMAN BEING, LANDFORM, WATER COURSE, DEATH, DAMAGE, LOSS OF LIFE/PROPERTY, PLANT, and EXPLOSIVE.

CATEGORY	LINGUISTIC REALIZATIONS INSTANTIATING THE CATEGORY
NATURAL DISASTER	'avalanche', 'drought', 'earthquake', 'quake', 'cyclone', 'hurricane', 'tropical cyclone', 'typhoon', 'tsunami', 'landslide', 'tornado', 'eruption', 'volcano', 'flood', 'flooding', 'fire', 'natural disaster', 'tropical depression', 'tropical storm'
	'avalancha', 'sequía', 'terremoto', 'seísmo', 'sismo', 'ciclón', 'ciclón tropical', 'huracán', 'tifón', 'tsunami', 'maremoto', 'deslizamiento de tierra', 'corrimiento de tierra', 'erupción', 'volcán', 'inundación', 'fuego', 'incendio', 'desastre natural', 'temblor', 'alud', 'depresión tropical', 'tormenta tropical'
ATMOSPHERIC AGENT	'wind'
	'viento'
WATER AGENT	'storm', 'wave', 'snow', 'hail'
	'tormenta', 'ola', 'diluvio', 'lluvia', 'granizo', 'nieve'
ATMOSPHERIC CONDITION	'easterly wave', 'tropical wave', 'easterly wave', 'tropical easterly wave', 'African easterly wave', 'upper-level through', 'cold wave', 'heat wave'
	'borrasca fría', 'onda tropical', 'ola de calor', 'ola de frío', 'granizada', 'helada'
MATERIAL ENTITY	'ash', 'sulfur dioxide', 'cyanide', 'lava', 'basaltic lava', 'pahoehoe', 'gas', 'magma', 'pumice', 'tephra', 'dust', 'rock', 'smoke', 'steam', 'sulphurous gas', carbón dioxide', 'debris', 'pyroclastic material', 'volcanic ejecta'
	'gas', 'ceniza', 'arena', 'piedra', 'lava basáltica', 'fragmento', 'dióxido de azufre', 'ceniza volcánica', 'ácido sulfúrico', 'ácido fluorhídrico', 'humo', 'CO2', 'magma', 'lava incandescente', 'piroclasto', 'aerosol', 'vapor', 'nitrógeno', 'escombro'
AREA	'region', 'country'
	'poblado', 'área'
CONSTRUCTION	'dam'. 'levee', 'seawall', 'barrier'
	'dique', 'presa',

ENERGY	'wave energy'
	'energía'
HUMAN BEING	'children', 'person', 'farmer', 'people', 'baby', 'firefighter', 'fire tender'
	'persona', 'gente', 'niño', 'bombero', 'efectivo de la guardia civil', 'dotación', 'voluntario'
LANDFORM	'mountain', 'river', 'cliff', 'coast'
	'sistema montañoso', 'río'
WATER COURSE	'river'
	'río'
DEATH	'fatality', 'death', 'casualty'
	'muerto', 'despararecido'
DAMAGE	'damage', 'destruction', 'devastation'
	'daño', 'estrago', 'destrozo', 'deformación'
LOSS OF LIFE/PROPERTY	'homeless', 'loss of life'
	'pérdida material', 'pérdida humana'
PLANT	'crop'
	'cultivo', 'árbol', 'vegetación', 'plantación'
EXPLOSIVE	'bomb'
	'bomba'

Table 73. Semantic categories established for the EXTREME SUBEVENT

Nevertheless, it should be underlined that this is only a preliminary classification, which is a first step towards a more definitive inventory. In the near future our aim is to establish a complete set of categories for the domain of the ENVIRONMENT, which could be useful for the implementation of the incipient ontology currently being built for Ecolexicon.

5.4.2.1.2 Assignment of roles to the arguments

As previously mentioned, semantic roles generally express the set of properties that a verb entails for a given argument. The set of semantic roles in our study largely coincides with the most general thematic relations provided by Role and Reference Grammar (§3.1.2.2) and the argument roles as proposed by Goldberg (§3.1.3.3). Others were taken from the inventory proposed by VerbNet (§3.2.2), ADESSE (§3.2.4) and Sensem (§3.2.5). Apart from semantic roles, we have also annotated each argument with a semantic tag or semantic label (§5.2.2.1.1), which is a more specific characterization

of its meaning. Along with semantic categories and semantic roles, an additional categorization in terms of macroroles, as proposed by RRG (i.e. ACTOR and UNDERGOER), is also provided. The number of macroroles that a predicate can take is 0, 1 or 2.

Table 74 is a non-exhaustive inventory of semantic roles for the EXTREME EVENT. As shown, for every role a definition is provided as well as a usage example extracted from the corpus. Furthermore, its typical phrase type is stated (NP, for noun phrases, PP, for prepositional phrases, and AVP, for adverbial phrases). When the phrase type is a prepositional phrase, the most common prepositions that are likely to head the phrase are given both for English and Spanish:

Roles	**Definition, example and syntactic category (phrase type)**
AGENT	Human entity that carries out an action with intentionality (e.g. 'The **man** started the fire intentionally')
	NP
NATURAL FORCE	Unvolitional agent that causes an action (e.g. 'The **hurricane** destroyed the city')
	NP
DESTINATION	Endpoint of path (e.g. 'At the time of the final eruption, ash was ejected into the atmosphere **towards the north-east**')
	PP (into, to)
	PP (a, hacia)
EXPERIENCER	Animate entity which continues to exist or ceases to exist in a natural disaster. It is usually a human being or a personified construction, plant, or area (e.g. 'There were few **people** who survived the earthquake').
	NP
FREQUENCY	The rate at which something occurs over a particular period of time (e.g. 'Many of Earth's volcanoes have erupted **dozens of times** in the past few thousand years')
	AVP
GEOGRAPHICAL LOCATION	Underspecified place that usually designates where a situation takes place (e.g. 'The tornado started at 3:30 pm **about 1.2 miles north of Poplar Grove in Boone county** and ended at 3:48 pm')

	PP (around, at, in, off, on, over, etc.)
	PP (a, en)
MANNER	The way in which an action is carried out. (e.g. 'This was the start zone where the avalanche **suddenly** originated').
	AVP
PATH	Trajectory of the motion. (e.g. 'Sufficiently large volumes of gases were lofted **into the stratosphere**')
	PP (across, along, from…to, in, into, out, over, through, to, up, upon)
	PP (por, hacia, a lo largo de, a través de, a, entre…y, sobre)
PATIENT	Entity that is affected by a natural force. (e.g. 'The hurricane devastated **the coast'**)
	NP
SITUATION/ EXPERIENCE	Situational context in which an event occurs. (e.g. 'In addition to the ash, a large amount of sulfur dioxide gas and aerosols were ejected **in the eruption**')
	PP (during, after, in)
	PP (en, por, durante, debido a, tras, como consecuencia de)
ORIGIN	The location origin, i.e. starting point, of a motion/trajectory or the entity from which another entity forms. (e.g. 'A hurricane developed **from a tropical storm'**) (e.g. 'The 1919 eruption of Mt Kelat on Java expelled water **from a crater lake**, covering 200 km2 of farmland')
	PP (from)
	PP (de, a partir de)
THEME	Entity that undergoes a change of state (e.g. 'The **tropical storm** evolved into a hurricane'), which develops from sth else (e.g. 'The **hurricane** developed from an easterly wave'), which simply exists without entailing an action (e.g. 'The **hurricane** lasted 3 days'), or the thing or substance with which an area is filled (e.g. 'Fields were inundated **with water**).
	NP
	PP (in, under, with)
	PP (de, con, en)
TIME	The time at which the action occurs. (e.g. 'The volcano erupted **in November 1995**')
	PP (after, ago, at, between, by, during, every, for, from, in, just, on, over,

	to)
	NP, PP (en, por, entre, desde, hasta, durante)
RESULT	An entity that is caused or produced by something else, a consequence or outcome. (e.g. 'The hurricane developed **into a tropical storm** in the central Atlantic')
	PP (into)
	PP (en, a)

Table 74. Description of the set of semantic roles used in our analysis

Finally, it should be mentioned that the same linguistic realization associated with same semantic category, can convey different semantic roles. This is the case of 'hurricane', for example. As shown in Table 75, a 'hurricane' is always understood as a NATURAL DISASTER. However, depending on the verb with which it collocates, it can activate the semantic role of THEME, when it merely exists but does not entail action or change, or, activate the role of NATURAL FORCE when it entails action and produces damage as instantiated in the sentence 'The hurricane struck the city of New York'.

	Semantic category	**Semantic roles**	**Examples**
'hurricane'	natural disaster	THEME	'The **hurricane** evolved into a tropical storm'
	natural disaster	NATURAL FORCE	'The **hurricane** struck the city of New York'

Table 75. Various semantic roles conveyed by the linguistic realization 'hurricane'

5.4.2.1.3 Morphosyntax of arguments (phrase types)

In our template proposal, the morphosyntactic structure associated with each argument has been specified. In this regard, the following phrases were distinguished: (i) *noun phrase* (NP), for phrases having a noun as the head of the phrase and which function as a noun in the sentence; (ii) *prepositional phrases* (PP), for phrases headed by a preposition, specifying the specific preposition(s) activated; (iii) adverbial phrases (AVP), for those phrases whose head is an adverb and fulfills the function of an adverb in the sentence.

The analysis of the morphosyntactic behavior of arguments, along with the semantic characterization of arguments was used for the subsequent establishment of

rules regarding the possible systematization of semantic categories. As shall be seen, in the verbal phrase template proposed for EcoLexicon, there is no separate module for morphosyntactic information. However, because of the usefulness of prepositional phrases, especially for encoding purposes and more specifically, for non-native speakers, they can be easily identified, thanks to the wide range of usage contexts included for each verb or verb sense. Table 76 shows as an example of some of the usage examples for the verb 'burst', which includes the various prepositions that can accompany this verb:

▪ 'The hurricane burst in all its fury' ▪ 'A tsunami burst across the northern coast of Japan' ▪ 'When the hurricane burst upon us, many people run out to see what has happened' ▪ 'The hurricane burst through the city's levees' ▪ 'The oil refinery burst into flames' ▪ 'The volcano burst back to life' ▪ 'The volcano burst into life in early June'

Table 76. Usage examples of 'burst'

5.4.2.2 Verb/predicate description

Our study focuses on combinations formed by verb + noun or noun + verb within EcoLexicon, which we have referred to as *collocations* in a broad sense (§2.1.4). These elements frequently occur in combination and in the majority of cases, keep their meaning. In this context, the verb is activated because of the frame evoked by the meaning of the noun. At the same time, the verb constrains the semantic nature of nouns that can combine with it. For example, in the collocation 'volcano spits', 'volcano', requires a verb designating something being forced out (e.g. 'spit'). However, it is also true that 'spit' only admits noun phrases designating something being forced out of a mouth or a mouth-like orifice (e.g. 'volcano', in which the crater is considered to be the mouth).

5.4.2.2.1 Differentiating verb senses and their English-Spanish correspondences

The meaning of each verb was formulated on the basis of information in dictionaries (§5.4.1), as well as by analyzing concordances. In this section, we explain the criteria

followed to differentiate the various senses for each verb and how the correspondences between verbs in English and Spanish were established.

For each verb, we analyzed the total number of concordances retrieved in our corpus. All the concordances lines for each verb provided by TermoStat and encoded in Table 68, were recorded manually in an Excel file. Subsequently, the arguments of each verb were identified and assigned a semantic label. This concordance analysis allowed us to discard those verbs which were not directly related to the EXTREME EVENT. In this regard, the concordance analysis helped us retained only verbs and verb senses that encoded relations and attributes relevant to the EXTREME EVENT.

Then, all the verbs having the same semantic categories were grouped together, and assigned a name or *frame label*. Although these frame labels resemble the lexical subdomains of the LGM (§3.1.2.1), we decided to name them *frames* because FBT is a frame-based approach. Each frame was defined with its underlying lexical domain in mind (e.g. EXISTENCE, POSITION, CHANGE, or MOVEMENT) and, according to the semantic categories assigned to its arguments.

Nevertheless, further evaluation was necessary since not all the verbs encoded with the same categories conveyed exactly the same meaning. Sometimes, there were only slightly differences regarding their register, as is the case for the Spanish verb 'pasar' and 'sobrevenir'. Although they both include a NATURAL DISASTER that exist in a particular LOCATION, at a given TIME or in a specified MANNER, and belong to the frame *to exist in time*, 'sobrevenir' is more formal than 'pasar'. Other times, although the various verbs labeled with the same categories activated the same underlying meaning, certain verbs imposed certain constraints on their arguments. This is the case for the frame *to cause motion*. As shall be seen in 5.4.3.4, this frame describes a NATURAL DISASTER which causes MATERIAL to move quickly and forcefully in a certain direction path. However the NATURAL DISASTER in this frame is usually constrained to NATURAL VOLCANIC DISASTERS. As such, it is correct to say 'the **volcano** ejected lava', but not 'the **hurricane** ejected rain*'. Such cases are clearly stated in the *note section* within the analysis tables provided for each verb or verb sense both in English and Spanish (§5.4.3). Nevertheless, it was found that verbs belonging to the same semantic category can evoke different frames. This is the case of the frames *to come against sth with sudden force* (see Tables 112 and 113) and *to cause to change for the worse* (see Tables 122 and 123).

The verb analysis was first done for English. The Spanish analysis was subsequently performed using the same methodology. However, once the semantic categories in each Spanish concordance line were identified, they were directly associated with the frames in English which had the same category labels. As such, it can be said that the Spanish analysis was, to a certain extent, a bit easier. This methodology is in consonance with that of Pimentel (2012) in her assignment of specialized verb equivalents within the legal domain. In line with Pimentel (2012), verbs which shared the same type and number of semantic tags in the two languages were potential correspondences for each other. However, in contrast to Pimentel (2012) who accomplished this task semi-automatically, we performed it manually by examining each verb in Spanish and looking for its possible correspondence in English, based on its annotation. When there were differences in meaning or usage between verbs in the same frame, this was clearly stated in the corresponding *note section* of the verb. This means that the number of English and Spanish verbs in a frame does not always coincide.

5.4.3 Verb Information tables

Each verb is encoded in a template for the subsequent implementation of phraseological information in EcoLexicon. Each template provides a frame description, composed of the following elements: (i) the lexical domain activated by the frame; (ii) the name of the frame; (iii) the definition of the frame; (iv) a user note, when needed. The definition of the frame is based on the semantic categories or semantic roles activated by the arguments which co-occur which each specific verb. In addition, when necessary, a *note*, which explains the meaning of the verb or its usage, is added.

Subsequently, the arguments activated by each verb are encoded in the template and classified and described according to their linguistic realizations, their semantic category, semantic roles and macroroles, and phrase type. Along with the description of the frame and the arguments conveyed by each verb, usage examples are also provided. The number of samples is not fixed, but rather depends on the complexity of the verb sense.

Sentences are annotated by using different colors to identify the arguments in consonance with FrameNet. However, in contrast to FrameNet, only the heads of the phrases are annotated. In other words, instead of coloring the whole phrase (e.g. 'the powerful hurricane'), only the head of the phrase (i.e. 'hurricane') is annotated. The

reason for this is that for our purposes, the identification of the head of the phrase (i.e. the term) was more helpful for the assignment of the semantic category and the subsequent establishment of frame-associated templates. For noun phrases and adverbial phrases with the roles of GEOGRAPHICAL LOCATION, TIME, and MANNER, the objective was not so much the identification of the linguistic realizations, but rather the identification of the set of prepositions with which the heads of the phrase generally occur. For this reason, even though the heads of the phrases designating the arguments with the role of GEOGRAPHICAL LOCATION, TIME, or MANNER are those listed on the table (e.g. for 'The hurricane developed **in the Atlantic ocean**', the template includes 'Atlantic ocean'), our interest was in the prepositions that are used to activate this meaning. Accordingly, the preposition is clearly specified in the phrase type section of the Table (i.e. PP (*in*)). Table 77 below provides the guidelines used for color assignment.

NATURAL DISASTER ATMOSPHERIC AGENT WATER AGENT ATMOSPHERIC CONDITION SITUATION/ EXPERIENCE EXPLOSIVE
MATERIAL ENTITY
PATH
AREA
TIME
MANNER
CONSTRUCTION
DESTINATION
ENERGY
ORIGIN
HUMAN BEING
LANDFIELD, WATER COURSE
DEATH
DAMAGE
FREQUENCY
LOSS (OF LIFE/PROPERTY)
GEOGRAPHICAL LOCATION
PLANT

Table 77. Color assignment in the template description

As described in Table 77, the arguments are designated using a color code. As shown, NATURAL DISASTER, WATER AGENT, ATMOSPHERIC AGENT, ATMOSPHERIC CONDITION, SITUATION/ EXPERIENCE and, EXPLOSIVE were all coded in red because they generally have a similar role in the frame activated. Table 78 shows the design of the template for verb analysis in our research:

1	Lexical domain: Frame: Frame definition: Note:					
	Semantic role	VERB				
	Macrorole					
	Labels					
	Linguistic realizations					
	Phrase type					
	Usage examples	1. 2. 3. 4. 5.				

Table 78. Verb template

Therefore, two tables are displayed for each frame, one that contains all the English verbs in the frame and another table containing the Spanish verbs. In addition, each table can have various verb templates. After each frame description, there is a frame template that gives the requirements and restrictions of the frame.

5.4.3.1 Domain of EXISTENCE

Within the domain of EXISTENCE, verbs belonging to the following frames were analyzed: *to begin to exist*¸ *to begin to exist from sth else*, *to begin to exist becoming sth else*, *to cause to exist/happen*, *to exist in time*, *to continue to exist (of people)*, *to continue to exist (of natural disaster)*, *to cease to exist*, *to cease to exist in the perception of others*, *to cause sb to cease to exist*, and *to cause to cease to exist (of fire disaster).*

To begin to exist

The frame *to begin to exist* includes English verbs such as 'start2', 'break out', 'originate', 'develop', 'evolve', 'blow up', 'form', 'burst2', 'erupt1', 'explode2', and Spanish verbs such as 'empezar', 'despertar', 'comenzar', 'originarse', 'desarrollarse', and 'formarse'. As shown in Table 79 and Table 80, the first argument is generally filled by English terms such as 'avalanche', 'drought', 'earthquake', 'fire', 'tsunami', 'landslide', 'volcano' and Spanish terms, such as 'huracán', 'terremoto', 'incendio', 'sequía'. These terms belong to the semantic category of NATURAL DISASTER, which at the same time fulfil the role of THEME and macrorole of UNDERGOER. The other arguments either indicate GEOGRAPHICAL LOCATION, TIME or MANNER. Morphologically speaking, the THEME is normally a noun phrase; the GEOGRAPHICAL LOCATION is a prepositional phrase; TIME is a prepositional phrase; and MANNER, an adverbial phrase:

1	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	START$_2$	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster, natural disaster season	_____	_____	_____
	Linguistic realizations		fire season, hurricane season, hurricane, tornado	small islands, Boone county	earlier, 14 April, 3.30 pm, twenty ninth of September	suddenly
	Phrase type		NP	PP (in, about)	PP (on, at)	AVP
	Usage examples	1. Fire seasons will start earlier and will finish later in the year. 2. The most active phase of the eruption started on 14 April, resulting in about 0.25 km3 of ejected tephra. 3. The tornado started at 3:30 pm about 1.2 miles north of Poplar Grove in Boone county and ended at 3:48 pm. 4. This tragic hurricane started suddenly in the small islands of Cape Verde on the twenty ninth of September. 5. The Atlantic hurricane season starts on June 1 and ends on Nov 30 of every year.				
2	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	BREAK OUT	Theme	Geographical	Time	Manner

				location		
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		avalanche, drought, earthquake, fire, tsunami, landslide, volcano	Are, boarder, Villagio Mall, Indonesia	a day after, 2005, Friday, 1897, Monday morning	_____
	Phrase type		NP	PP (in, at)	PP (in, on)	_____
	Usage examples	1. An avalanche broke out Monday in Are just a day after alert levels were raised in the area. 2. In 2005, a severe drought broke out affecting fisherman and those who rely on the river for not only food, but basic water needs as well. 3. The 6.6-magnitude earthquake broke out in the boarder of Xinjiang's Xinyuan County and Jing County. 4. A fire unrelated to the hurricane broke out but was quickly brought under control. 5. As a result of the earthquake, a tsunami broke out as well. 6. About 1,400 people remained buried by thick mud and debris since the landslide broke out last Friday. 7. In 1897 a submarine volcano broke out near the north coast of Tonga. 8. A massive fire broke out at the Villaggio Mall around 11am on Monday morning. 9. A huge forest fire has broken out in Indonesia. 10. A major fire broke out in Central Delhi on Monday morning.				
3	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER]. Note: It normally focuses on the location where it begins to exist.					
	Semantic role	ORIGINATE	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		tropical cyclones, tsunamis, disasters, hazards	warm tropical seas, Japan and Kamchatka Peninsula, steep mountain areas	every summer	sudden-ly
	Phrase type		NP	PP (over, between, in)	AVP	AVP
	Usage examples	1. Tropical cyclones are defined as intense cyclonic storms that originate over warm tropical seas every summer. 2. In the Pacific, these disasters originate between Japan and the Kamchatka Peninsula inclusive. 3. Almost all landslides have originated in steep mountain areas. 4. This was the start zone where the avalanche suddenly originated. 5. On average, 850 tornadoes are reported annually of which 600 originate in the United States.				
4	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a					

	given [TIME] or in a specified [MANNER].					
	Semantic role	DEVELOP	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		drought, hurricane, landslide	parts of East Africa, Indian ocean	late 2010, September 1, late December	quickly
	Phrase type		NP	PP (in)	PP (on, in)	AVP
	Usage examples	1. Severe drought developed quickly in parts of East Africa in late 2010. 2. The hurricane developed in the western Caribbean Sea on September 1, and tracked northeastward. 3. Landslide developed north-east of Lansjärv in northern Sweden. 4. In late December of 2004 after a terrible tsunami developed in the Indian ocean, many lives were changed forever.				
5	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER]. Note: This is not the most common meaning of *evolve*. It normally comes followed by *from* or *into*. The NATURAL DISASTER is generally associated with disasters that develop and change, but not with disasters that occur suddenly (e.g. an earthquake).					
	Semantic role	EVOLVE	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		tropical cyclone, tornado	tropical and subtropical waters, the inland continental areas	summer through late fall,	rapidly
	Phrase type		NP	PP (in, over)	PP (in, through)	AVP
	Usage examples	1. Tropical cyclones evolve in tropical and subtropical waters, in summer through late fall. 2. Tornados evolve over the inland continental areas. 3. Waterspouts are tornadoes that evolve rapidly over warm waters, most frequently along the southeastern shores and the Gulf coast. 4. Tropical cyclones evolve in baroclinic regions, i.e., regions with a large-scale, typically meridionally oriented, horizontal temperature contrast.				
6	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a certain [MANNER]. Note: The NATURAL DISASTER can also be an ATMOSPHERIC AGENT, but it is usually associated with wind events.					
	Semantic role	BLOW UP	Theme	Geographical location	Time	Manner

	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster [wind], atmospheric agent [wind]	_____	_____	_____
	Linguistic realizations		hurricane, storm	Gulf	overnight, couple of days back, very night of the first day	quickly, like an atom bomb
	Phrase type		NP	PP (in)	PP (back)	AVP
	Usage examples	1. A hurricane had blown up in the Gulf. 2. Katrina, a "category one" hurricane, had blown up overnight into a massive killer. 3. A big storm had blown up a couple of days back. 4. A second storm had blown up to the south of this one and was becoming very strong. 5. Unfortunately, the wind blew up very quickly and so we had a rather fraught move onto the pontoon! 6. Tornado blows up like an Atom Bomb. 7. The very night of the first day a typhoon blew up.				

7	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	FORM	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer		_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		volcano, tropical storm, landslide, flood, hurricane, tornado	ocean, Florida, coast of Congo	1887, 1993, 27 October	in a number of other ways
	Phrase type		NP	PP (in, over, off)	PP (during, in, on)	AVP (in)
	Usage examples	1. Volcanoes that have formed in the ocean planed off at the top because of close proximity to the ocean's surface. 2. During the 1887 season, 19 tropical storms formed, of which a record 4 occurred after 1 November. 3. When it does reach the surface, a volcano is formed. 4. Twenty-one Atlantic tropical storms formed in 1933, a record only recently exceeded in 2005. 5. The smallest storm on record, Tropical Storm Marco, formed during October 2008, and made landfall in Veracruz. 6. Hurricane Grace formed on 27 October and moved northward along the east coast. 7. Fifteen tornadoes formed over Florida, accounting for 44 of those killed. 8. The tropical storm formed off the Coast of Congo in mid-April of				

		1991. 9. Tsunamis may be formed in a number of other ways, including non-seismogenic submarine landslides, sub-aerial landslides into the ocean or other body of water, and volcanic eruptions.

8 Lexical domain: EXISTENCE
Frame: to_begin_to_exist
Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].
Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake.

Semantic role	$BURST_2$	Theme	Geo. location	Time	Manner
Macrorole		Undergoer	_____	_____	_____
Labels		natural disaster [earth]	_____	_____	_____
Linguistic realizations		volcano, hurricane, fire	Italy	May 20th 1883, night of August 26	fiercely, in all its fury, with a tremendous explosion
Phrase type		NP	PP (in)	PP (on)	AVP

Usage examples
1. The volcano burst out fiercely in Italy.
2. On May 20th 1883 the volcano burst out with a tremendous explosion.
3. It happened before the volcano burst out in all its fury, on the night of August 26.
4. The volcano burst again.
5. The volcano burst into life in early June.
6. The fire burst into life.
7. The hurricane burst in all its fury.

9 Lexical domain: EXISTENCE
Frame: to_begin to exist
Frame definition: [NATURAL DISASTER] begin to exist/happen in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].
Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake.

Semantic role	$ERUPT_1$	Theme	Geo. location	Time	Manner	Frequency
Macrorole		Undergoer	_____	_____	_____	_____
Labels		natural disaster [earth]	_____	_____	_____	_____
Linguistic realizations		volcano	Philippi-nes, off the coast	past few thousand years, Holocene times, 1991, May 10	explos-ively, immed-iately, conti-nuously,	several times, relatively frequently, median rate of once every 220 years
Phrase type		NP	PP (in, off)	PP (in, on, at, ago, for)	AVP	NP (at)

	Usage examples	1. For example, many of Earth's volcanoes have erupted dozens of times in the past few thousand years. 2. Izalco in El Salvador erupted on 10 May, Masaya in Nicaragua erupted in July, and Santa Maria in Guatemala exploded on 24 October. 3. Over the last 500 years, individual volcanoes have erupted at the median rate of once every 220 years. 4. For example on 29 November 1975, Kilauea volcano erupted 30 minutes after a 7.5 magnitude earthquake. 5. Most scientists consider a volcano active if it has erupted in holocene times. 6. There is evidence that Santorini may have erupted several times over a time span of two hundred years. 7. The Mount Pinatubo erupted in the Philippines in 1991. 8. The volcano has erupted explosively at least 12 times during the last two hundred years. 9. Cordon Caulle volcano in Central Chile erupted two days after the great 9.5 earthquake. 10. Immediately, the volcano erupted and sent out a lateral explosion, which was felt 425 km away. 11. During the morning of 17 January 2002, Mount Nyiragongo erupted with lava flows spewing in several directions. 12. The Stromboli strato volcano off the Coast of Sicily has erupted continuously for thousands of years. 13. Among them, the Eyjafjallajökull volcano has erupted relatively frequently in the last millennia 14. In the former case, volcanoes have erupted in Australia in the past 10,000 years.

10 Lexical domain: EXISTENCE
Frame: to_begin to exist
Frame definition: [NATURAL DISASTER] begin to exist/happen in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].
Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake.

Semantic role	$EXPLODE_2$	Theme	Geographical location	Time	Manner	Freq.
Macrorole		Undergoer	_____	_____	_____	_____
Labels		natural disaster [earth]	_____	_____	_____	_____
Linguistic realizations		volcano	Guatemala, the islands of Java and Sumatra	24 October, August of this year	suddenly, violently	eight times
Phrase type		NP	PP (in, between)	PP (on, by)	AVP	AVP

Usage examples	1. Masaya in Nicaragua erupted in July, and Santa Maria in Guatemala exploded on 24 October. 2. On August 26-27, 1883, the Krakatoa volcano exploded between the islands of Java and Sumatra. 3. The volcano exploded suddenly and violently, killing six scientists and three tourists. 4. This volcano had exploded eight times by August of this year.

Table 79. Analysis of the frame *to begin to exist* (English)

<table>
<tr><td>1</td><td colspan="5">Lexical domain: EXISTENCE
Frame: to_begin_to_exist
Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a certain [MANNER].
Note: The NATURAL DISASTER can also be an ATMOSPHERIC AGENT, but it is usually associated with wind events.</td></tr>
<tr><td></td><td>Semantic role</td><td>EMPEZAR</td><td>Theme</td><td>Geographical location</td><td>Time</td><td>Manner</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, atmospheric agent</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>deslizamiento, fuego, erupción, sequía, tornado, terremoto, tsunami</td><td>oeste amazónico</td><td>1956, una misma semana, 19 de octubre, 4h 31 de la madrugada, medianoche, julio del año pasado</td><td>de pronto, con violencia</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>PP (en)</td><td>PP (en, durante, entre…y) NP</td><td>ADV</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="4">1. El deslizamiento empezó en 1956 durante la construcción de una carretera local.
2. Doce incendios importantes empezaron en una misma semana.
4. El incendio empezó el 19 de octubre.
5. El terremoto empezó de pronto y duró dos minutos.
6. El terremoto empezó con una violencia atroz.
7. Este terremoto empezó a las 4h 31 de la madrugada y duró unos 40 segundos.
8. El tsunami empezó 20 minutos después del terremoto.
9. La erupción empezó cerca de la medianoche y causó el incendio de matorrales en los alrededores sin poner en riesgo a los vecinos.
10. La sequía empezó en el oeste amazónico y ya se propaga a otros estados brasileños como Pará, al este.
11. La sequía empezó en julio del año pasado y ha provocado una escasez de agua potable que afecta a cerca de seis millones de personas.
12. El tornado empezó entre las seis y las seis y media de la mañana.
13. La temporada de huracanes en el Atlántico empieza el 1 de junio y termina el 30 de noviembre.</td></tr>
<tr><td>2</td><td colspan="5">Lexical domain: EXISTENCE
Frame: to_begin to exist
Frame definition: [NATURAL DISASTER] begin to exist/happen in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].
Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake.</td></tr>
<tr><td></td><td>Semantic role</td><td>DESPERTAR</td><td>Theme</td><td>Geo. location</td><td>Time</td><td>Manner</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster</td><td>_____</td><td>_____</td><td>_____</td></tr>
</table>

			[earth]			
	Linguistic realizations		volcán	islas	semana pasada, 27 de noviembre, pasada noche	con poder aterrador
	Phrase type		NP	PP (en)	NP	AVP
	Usage examples	1. El vulcanólogo dijo que el volcán despertó la semana pasada. 2. El volcán despertó el 27 de noviembre, tras 36 años de reposo. 3. El volcán despertó la pasada noche en las islas tras permanecer inactivo durante varios años. 4. Después de un siglo largo letargo, el volcán despertó con poder aterrador en una erupción que enterró varias ciudades romanas como Pompeya y Herculano.				

3	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	COMENZAR	Theme	Geo. location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster, atmospheric condition	_____	_____	_____
	Linguistic realizations		actividad de terremotos, incendio, tormenta, terremoto, tsunami	Cobija	mitad de marzo, segunda semana de octubre, costas del noroeste de Japón, 16.00 horas, 5 minutos después	repentina-mente
	Phrase type		NP	PP (en)	PP (a) NP	AVP
	Usage examples	1. En el caso del monte Santa Helena, la actividad de terremotos comenzó a mitad de marzo. 2. El incendio de Cedar comenzó el 25 de octubre de 2003. 3. Las inundaciones comenzaron la segunda semana de octubre. 4. La tormenta comenzó en torno a las 16.00 horas y durante una hora el agua cayó con rachas de fuerte intensidad. 5. El terremoto comenzó repentinamente en las costas del noroeste de Japón y llegó a sentirse en Tokio. 6. En Cobija, el tsunami comenzó 5 minutos después del terremoto, provocando un lento ascenso del nivel del mar que alcanzó entre 10 y 15 metros.				

4	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	ORIGINAR-SE	Theme	Geo. location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster, atmospheric	_____	_____	_____

			condition			
	Linguistic realizations		ciclón tropical, huracán, terremoto, tornado, erupción	cerca de la línea del Ecuador, océanos, tierra, 610 km de profundidad, Caribe, una larga dislocación de la corteza terrestre	15.50 horas, antes de las 11 horas, día de ayer	de repente
	Phrase type		NP	PP (a, en)	PP (a, en) AVP	AVP
	Usage examples	1. Los ciclones tropicales se originen cerca de la línea del Ecuador. 2. El huracán se origina en los océanos y los tornados en la tierra. 3. El reciente terremoto de Albuñuelas (Granada) de 11 abril 2010, con magnitud 6.3 se originó a 610 km de profundidad. 4. Las erupciones fisurales se originan en una larga dislocación de la corteza terrestre, que puede ser desde apenas unos metros hasta varios km. 5. El huracán se originó en el Caribe para seguir una trayectoria hacia el noreste del Atlántico. 6. Este huracán se originó en la región central del Golfo de México y alcanzó la categoría de huracán de moderada intensidad en la tarde del 27 de octubre. 7. La erupción se originó a las 15.50 horas de ese día domingo, iniciando un ciclo eruptivo que se prolongó por 9 meses. 8. El incendio, se originó poco antes de las 11.00 horas en una zona próxima al club de golf. 9. Un tornado se originó de repente en el día de ayer en el Estrecho de Gibraltar.				
5	Lexical domain: EXISTENCE Frame: to_begin_to_exist Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	DESARRO-LLARSE	Theme	Geo. location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster, atmospheric condition	_____	_____	_____
	Linguistic realizations		huracán, tormenta tropical, borrasca, incendio, tornado	medio del océano, niveles altos	14:07, la noche de este domingo	con fuerza
	Phrase type		NP	PP (en)	NP	AVP
	Usage examples	1. El NHC no lanzó ninguna alerta específica dado que el huracán se desarrolla en medio del océano. 2. El incendio se desarrolló a las 14:07 del lunes 7 de mayo de 2012, provocado por un tanquero que transportaba químicos inflamables. 3. Un incendio se desarrolló con fuerza la noche de este domingo en la Población Libertad de Puerto Montt. 4. Poco después del mediodía hora local de este martes, un enorme				

<table>
<tr><td></td><td></td><td colspan="5">tornado se desarrolló en la región de Dallas-Fort Worth.
5. La séptima tormenta tropical se desarrolla sobre el atlántico norte ganando fuerza rápidamente.</td></tr>
<tr><td>6</td><td colspan="6">Lexical domain: EXISTENCE
Frame: to_begin_to_exist
Frame definition: [NATURAL DISASTER] begins to exist in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].
Note: The NATURAL DISASTER is generally associated with disasters that develop and change, but not with disasters that occur suddenly (e.g. an earthquake).</td></tr>
<tr><td></td><td>Semantic role</td><td>FORMARSE</td><td>Theme</td><td>Geo. location</td><td>Time</td><td>Manner</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, atmospheric condition</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>tormenta tropical, tifón, huracán, depresión tropical, sistema tormentoso</td><td>Atlántico tropical, costa</td><td>23 de noviembre, 1996, marzo de 2004</td><td>_____</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>PP (a, en)</td><td>PP (a, en)
AVP</td><td>AVP</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="5">1. La tormenta tropical Delta se formó el 23 de noviembre en el Atlántico tropical.
2. El Tifón Nancy se formó el 12 de septiembre de 1961.
3. Un sistema tormentoso que parecía similar a un huracán se formó en 1996, en el lago Hurón.
4. En marzo de 2004, un huracán sí se formó en el Océano Atlántico Sur y tocó tierra en Brasil.
5. Una fuerte depresión tropical/tormenta tropical débil se formó fuera de la costa del Congo a mediados de abril de 1991.
6. El tifón se formó a poca distancia de la costa noroeste de la isla de indonesia.</td></tr>
</table>

Table 80. Analysis of the frame *to begin to exist* (Spanish)

It is true that the various verbs analyzed here are not completely synonyms, but rather activate the same meaning. As previously mentioned, as the hierarchy becomes more specific, the information becomes more focalized, and the contexts of use are thus more restricted. When this happens, it is stated in the *note section* within the frame description of the template. For example, 'erupt1' and 'explode2' belong to the frame of *to begin to exist.* However, unlike 'start', which is used in a wide variety of contexts, 'erupt1' and 'explode2' are restricted to volcanic events.

The generalizations that can be specified for the frame *to begin to exist* are displayed in Table 81. As shown, when a verb has as its first argument a NATURAL DISASTER or ATMOSPHERIC CONDITION, which fulfils the role of THEME and macrorole of

UNDERGOER, and can either have as second argument a GEOGRAPHICAL LOCATION, TIME or MANNER, it activates the meaning and syntax of the frame *to begin to exist*.

Frame: **to begin to exist**					
Semantic role	Theme	*start2* *break out* *originate* *develop* *evolve* *blow up* *form* *burst2* *erupt1* *explode2* *empezar* *despertar* *comenzar* *originarse* *desarrollarse* *formarse*	Geographical location	Time	Manner
Macrorole	Undergoer		_____	_____	_____
Conceptual class	natural disaster, atmospheric condition		_____	_____	_____
Phrase type	NP		PP (around, at, from, in, off, on, over) (English) PP (a, en) (Spanish)	PP (after, ago, at, between, by, during, every, for , from, in , just, on, over, to) (English) NP, PP (en, por, entre, desde, hasta, durante) (Spanish)	AVP

Table 81. Template for the frame *to begin to exist*

To begin to exist from sth else

Certain verbs in the previously described frame also appear in *to begin to exist from sth else* (Table 82 and Table 83) and *to begin to exist becoming sth else* (Table 85 and Table 86). However, they have different semantic arguments because of the specification of their origin ('from X' in the frame *to begin to exist from sth else*) and of the new entity that they become ('into X' in the frame *to begin to exist becoming sth else*). In *to begin to exist from sth else*, the first argument is also a NATURAL DISASTER, conveying the meaning of THEME and macrorole of UNDERGOER. However, the second argument is a NATURAL DISASTER or ATMOSPHERIC CONDITION, which has the role of ORIGIN and macrorole of ACTOR.

1	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	ORIGINATE FROM	Theme	Origin
	Macrorole		Undergoer	Actor
	Labels		natural disaster	natural disaster, atmospheric condition
	Linguistic realizations		tsunami, tropical cyclone, tornado	easterly wave, earthquake, storm
	Phrase type		NP	NP
	Usage examples	1. Most tsunamis originate from submarine seismic disturbances. 2. Tropical cyclones originate from an easterly wave. 3. Most tsunamis originate from earthquakes. 4. These types of floods originate from storms.		
2	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	DEVELOP FROM	Theme	Origin
	Macrorole		Undergoer	Actor
	Labels		natural disaster	natural disaster, atmospheric condition
	Linguistic realizations		hurricane, tsunami	earthquake, tropical wave
	Phrase type		NP	NP
	Usage examples	1. The hurricane developed from a tropical wave that moved off the coast of Africa on August 14. 2. The tsunami developed from an earthquake that occurred at a subduction zone off the Sumatran coast.		
3	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	EVOLVE FROM	Theme	Origin
	Macrorole		Undergoer	Actor
	Labels		natural disaster,	natural disaster, atmospheric condition
	Linguistic realizations		tropical cyclone, tornado	depression, storm
	Phrase type		NP	NP
	Usage examples	1. The third tropical cyclone evolved from a depression which had originated just west of the Azores. 2. This tornado evolved from the same storm that produced the Carter, SD tornado.		
4	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	FORM FROM	Theme	Origin
	Macrorole		Undergoer	Actor

	Labels		natural disaster	natural disaster, atmospheric condition
	Linguistic realizations		tsunami, tropical cyclone, tornado	easterly wave, upper-level through
	Phrase type		NP	NP
	Usage examples	1. The Indian Ocean tsunami, formed from an undersea earthquake on December 26, 2004, was indeed one of the deadliest natural disasters in recent times. 2. This tropical cyclone formed from the combination of an easterly wave and an upper-level through.		

Table 82. Analysis of the frame *to begin to exist from sth else* (English)

1	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	ORIGINARSE A PARTIR DE	Theme	Origin
	Macrorole		Undergoer	Actor
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		huracán, ciclón	onda tropical
	Phrase type		NP	NP
	Usage examples	1. El huracán Earl se originó a partir de una onda tropical el 25 de agosto. 2. El ciclón se originó a partir de una onda tropical el 4 de junio en el Golfo de México.		
2	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	DESARROLLARSE A PARTIR DE	Theme	Origin
	Macrorole		Undergoer	Actor
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		huracán, borrasca fría	depresión fría, ondas tropicales
	Phrase type		NP	NP
	Usage examples	1. La borrasca fría se desarrolló a partir de una depresión fría en niveles altos. 2. Este tipo de huracán se desarrolla típicamente a partir de las ondas tropicales que se forman en la sabana africana durante la estación lluviosa.		
3	Lexical domain: EXISTENCE Frame: to_begin_to_exist_from_sth_else Frame definition: [NATURAL DISASTER] begins to exist from another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	FORMARSE DE/ A PARTIR DE	Theme	Origin
	Macrorole		Undergoer	Actor
	Labels		natural disaster, atmospheric	natural disaster, atmospheric condition

			condition	
	Linguistic realizations		huracán, depresión tropical, tormenta tropical	depresión tropical, onda tropical
	Phrase type		NP	NP
	Usage examples	1. El tercer huracán se formó de una tormenta tropical en las Antillas Menores, fenómeno que comenzó el 5 de agosto y se convirtió en huracán el día 8. 2. La sexta depresión tropical, quinta tormenta tropical, y primer huracán, se formó a partir de una onda tropical el 1 de septiembre. 3. El Huracán Rafael fue la decimoséptima tormenta tropical de la temporada de huracanes en el Atlántico de 2012 y se formó a partir de una depresión originada a unos 800 km al SW de las islas de Cabo Verde del 5 al 8 de octubre de 2012.		

Table 83. Analysis of the frame *to begin to exist from sth else* (Spanish)

This analysis shows that the same verbs contained in the previous frame can activate different arguments with different semantic information. This new semantic content is syntactically represented in English and Spanish by a prepositional phrase. In this regard, whenever a verb fulfils the conditions displayed in the template below (Table 84), it will convey the meaning of the frame *to begin to exist from sth else*.

Frame: **to begin to exist from sth else**			
Semantic role	Theme	*originate* *develop* *evolve* *form* *originarse* *desarrollarse* *formarse*	Origin
Macrorole	Undergoer		Actor
Conceptual class	natural disaster		natural disaster, atmospheric condition
Phrase type	NP		PP (from) (English) PP (a partir de, de) (Spanish)

Table 84. Template for the frame *to begin to exist from sth else*

To begin to exist becoming sth else

This frame is similar to the preceding one. The first argument is the same (NATURAL DISASTER, THEME, UNDERGOER). However, the second argument is a NATURAL DISASTER or ATMOSPHERIC CONDITION which is the transformed entity or RESULT (see Tables 85 and 86):

1	Lexical domain: EXISTENCE Frame: to_begin_to_exist_becoming_sth_else Frame definition: [NATURAL DISASTER] or [ATMOSPHERIC CONDITION] begins to exist, becoming another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	DEVELOP INTO	Theme	Result
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		drought, landslide, hurricane	tropical storm, famine
	Phrase type		NP	NP
	Usage examples	1. In 2011, the world watched as a drought developed into a fully-blown famine because of the behavior of Somali militants Al-Shabaab. 2. The hurricane developed into a tropical storm in the central Atlantic.		
2	Lexical domain: EXISTENCE Frame: to_begin_to_exist_becoming_sth_else Frame definition: [NATURAL DISASTER] or [ATMOSPHERIC CONDITION] begins to exist, becoming another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	EVOLVE INTO	Theme	Result
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		tropical cyclone, funnel clouds	extratropical cyclone, tornado
	Phrase type		NP	NP
	Usage examples	1. A tropical cyclone can cease to have tropical characteristics and evolve into extratropical cyclones. 2. In addition, its interaction with the main belt of the Westerlies, by means of merging with a nearby frontal zone, can cause tropical cyclones to evolve into extratropical cyclones. 3. Tornadoes often begin as funnel clouds with no associated strong winds at the surface, and not all funnel clouds evolve into tornadoes.		

Table 85. Analysis of the frame *to begin to exist becoming sth else* (English)

1	Lexical domain: EXISTENCE Frame: to_begin_to_exist_becoming_sth_else Frame definition: [NATURAL DISASTER] or [ATMOSPHERIC CONDITION] begins to exist, becoming another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	TRANSFORMAR-SE EN	Theme	Result
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		huracán, tormenta tropical	tormenta tropical, depresión tropical, huracán
	Phrase type		NP	NP
	Usage examples	1. El huracán se transformó en tormenta tropical (no menos destructiva), con lluvias torrenciales que causaron serios daños en los pueblos costeros. 2. Durante las 24 horas siguientes a su entrada en tierra firme, el huracán se transformó en tormenta tropical y luego en depresión tropical. 3. La tormenta tropical se transformó en huracán el 24 de octubre y a los 2 días ya había alcanzado la categoría 5 en la Escala Saffir-Simpson.		
2	Lexical domain: EXISTENCE Frame: to_begin_to_exist_becoming_sth_else Frame definition: [NATURAL DISASTER] or [ATMOSPHERIC CONDITION] begins to exist, becoming another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	CONVERTIRSE EN	Theme	Result
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		huracán, tormenta tropical	huracán de categoría 2, tormenta post-tropical, huracán de categoría 1, huracán
	Phrase type		NP	NP
	Usage examples	1. Más tarde esa noche, el huracán se convirtió en un huracán de categoría 2 con vientos de 105 mph sostenidos. 2. Con vientos de 128 Km/h, a las 19:00 horas de la tarde el huracán se convirtió en tormenta post-tropical, con el centro a 10 Km de Atlantic City. 3. A punto de llegar a la isla caribeña de Jamaica, la tormenta tropical se convirtió en el décimo huracán de categoría uno. 4. La tormenta tropical se convirtió en huracán con rumbo a Cuba.		
3	Lexical domain: EXISTENCE Frame: to_begin_to_exist_becoming_sth_else Frame definition: [NATURAL DISASTER] or [ATMOSPHERIC CONDITION] begins to exist, becoming another [NATURAL DISASTER] or [ATMOSPHERIC CONDITION].			
	Semantic role	EVOLUCIONAR A	Theme	Result
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric condition
	Linguistic realizations		tormenta tropical, depresión tropical	tormenta tropical, huracán
	Phrase type		NP	NP

	Usage examples	1. Al entrar a las aguas calientes del Golfo de México la entonces tormenta tropical evolucionó a un huracán que alcanzó la categoría III en la escala Saffir-Simpson. 2. La depresión tropical evoluciona a tormenta y se llama "Nicole".

Table 86. Analysis of the frame *to begin to exist becoming sth else* (Spanish)

As shown in Table 87, the English verbs in *to begin to exist becoming sth else* are 'develop into', 'evolve into' and the Spanish verbs are 'transformarse en', 'convertirse en', and 'evolucionar a':

Frame: to begin to exist becoming sth else			
Semantic role	Theme	*develop* *evolve* *transformarse* *convertirse* *evolucionar*	Result
Macrorole	Undergoer		_____
Conceptual class	natural disaster, atmospheric condition		natural disaster, atmospheric condition
Phrase type	NP		PP (into) English PP (en, a) Spanish

Table 87. Template of the frame *to begin to exist becoming sth else*

To cause to exist/happen

The frame *to cause to exist/happen* is defined as a NATURAL FORCE or volitional AGENT that causes a PATIENT to exist or happen. The PATIENT can be another NATURAL DISASTER, ATMOSPHERIC CONDITION, ATMOSPHERIC AGENT, WATER AGENT, DAMAGE, LOSS OF LIFE PROPERTY, or DEATH, depending on the verb. The English verbs in the frame are 'cause', 'start1', 'form2', 'generate', 'produce', 'spawn', 'result (from/in)', 'trigger', 'spread2' (Table 88), and those in Spanish are 'causar', 'generar', 'producir1', 'provocar', 'ocasionar', 'originar' (Table 89):

1	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE/AGENT] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent, water agent, or atmospheric condition. The AGENT is usually a human being. The PATIENT is usually a natural disaster, atmospheric agent, water agent, atmospheric condition, damage, loss of property/life or death entity.			
	Semantic role	CAUSE	Natural force/Agent	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric agent, water agent, atmospheric condition, human being	natural disaster, atmospheric agent, water agent, atmospheric condition damage, loss of property/life, death
	Linguistic		earthquake, floods,	damage, fatalities, death,

	realizations		storm, rain, landslide, forest fires, fires, humans	tsunami, disasters, floods, flooding, landslides
	Phrase type		NP	NP
	Usage examples	1. This chapter focuses on direct damage caused by earthquakes. 2. Most destructive tsunamis are caused by earthquakes of magnitude 7.5 or more. 3. Most of the disasters are thereby caused by either floods or storms. 4. This figure shows the number of fatalities caused by floods in the period 1970-2008. 5. Italy spent approximately EUR 146 billion between 1957 and 2000 as a result of damage caused by landslides and floods. 6. Three days of heavy rain caused floods that forced evacuation. 7. Heavy rains in late May and early June caused extensive flooding. 8. The country has failed to bring down the number of fatalities or the damage caused by forest fires. 9. Fires caused the deaths of 15 people in Italy and 12 people in Croatia. 10. The fires in Greece caused a high number of fatalities. 11. The same rainfall event caused numerous landslides. 12. In Rio de Janeiro 1966, record-breaking rainfall in January and March caused catastrophic landslides. 13. It is evident that the potential damage caused by landslides is often aggravated as a result of land use management. 14. Fires in Europe are caused especially by humans. 15. Earthquakes are caused mostly by rupture of geological faults. 16. The hurricane caused severe damage to shipping in Halifax Harbour. 17. This hurricane caused heavy casualties and extensive destruction along its path from the Leeward Islands to Florida. 18. The hurricane caused 200 deaths and approximately 200,000 homeless. 19. The hurricane caused severe destruction across the entire Mississippi coast and into Alabama. 20. A tsunami is a series of huge waves that can cause great devastation and loss of life when they strike a coast. 21. The hurricane caused winds gusts of up to 105 mph (169 km/h), heavy rainfall, and a storm surge of 6 feet (2 m).		
2	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent, water agent, or atmospheric condition. The PATIENT is usually a natural disaster, atmospheric agent, water agent, damage, death, material entity.			
	Semantic role	PRODUCE	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric agent, water agent, atmospheric condition	natural disaster, atmospheric agent, water agent, atmospheric condition damage, loss of property/life, death
	Linguistic realizations		earthquake, landslide, floods, volcanic eruption, eruption, el Niño event	waves, tsunami, hazardous phenomena, fatalities, tephra, liquid rock

	Phrase type		NP	NP
	Usage examples	1. Large waves produced by an earthquake or a submarine landslide can overrun nearby coastal areas. 2. Many past earthquakes in southern California produced little damage because they occurred in unpopulated area. 3. The tsunami events of 21–22 May 1960 were produced in Chile by over four dozen earthquakes. 4. It corresponded to the catastrophic tsunamis produced by the great Sumatra earthquake on December 26, 2004. 5. In Switzerland, floods produced 6 fatalities and some EUR 2 billion in losses. 6. There are many hazardous phenomena produced directly, or as secondary effects, by volcanic eruption 7. The 186 AD eruption of Taupo produced 60 100 km3 of tephra. 8. It refers to the most common liquid rock produced by a volcano, containing mainly magnesium, iron and calcium. 9. Hurricane Ivan of 2004 produced 118 tornadoes in three days. 10. Earthquakes of magnitude 7 or larger on the Richter scale have been produced by major explosive eruptions. 11. The 1982–1983 El Niño–Southern Oscillation event produced the worst drought in Australia to that point in time. 12. The strong or moderate ENSO events between 1851 and 1974 produced major droughts in eastern Australia. 13. The Lisbon earthquake of 1755 produced a 3–4 m wave that was felt on all sides of the Atlantic. 14. The historical 1821 hurricane produced winds reaching 180–210 km hr-1. 15. The tropical storm produced loss of life, and destruction of property on a scale never before seen in the history of the United States.		
3	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent, water agent, or atmospheric condition. The PATIENT is usually a natural disaster, death, or loss of life/property.			
	Semantic role	RESULT (IN, FROM)	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric agent, water agent, atmospheric condition	natural disaster, atmospheric agent, water agent, atmospheric condition damage, loss of property/life, death
	Linguistic realizations		earthquake, natural hazards, disaster, drought, tremor	fatalities, death, losses, landslide, floods
	Phrase type		NP	NP
	Usage examples	1. Some meteorological factors have resulted in landslides and floods that have caused loss of lives. 2. The earthquake and related hazards resulted in hundreds of thousands of fatalities. 3. Natural hazards have resulted in human losses. 4. The Kobe earthquake resulted in larger financial and economic losses than any disaster ever. 5. The disaster resulted in the deaths of 20,000 people.		

		6. In Ethiopia, drought resulted in starvation followed by massive international appeals 7. This tremor resulted in 69,227 fatalities. 8. Since 684 AD, in the Japan region, 73 tsunami have resulted in over 200 000 deaths. 9. Unfortunately, 25 000 deaths also resulted from this natural disaster along the coasts of Chile and Peru. 10. The tsunami resulted from sudden land displacement of 2–4 m along the fault. 11. The deaths resulted from the fires that immediately broke out in the cities. 12. In recent years, the vicissitudes wrought by harsh weather patterns and natural disasters have resulted in hurricanes and earthquakes and had a catastrophic impact on people's lives. 13. During this year there where fifteen storms identified that could have resulted in hurricanes. 14. NOAA'S National Hurricane Center estimates that $85 billion of total damages resulted from Hurricanes Katrina and Rita alone. 15. In 1992 in Florida, Hurricane Andrew resulted in death, damage and destruction to housing and business. 16. The passage of the hurricane resulted in damage to houses, bridges, and electrical and water supply.

4 Lexical domain: EXISTENCE
Frame: to_cause_to_exist/happen
Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen.
Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent, water agent, or atmospheric condition. The PATIENT is usually a natural disaster, water agent, atmospheric agent, or death entity.

Semantic role	TRIGGER	Natural force	Patient
Macrorole		Actor	Undergoer
Labels		natural disaster, atmospheric agent, water agent, atmospheric condition	natural disaster, water agent, atmospheric agent, atmospheric condition, atmospheric agent, death
Linguistic realizations		earthquake, volcanoes, landslides, rainfall, natural hazard	earthquake, flash flooding, landslide, tsunami, avalanche, disaster, death,
Phrase type		NP	NP

Usage examples

1. California' earthquake triggered 14 other earthquakes up to 1200 km away.
2. Flash flooding triggered by record rainfall in Istanbul killed 31 people.
3. Turkey landslide triggered by heavy rainfall brought down a house in northern Turkey.
4. In Bergen (Norway), heavy rainfall triggered landslides that swept away seven people working on a house.
5. The earthquake that triggered the tsunami collapsed the city.
6. 15 millions more people have been affected by disasters triggered by natural hazards.
7. The HDI again turns out to be the best predictor of deaths triggered by extreme natural events.
8. The debris avalanche was triggered by an earthquake registering 7 on the Richter's scale.
9. Such landslides might be triggered by earthquakes or by rainfall.
10. The climactic eruption of Mount St Helens during May 11 was

		triggered by a spectacular landslide. 11. There is also evidence that tropical cyclones have triggered earthquakes in other places along the western margin of the Pacific plate. 12. It occurred when an earthquake triggered a large avalanche that overran the camp. 13. Coastal areas are also at risk of tsunami, the waves triggered by earthquakes, volcanoes or undersea landslides that can cross oceans at 800 km/h and rise to enormous heights. 14. In 1916 remnants of a hurricane triggered severe flooding along the French Broad River in July of that year.
5	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster or movement of material. The PATIENT is usually a natural disaster, atmospheric agent, water agent, energy, material entity.	
	Semantic role	GENERATE: Natural force / Patient
	Macrorole	Actor / Undergoer
	Labels	natural disaster, material entity / natural disaster, atmospheric agent, water agent, energy, material entity
	Linguistic realizations	hurricanes, cyclone, volcano, earthquake, eruption, fire, flow / tsunami, landslide, tornado, strong winds, great waves
	Phrase type	NP / NP
	Usage examples	1. Either method is an enormous amount energy being generated by hurricanes. 2. A submarine explosion could have generated tsunami 15 m high throughout the Strait. 3. These flows probably generated the tsunami that preceded the final explosion. 4. Cyclone Namu, which passed over the Solomon Islands in 1986, generated hundreds of landslides with an average volume of mater. 5. The volcano generated massive pyroclastic flows. 6. The tsunami was generated by an earthquake with a surface magnitude of 8.3 in the Richter scale. 7. Tsunamis are generated mainly by earthquakes or volcanoes. 8. The eruption of the Eyjafjallajökull volcano in Iceland generated a large ash cloud, thus causing enormous problems for the traffic. 9. The Santorini eruption of 1470 BC generated a tsunami that must have destroyed all coastal towns. 10. The undersea earthquake generated a great wave. 11. The most intense and hazardous fire tornadoes are generated by forest fires. 12. An eruption in 1966 generated lahars that took hundreds of lives. 13. The fires were exacerbated by the development of strong winds generated by the cyclone in the lee of surrounding mountains. 14. 1098 offshore earthquakes have generated only 19 tsunamis. 15. It is interesting to measure the energy contained in any tsunami wave generated by an eruption. 16. Fortunately, due to the compact size of Hurricane Charley, only limited significant storm surge was generated by the hurricane.
6	Lexical domain: EXISTENCE	

	Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster or atmospheric agent. The PATIENT is usually a natural disaster.			
	Semantic role	SPAWN	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric agent	natural disaster
	Linguistic realizations		earthquake, tropical cyclone, hurricane, winds, thunderstorms	tsunami, tornado
	Phrase type		NP	NP
	Usage examples	1. Earthquakes spawn tsunami in Samoa, Rock Indonesia. 2. Why the Indonesian earthquake didn't spawn a tsunami. 3. How hurricanes spawn tornadoes. 4. Tropical cyclones spawn tornadoes when certain instability and vertical shear criteria are met. 5. The data sets used report damage values for two Hawaiian tsunamis spawned by Alaskan earthquakes April 1, 1946, Unimak Island. 6. A hurricane's high winds are also destructive and may spawn tornadoes. 7. Relatively weak hurricanes like Danny (1985) have spawned significant super cell tornadoes well inland. 8. It is interesting to highlight that not every thunderstorm spawns a tornado.		
8	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] or [AGENT] causes a [PATIENT] to exist. Note: The NATURAL FORCE is usually a natural disaster, or human being. The PATIENT is usually a natural disaster.			
	Semantic role	$START_1$	Natural force/Agent	Patient
	Macrorole		Actor	Undergoer
	Labels		human being, natural disaster	natural disaster
	Linguistic realizations		people, children, earthquake, hurricane	fire, landslide
	Phrase type		NP	NP
	Usage examples	1. The majority of forest fires are started intentionally. 2. On 18 May, an earthquake of magnitude 5.1 on the Richter scale started landslides in the Bulge Region. 3. Four out of five wildfires are started by people. 4. Between 1940 and 1975, over 200 000 fires were started by lightning in the western United States. 5. The hurricane of 1938 that devastated New England, in the United States, started fires in New London. 6. Nearly a third of all fires are started by children, most of whom are boys playing with matches. 7. The largest such event, in June 1940, started 1488 fires.		
9	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes a [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster, or human being. The PATIENT is			

	usually a natural disaster or construction.			
	Semantic role	FORM$_2$	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster	natural disaster, construction
	Linguistic realizations		landslide, earthquake, floods	tsunami, landslide, dams
	Phrase type		NP	NP
	Usage examples	1. These dams are formed by landslides blocking a valley, or glaciers holding back. 2. The latter floods over time have formed an outwash plain 1000 km2 in area. 3. The landslide was formed by the earthquake. 4. The earthquake formed a tsunami which hit the coast line of nearby Okushiri Island and the central west coast of Hokkaido within minutes.		
9	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually an atmospheric agent, and the PATIENT a natural fire disaster.			
	Semantic role	SPREAD$_2$	Agent	Patient
	Macrorole		Actor	Undergoer
	Labels		atmospheric agent	natural disaster
	Linguistic realizations		wind	Fire
	Phrase type		NP	NP
	Usage examples	1. Cyclonegenerated winds spread fires that destroyed over 3300 buildings in Hokkaido. 2. Most severe fires have been associated with strong winds that spread the fire during hot, dry periods.		

Table 88. Analysis of the template *to cause to exist/happen* (English)

1	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] or [AGENT] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent, human being or atmospheric condition. The PATIENT is usually a natural disaster, damage or death entity.			
	Semantic role	CAUSAR	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric agent, human being	damage, death, natural disaster
	Linguistic realizations		earthquake, floods, storm, rain, landslide, forest fires, fires, humans	damage, fatalities, death, tsunami, disasters, floods, flooding, landslides
	Phrase type		NP	NP
	Usage examples	1. Este fue el mayor daño causado por un ciclón tropical. 2. Las sacudidas en el suelo son las causas directas de los daños más graves causados por un terremoto. 3. Las inundaciones fueron causadas por el Río Pánuco y devastaron vastas hectáreas de tierra. 4. Los riesgos naturales son responsables de causar muchas muertes y daños en todo el mundo cada año. 5. ¿Qué ciclones tropicales han causado más muertes y más daños? 6. La principal causa de las crecidas fluviales son las lluvias torrenciales causadas por el monzón, asociadas muchas veces con tifones. 7. Las inundaciones fueron causadas por lluvias abundantes. 8. En latitudes medias y altas, las inundaciones están causadas por temporales de lluvia de origen frontal. 9. Estos terremotos fueron causados por desprendimientos o deslizamientos de grandes masas. 10. Una razón importante de que los terremotos causen un gran daño y pérdidas de vida es que a menudo golpean sin previo aviso. 11. Las lluvias torrenciales que azotan a la Comunidad Valenciana causan graves pérdidas económicas. 12. Aunque algunas presas se derrumben y causen inundaciones peligrosas aguas abajo, si una presa permanece estable puede proporcionar un valioso depósito de agua. 13. La sequía causó elevadas pérdidas económicas y favoreció el desarrollo de enfermedades. 14. Este es un ejemplo del daño causado por un tornado de este tipo. 15. Las inundaciones causadas por lluvias intensas o torrenciales son las más frecuentes. 16. Los tsunamis también pueden estar causados por deslizamientos de tierra subterráneos o erupciones. 17. Donde ocurren los deslizamientos pueden causar importantes daños y pérdida de vidas. 18. Los ciclones tropicales en el mar abierto causan grandes olas, lluvias torrenciales y fuertes vientos. 19. Las lluvias causan importantes inundaciones y deslizamientos de tierra en Brasil. 20. Explosiones, lluvias de ceniza y tsunamis causados por la erupción provocaron la muerte de unas 90 000 personas en Indonesia.		
2	Lexical domain: EXISTENCE			

	Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster or movement of material. The PATIENT is usually a natural disaster, atmospheric agent, water agent, energy, material entity, damage or loss of life/property.			
	Semantic role	GENERAR	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, material entity	natural disaster, atmospheric agent, water agent, energy, material entity, damage, loss of life/property
	Linguistic realizations		huracán, inundaciones, tsunami, terremoto, sequía, tornado, incendio, tormenta, volcán	graves inundaciones, olas de dos metros, tsunami, pérdidas, explosiones, destrucción total, deformaciones, destrozos, pánico
	Phrase type		NP	NP
	Usage examples	1. El huracán generó graves inundaciones que provocaron el cierre de la red del metro de Nueva York. 2. El huracán generó pérdidas por 150.000 millones de dólares. 3. El huracán generó la destrucción casi total en zonas costeras. 4. Tsunami generó olas de dos metros que afectaron isla japonesa. 5. El terremoto generó un tsunami que impactó las costas de las ya devastadas. 6. Este terremoto generó deformaciones en la superficie de más de tres mes que aumentó la probabilidad de ocurrencia de futuros eventos hacia el noreste. 7. La sequía generó pérdidas de 650.000 millones. 8. La sequía generó una pérdida de 35.000 bovinos. 9. El tornado generó destrozos en el este. 10. El tornado generó pánico en todo el estado. 11. Un tornado generó múltiples destrozos en Monte Nievas. 12. El incendio generó pérdidas por más de 400 millones de dólares. 13. Un incendio generó muy severos daños en la Catedral de Holy Day. 14. Las tormentas e inundaciones generaron cuantiosos daños en 2010 en España. 15. El volcán generó pérdidas de 8 millones de dólares. 16. El volcán generó 16 explosiones el 21 de agosto.		
3	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster or atmospheric condition. The PATIENT is usually a natural disaster, atmospheric agent, water agent, damage, death, material entity.			
	Semantic role	PRODUCIR$_1$	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric condition	natural disaster, atmospheric agent, water agent, damage, death, material entity
	Linguistic realizations		volcán, tsunami, huracán, tifón, avalancha, sequía	olas, enorme inundación, vientos de más de 140

				km/hora, marejada, muertos, daños materiales, pérdidas, víctimas
	Phrase type		NP	NP
	Usage examples	1. El volcán produjo daños materiales en Canarias y destruyó la playa. 2. El tsunami produjo solamente olas pequeñas en México. 3. El tsunami produjo 23 000 muertos. 4. El tsunami produjo una enorme inundación. 5. El huracán produjo olas altas en Wilmington Carolina del Norte y marejadas anormalmente altasa través de Nueva Inglaterra. 6. El tifón produjo vientos de más de 140 km/hora. 7. La avalancha de nieve produjo al menos 25 muertos. 8. La sequía produjo pérdida de un 50% en soja y maíz. 9. El tremendo descarrilamiento fue producido por el tornado. 10. Ecologistas alertan del "desastre ecológico" producido por el incendio. 11. Declaración de daños en explotaciones agrícolas y ganaderas poducidas por las inundaciones del 28/09/2012. 12. Hasta el momento, el número de víctimas producidas por las inundaciones asciende a nueve.		
4	Lexical domain: EXISTENCE Frame: to_cause_to_exist/happen Frame definition: [NATURAL FORCE] or [AGENT] causes [PATIENT] to exist/happen. Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent. The PATIENT is usually a natural disaster, water agent, atmospheric agent, or damage, loss of property/life, or death entity.			
	Semantic role	PROVOCAR	Natural force/Agent	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, atmospheric agent, human being	natural disaster, water agent, atmospheric agent, death
	Linguistic realizations		inundaciones, incendio, huracán, avalancha, terremoto, deslizamiento de tierra, tsunami, erupción de un volcán, tornado	pérdidas, olas de hasta 4 metros de altura, lluvias e inundaciones, muertes, daños, destrozos, tsunami, heridos
	Phrase type		NP	NP
	Usage examples	1. Pakistán cifra en 33.500 millones las pérdidas provocadas por las inundaciones. 2. Las instituciones canarias han hecho un primer balance de las pérdidas provocadas por el incendio de La Gomera. 3. En la costa sur de Cuba, el huracán provocó olas de hasta 4 m de altura. 4. “Ernesto”, convertido en huracán, provocó este miércoles fuertes lluvias e inundaciones al atravesar la península de Yucatán. 5. Hasta el 11 de octubre de 2005, el huracán provocó al menos 1620 muertes. 6. La avalancha provocó daños en Suratá, Santander. 7. Una avalancha provocó la muerte de trece montañistas. 8. Un fuerte terremoto provocó un tsunami. 9. Algunas fuentes afirman que el terremoto provocó unas 65.000		

<table>
<tr><td></td><td></td><td colspan="3">muertes, y unos 375.000 heridos.
10. Poderoso tsunami provoca más de 120 muertes.
11. Un tornado provoca destrozos en Córdoba.
12. Un deslizamiento de tierra provocó la muerte de 11 personas.
13. En Indonesia, la erupción de un volcán provocó 16 muertos y 13 heridos graves.
14. El incendio fue provocado por un pirómano.</td></tr>
<tr><td>5</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_cause_to_exist/happen
Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen.
Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent.
The PATIENT is usually a natural disaster, water agent, atmospheric agent, or dame, death or loss of life/property entity.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="4">OCASIO-NAR</td><td>Natural force</td><td>Patient</td></tr>
<tr><td></td><td>Macrorole</td><td>Actor</td><td>Undergoer</td></tr>
<tr><td></td><td>Labels</td><td>natural disaster, atmospheric agent</td><td>natural disaster, water agent, atmospheric agent, death, damage, loss of property/life</td></tr>
<tr><td></td><td>Linguistic realizations</td><td>huracán, terremoto, tsunami, sequía, incendio, tornado, erupción, movimiento de masa de tierra</td><td>daños, destrozos, muertes, pérdidas, estragos, tsunami</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>NP</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. El huracán ocasionó graves daños en Santiago de Cuba.
2. Este huracán ocasionó alrededor de 2000 muertos.
3. Un fuerte terremoto ocasiono grandes daños en Haití.
4. El terremoto ocasionó grandes destrozos en Italia.
5. Un terremoto de 9 grados Richter, seguido de un gigantesco tsunami, ocasionó 18.000 muertos (15.873 muertos y 2.768 desaparecidos).
6. El Estado estudio cómo compensar los más de 1.500 millones de euros en pérdidas ocasionados por la sequía de estos últimos meses.
7. Esto son los estragos ocasionados por el incendio en La Gomera.
8. Los daños ocasionados por el incendio han sido cifrados por los responsables municipales en “miles de euros”.
9. A día de hoy, sin embargo, todavía no pueden hacer un balance oficial de los daños ocasionados por el tornado.
10. La erupción de un volcán ocasiona seis muertes en Colombia.
11. El movimiento de masa de tierra ocasionó importantes pérdidas materiales y humanas.
12. El terremoto ocasionó un tsunami, con olas de un promedio de 10 metros de altura que sumergió más de 443 kilómetros cuadrados de áreas costeras.</td></tr>
<tr><td>6</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_cause_to_exist/happen
Frame definition: [NATURAL FORCE] causes [PATIENT] to exist/happen.
Note: The NATURAL FORCE is usually a natural disaster, atmospheric agent.
The PATIENT is usually a natural disaster, water agent, atmospheric agent, or damage, loss of property/life, or death entity.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="3">ORIGINAR/ ORIGINAR-SE</td><td>Natural force</td><td>Patient</td></tr>
<tr><td></td><td>Macrorole</td><td>Actor</td><td>Undergoer</td></tr>
<tr><td></td><td>Labels</td><td>natural disaster, atmospheric agent</td><td>natural disaster, water agent, atmospheric agent,</td></tr>
</table>

				damage, death
	Linguistic realizations		inundaciones, catástrofes de causa atmosférica, maremoto, terremoto, temporales, olas de frío y de calor, deslizamientos, huracán, sequía, depresión tropical, huracán, erupción	daño, maremoto, perturbaciones atlánticas muy desarrolladas, terremotos, volcanes
	Phrase type		NP	NP
	Usage examples	1. El daño originado por las inundaciones y otras catástrofes de causa atmosférica es cada vez mayor en toda Europa, a consecuencia probablemente de intervenciones humanas. 2. Para que un terremoto origine un maremoto, el fondo marino debe ser movido abruptamente. 3. En Europa, los temporales de viento fuerte son originados por perturbaciones atlánticas muy desarrolladas. 4. Las olas de frío y de calor están originadas por los desplazamientos de las masas de aire ártico y polar continental, y tropical continental, respectivamente. 5. Los deslizamientos son peligros naturales de causa múltiple, esto es, se originan por la acción combinada de fenómenos atmosféricos diversos. 6. Como consecuencia del terremoto se originaron maremotos que arrasaron las costas del Japón. 7. La mayoría de los maremotos son originados por terremotos de gran magnitud bajo la superficie acuática. 8. Las sequías no están originadas por la misma situación sinóptica en todas las regiones. 9. La "depresión tropical" que azotó Canarias entre el 18 y 21 de septiembre de 1951, originó precipitaciones muy abundantes en diferentes áreas del archipiélago. 10. El último tsunami pequeño originó olas de dos metros que se sintieron en el puerto de Mahón, en la isla de Menoría, causando daños a los barcos que allí se encontraban. 11. Este es el tipo de erupción con la que se originan muchos volcanes, de ahí su denominación genérica. 12. El conocido "terremoto de Lisboa" afectó a todo el suroeste peninsular y llegó a originar un tsunami intenso, de efectos desastrosos en las costa 13. La evaluación de la probabilidad de daños originados por terremotos ha de tener en cuenta la peligrosidad sísmica y la vulnerabilidad de los elementos del territorio y del terreno. 14. En realidad Saffir y Simpson diseñaron una escala de daños potenciales originados por los huracanes.		

Table 89. Analysis of the frame *to cause to exist/happen* (Spanish)

As encoded in the template in Table 90, the verbs in *to cause to exist/happen* generally have two arguments: NATURAL FORCE and PATIENT. However, certain verbs (those with an asterisk), such as 'cause', 'start1', 'causar', 'provocar', can extend the

NATURAL FORCE to a volitional AGENT in some contexts (e.g. 'Fires in Europe are caused especially by humans'):

Frame: **to cause to exist/happen**				
Semantic role	Natural force/Agent*	*cause** *start1** *form2* *generate* *produce* *spawn* *result* *(from/in)* *trigger* *spread2*	*causar** *generar* *producir1* *provocar** *ocasionar* *originar*	Patient
Macrorole	Actor			Undergoer
Conceptual class	natural disaster, atmospheric condition, water agent, atmospheric agent, human being			natural disaster, atmospheric agent, water agent, atmospheric condition, damage, loss of life/property, death
Phrase type	NP			NP

Table 90. Template of the frame *to cause to exist/happen*

To exist in time

In the frame *to exist in time,* in the same way as *to begin to exist,* GEOGRAPHICAL LOCATION, TIME and MANNER were potential semantic argument roles activated by the predicates in the frame. As shown in Tables 91 and 92, English verbs that activate this frame are 'happen', 'take place', 'occur', 'recur', whereas the Spanish verbs are 'producirse', 'ocurrir', 'tener lugar', 'suceder', 'sobrevenir', and 'pasar'. In many cases, differences are due to register changes (e.g. 'sobrevenir' and 'pasar'). This information is included in the *note* section for each verb sense

1	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	HAPPEN	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		avalanche, landslide, flood, tornado, hurricane, eruption, flooding	Col des Mosses, island, 149 km from Oaxaca	weekend, 1978, daytime, June 1 and November 30, 2004	sudden-ly
	Phrase type		NP	PP (in, at, from)	PP (during, in, at, between)	AVP

	Usage examples	1. One of the more recent avalanches happened in 1978 at Col des Mosses, Switzerland, when the avalanche overwhelmed a ski-lift and killed 60 people. 2. The landslides and floods that happened in the island together with the resulting damages, were the worst that people have seen and experienced. 3. The flooding happened at daytime and during the weekend, when most families were at home. 4. Multiple tornadoes happened; but there was no survey done to precisely separate their paths 5. Tropical cyclones happen in the south east of the United States, between June 1 and November 30. 6. The earthquake happened 149 km from Oaxaca. 7. How the 2004 tsunami happened? 8. The drought happened suddenly - what is called a "flash drought". 9. The previous eruption of the Puyehue volcano happened in 1960. 10. Where did Katrina hurricane happened?

2	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	TAKE PLACE	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		fire, landslide, earthquake, flood, hurricane	this zone, Europe, base camp	every year, recent decades, three years after	suddenly
	Phrase type		NP	PP (in, at)	PP (over, after)	AVP
	Usage examples	1. An average of 70 000 fires take place every year burning more than half a million hectares. 2. The main landslide took place west of the old center of Corniglio. 3. From the millions of earthquakes that take place every year, few are actually deadly. 4. It was in this zone that some of the deadliest floods took place. 5. The floods that took place in Europe over the recent decades have shown an increase in economic losses. 6. The hurricane took place three years after Hurricane Katrina. 7. It is thought the avalanche took place suddenly at a base camp around 4,000 ft from Manaslu's 26,760ft high summit. 8. Hurricane Mitch took place in 1998 in the Atlantic Ocean.				
3	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	OCCUR	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____

	Linguistic realizations		landslide, tsunami, eruption, avalanche, earthquake	Europe, Chicxulub, mountain areas, Austria, Peru, island	period 2003-2009, past years, sixty-five million years ago	simulta-neously
	Phrase type		NP	PP (in, at, around, across)	PP (during, in)	AVP
	Usage examples	1. The major landslide phenomena that occurred in Europe during the period 2003–2009 were compiled from an analysis of global disasters databases. 2. One of the largest impact-induced tsunami occurred at Chicxulub, Mexico, sixty-five million years ago. 3. In the South-eastern part of Poland floods occurred in the mountain and sub mountain areas. 4. These eruptions occurred simultaneously. 5. The avalanches which occurred in Austria in the past years have not been extraordinary. 6. Other noteworthy earthquakes occurred in 1769 south-east of Los Angeles; on 10 June 1836. 7. The most fatal floods occurred in Romania with 85 people killed in 2005. 8. These fires that occurred mainly in the Mediterranean Region, led to land degradation. 9. Many of these earthquakes occurred in the southern part of the Long Valley Caldera. 10. Similar floods had occurred in 1798, 1912, 1914, 1938, 1944, 1948, and 1951. 11. Hurricane Mitch occurred in 1998, early in the rainy season. 12. By the beginning of August, a dozen Vesuvian-type eruptions had occurred across the island.				
4	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists again in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	RECUR	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		earthquake, eruption,	same location, same place	today, date	same way
	Phrase type		NP	PP (at, in)	PP (to)	PP (in)
	Usage examples	1. Since then no strong earthquake recurred to date at the same locality, though two earthquakes, one in 1597 and the other in 1969, occurred nearby. 2. If the 79 AD eruption recurred today, over 1,500,000 people would be affected. 3. One of the aspects of such systems is that earthquakes recur at the same location. 4. The fact that avalanches tend to recur in the same place and in the same way has aided this mapping.				

Table 91. Analysis of the frame *to exist in time* (English)

1	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	PRODUCIRSE	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		avalancha, sequía, terremoto, ciclón, tsunami, tornado, inundación, incendio	sur de Louisiana, zonas sísmicas, Estados Unidos, montañas Rocosas y Apalaches	los meses de agosto y septiembre, las 3.00 y las 7.00	de repente
	Phrase type		NP	PP (en, entre)	NP, PP (en, por, entre)	AVP
	Usage examples	1. La segunda inundación se produjo en el domo de sal de Jefferson Island, en el sur de Louisiana. 2. La mayoría de los terremotos se producen en zonas sísmicas o fallas geológicas. 3. Estos ciclones tropicales se producen normalmente en los meses de agosto y septiembre. 4. En promedio, en los Estados Unidos se producen unos 1.200 tornados por año. 5. ¿Por qué los ciclones tropicales se producen primordialmente en el verano y otoño? 6. La mayoría de los tornados en Estados Unidos se producen en los estados de las llanuras, entre las montañas Rocosas y Apalaches. 7. A nivel mundial, la mayoría de los tornados se producen durante la tarde, entre las 3:00 pm y las 7:00 pm del tiempo local, siendo el punto más alto a las 5:00 pm. 8. La avalancha se produjo al norte de Nepal. 9. La sequía se produjo de repente - lo que se denomina una "sequía flash", ya que se ha producido en cuestión de meses, en lugar de temporadas o años. 10. El tsunami se produjo exactamente un año después del terremoto de 2003 que devastó la ciudad iraní meridional de Bam. 11. El primer incendio se produjo el pasado miércoles 5 de diciembre, y desde entonces se han registrado 7 incendios más.				
2	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	OCURRIR	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		avalanche, sequía, terremoto, ciclón,	sur de Louisiana, región de Murcia, zonas	los meses de agosto y septiembre, 2 de	de repente

			tsunami, tornado, inundación, incendio	sísmicas, Estados Unidos, montañas Rocosas y Apalaches	febrero de 1999, las 3.00 y las 7.00	
	Phrase type		NP	PP (en, entre)	NP, PP (en, por, entre)	AVP
	Usage examples	1. La segunda inundación ocurrió en el domo de sal de Jefferson Island, en el sur de Louisiana. 2. Por último, destacan los terremotos ocurridos en la región de Murcia el 2 de febrero de 1999. 3. En fechas más recientes, la inundación de 1994 ocurrida en este río dejó sin vivienda a 200 millones de personas. 4. La mayoría de los terremotos ocurren en zonas sísmicas o fallas geológicas. 5. Estos ciclones tropicales ocurren normalmente en los meses de agosto y septiembre. 6. En promedio, en los Estados Unidos ocurren unos 1.200 tornados por año. 7. ¿Por qué los ciclones tropicales ocurren primordialmente en el verano y otoño? 8. La mayoría de los tornados en Estados Unidos ocurre en los estados de las llanuras, entre las montañas Rocosas y Apalaches. 9. A nivel mundial, la mayoría de los tornados ocurren durante la tarde, entre las 3:00 pm y las 7:00 pm del tiempo local, siendo el punto más alto a las 5:00 pm. 10. La avalancha ocurrió al norte de Nepal. 11. La sequía ocurrió de repente - lo que se denomina una "sequía flash", ya que se ha producido en cuestión de meses, en lugar de temporadas o años. 12. El tsunami ocurrió exactamente un año después del terremoto de 2003 que devastó la ciudad iraní meridional de Bam. 13. El primer incendio ocurrió el pasado miércoles 5 de diciembre, y desde entonces se han registrado 7 incendios más.				
3	Lexical domain: EXISTENCE Frame: to_exist_in_time Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].					
	Semantic role	TENER LUGAR	Theme	Geographical location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster	_____	_____	_____
	Linguistic realizations		avalancha, sequía, terremoto, ciclón, tsunami, tornado, inundación, incendio	sur de Louisiana, zonas sísmicas, Estados Unidos, montañas Rocosas y Apalaches	los meses de agosto y septiembre, las 3.00 y las 7.00	de repente
	Phrase type		NP	PP (en, entre)	NP, PP (en, por, entre)	AVP

	Usage examples	1. La segunda inundación tuvo lugar en el domo de sal de Jefferson Island, en el sur de Louisiana. 2. La mayoría de los terremotos tienen lugar en zonas sísmicas o fallas geológicas. 3. Estos ciclones tropicales tienen lugar normalmente en los meses de agosto y septiembre. 4. En promedio, en los Estados Unidos tienen lugar unos 1.200 tornados por año. 5. ¿Por qué los ciclones tropicales tienen lugar primordialmente en el verano y otoño? 6. La mayoría de los tornados en Estados Unidos tienen lugar en los estados de las llanuras, entre las montañas Rocosas y Apalaches. 7. A nivel mundial, la mayoría de los tornados tienen lugar durante la tarde, entre las 3:00 pm y las 7:00 pm del tiempo local, siendo el punto más alto a las 5:00 pm. 8. La avalancha tuvo lugar al norte de Nepal. 9. La sequía tuvo lugar de repente - lo que se denomina una "sequía flash", ya que se ha producido en cuestión de meses, en lugar de temporadas o años. 10. El tsunami tuvo lugar exactamente un año después del terremoto de 2003 que devastó la ciudad iraní meridional de Bam. 11. El primer incendio tuvo lugar el pasado miércoles 5 de diciembre, y desde entonces se han registrado 7 incendios más.

4 Lexical domain: EXISTENCE
Frame: to_exist_in_time
Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].

Semantic role	SUCEDER	Theme	Geographical location	Time	Manner
Macrorole		Undergoer	_____	_____	_____
Labels		natural disaster	_____	_____	_____
Linguistic realizations		avalancha, sequía, terremoto, ciclón, tsunami, tornado, inundación, incendio	sur de Louisiana, zonas sísmicas, Estados Unidos, montañas Rocosas y Apalaches	los meses de agosto y septiembre, las 3.00 y las 7.00	de repente
Phrase type		NP	PP (en, entre)	NP, PP (en, por, entre)	AVP

Usage examples	1. La segunda inundación sucedió en el domo de sal de Jefferson Island, en el sur de Louisiana. 2. La mayoría de los terremotos suceden en zonas sísmicas o fallas geológicas. 3. Estos ciclones tropicales suceden normalmente en los meses de agosto y septiembre. 4. En promedio, en los Estados Unidos suceden unos 1.200 tornados por año. 5. ¿Por qué los ciclones tropicales suceden primordialmente en el verano y otoño? 6. La mayoría de los tornados en Estados Unidos suceden en los estados de las llanuras, entre las montañas Rocosas y Apalaches. 7. A nivel mundial, la mayoría de los tornados suceden durante la

		tarde, entre las 3:00 pm y las 7:00 pm del tiempo local, siendo el punto más alto a las 5:00 pm. 8. La avalancha sucedió al norte de Nepal. 9. La sequía sucedió de repente - lo que se denomina una "sequía flash", ya que se ha producido en cuestión de meses, en lugar de temporadas o años. 10. El tsunami sucedió exactamente un año después del terremoto de 2003 que devastó la ciudad iraní meridional de Bam. 11. El primer incendio sucedió el pasado miércoles 5 de diciembre, y desde entonces se han registrado 7 incendios más.

5 Lexical domain: EXISTENCE
Frame: to_exist_in_time
Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].
Note: Formal register

Semantic role	SOBREVENIR	Theme	Geographical location	Time	Manner
Macrorole		Undergoer	_____	_____	_____
Labels		natural disaster	_____	_____	_____
Linguistic realizations		sequía, terremoto, tormenta, huracán, incendio, tsunami, inundaciones	refinería, estado sureño de Oaxaca	noche del 28 de julio, dos días más tarde, día 18, año pasado	de repente, con fuertes vientos
Phrase type		NP	PP (en)	NP	AVP

Usage examples

1. Sobrevino una sequía, que se prolongó durante años.
2. Entonces, de repente sobrevino un fuerte terremoto, de manera que los cimientos de la cárcel fueron sacudidos.
3. La noche del 28 de julio sobrevino una tormenta de gran intensidad, con fuertes vientos arremolinados y lluvia.
4. Dos días más tarde sobrevino un huracán que arrasó Florida.
5. El día 18 sobrevino un huracán tan violento, que en menos de 3 horas echó á pique todas las Embarcaciones que había en el río.
6. Tras el terremoto, sucedido durante la madrugada, sobrevino un tsunami que afectó las costas chilenas.
7. El jueves 18 de octubre, sobrevino un incendio en la refinería Dovalí Jaime, de Salina Cruz, en el sureño estado de Oaxaca.
8. El año pasado, además de prestar socorro inmediato cuando sobrevinieron inundaciones, la Cruz Roja China reconstruyó 1.200 casas.

6 Lexical domain: EXISTENCE
Frame: to_exist_in_time
Frame definition: [NATURAL DISASTER] exists in a particular [LOCATION], at a given [TIME] or in a specified [MANNER].

Semantic role	PASAR	Theme	Geographical location	Time	Manner
Macrorole		Undergoer	_____	_____	_____
Labels		natural disaster	_____	_____	_____
Linguistic realizations		avalanche, sequía, terremoto, ciclón, tsunami,	sur de Louisiana, zonas sísmicas, Estados	los meses de agosto y septiembre, las 3.00 y las 7.00	de repente

			tornado, inundación, incendio	Unidos, montañas Rocosas y Apalaches		
	Phrase type		NP	PP (en, entre)	NP, PP (en, por, entre)	AVP
	Usage examples	1. La segunda inundación pasó en el domo de sal de Jefferson Island, en el sur de Louisiana. 2. La mayoría de los terremotos pasan en zonas sísmicas o fallas geológicas. 3. Estos ciclones tropicales pasan normalmente en los meses de agosto y septiembre. 4. En promedio, en los Estados Unidos pasan unos 1.200 tornados por año. 5. ¿Por qué los ciclones tropicales pasan primordialmente en el verano y otoño? 6. La mayoría de los tornados en Estados Unidos pasan en los estados de las llanuras, entre las montañas Rocosas y Apalaches. 7. A nivel mundial, la mayoría de los tornados pasan durante la tarde, entre las 3:00 pm y las 7:00 pm del tiempo local, siendo el punto más alto a las 5:00 pm. 8. La avalancha pasó al norte de Nepal. 9. La sequía pasó de repente - lo que se denomina una "sequía flash", ya que se ha producido en cuestión de meses, en lugar de temporadas o años. 10. El tsunami pasó exactamente un año después del terremoto de 2003 que devastó la ciudad iraní meridional de Bam. 11. El primer incendio pasó el pasado miércoles 5 de diciembre.				

Table 92. Analysis of the frame *to exist in time* (Spanish)

As shown in Table 93, verbs contained in the frame *to exist in time* activate the same template as the frame *to begin to exist* since their first argument is a NATURAL DISASTER, which fulfils the role of THEME and macrorole of UNDERGOER. Predicates can optionally include arguments that designate GEOGRAPHICAL LOCATION, TIME, or MANNER.

Frame: to exist in time					
Semantic role	Theme	*happen* *take place* *occur* *recur* *producirse* *ocurrir* *tener lugar* *suceder* *sobrevenir* *pasar*	Geographical location	Time	Manner
Macrorole	Undergoer		_____	_____	_____
Conceptual class	natural disaster		_____	_____	_____
Phrase type	NP		PP (around, at, in, off, on, over, etc.) (English) PP (a, en) (Spanish)	PP (after, ago, at, between, by, during, every, for, from, in, just, on, over, to) (English) PP (en, por, entre, desde, hasta, durante) (Spanish)	AVP

Table 93. Template for the frame *to exist in time*

To continue to exist (of natural disaster)

In *to continue to exist* (when referring to a natural disaster), the semantic role of TIME becomes compulsory since the frame describes those verbs linked to a NATURAL DISASTER that exists over a period of TIME. It is instantiated by verbs, such as 'last' or 'persist', in English. The verb 'extend2' has also been included in this template. However, 'extend2' is more likely to be applied to the consequences or effects of NATURAL DISASTERS than to disaster themselves. The same can be said of the verb 'extenderse2', for Spanish (see Tables 94 and 95):

<table>
<tr><td>1</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_continue_to_exist_(of natural disaster)
Frame definition: [NATURAL DISASTER] continues to exist in [TIME].
Note: The NATURAL DISASTER can also include atmospheric agent, and atmospheric condition.</td></tr>
<tr><td></td><td>Semantic role</td><td>LAST</td><td>Theme</td><td>Time</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, atmospheric agent, atmospheric condition</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>cyclone season, storm, landslide, eruption, tsunami</td><td>several days, December to May, five days, 45 minutes, several hours, 500 and 600 seconds, 24 hours, a minute</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>PP (from, to, just, for, between)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. Large fires that lasted several days occurred recently in Portugal.
2. The cyclone season around Australia usually lasts from 1 December to 1 May.
3. This storm lasted about five days.
4. Although the storm lasted just 45 minutes, flood waters up to 2 m high surged through.
5. It caused a catastrophic landslide and a major explosive eruption that lasted for several hours.
6. This earthquake lasted between 500 and 600 seconds.
7. This 1667 hurricane lasted about 24 hours and was accompanied by very violent winds and tides.
8. It seems the tsunami lasted 18 hours at least.
9. The storm lasted only a minute.
10. Estimated 50-60 mph wind lasted about 5 minutes.</td></tr>
<tr><td>2</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_continue_to_exist_(of natural disaster)
Frame definition: [NATURAL DISASTER] continues to exist in [TIME].
Note definition: The NATURAL DISASTER is usually a lengthy process (e.g. drought or fire). It can also include atmospheric agent, water agent, and atmospheric condition.</td></tr>
<tr><td></td><td>Semantic role</td><td>PERSIST</td><td>Theme</td><td>Time</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, atmospheric condition, atmospheric agent</td><td>_____</td></tr>
<tr><td></td><td>Linguistic</td><td></td><td>fire, drought, rainfall,</td><td>hours, two days, five</td></tr>
</table>

	realizations		storm, wind	times longer, several days
	Phrase type		NP	PP (for)
	Usage examples	1. The fire persisted for hours. 2. As a drought persists, the conditions surrounding it gradually worsen and its impact on the local population gradually increases. 3. The extremely heavy rainfall in Roxas City persisted for 2 days, with a recorded daily rainfall of 232.5 mm. 4. Smaller storms, even though they can persist five times longer, do not cross this threshold. 5. These winds can persist for several days.		
3	Lexical domain: EXISTENCE Frame: to_continue_to_exist_ (of natural disaster) Frame definition: [NATURAL DISASTER] continues to exist in [TIME]. Note: The NATURAL DISASTER is usually associated with the effect, impact, season or prediction of the natural disaster.			
	Semantic role	$EXTEND_2$	Theme	Time
	Macrorole		Undergoer	_____
	Labels		natural disaster	_____
	Linguistic realizations		effect, impact, season, prediction	hundreds of years, the year 2000, end of May, this time span
	Phrase type		NP	PP (to)
	Usage examples	1. The effect of an earthquake may extend to hundreds of years afterwards up to 1000 km away from epicenter. 2. However, this cartography of forest fire impact extends back only to the year 2000. 3. The cyclone season for Australia extends to the end of May. 4. The prediction of bush or forest fires may even extend beyond this time span.		

Table 94. Analysis of the frame *to continue to exist (of natural disaster)* (English)

1	Lexical domain: EXISTENCE Frame: to_continue_to_exist_(of natural disaster) Frame definition: [NATURAL DISASTER] continues to exist in [TIME]. Note: The NATURAL DISASTER can also include ATMOSPHERIC AGENT, and ATMOSPHERIC CONDITION.			
	Semantic role	DURAR	Theme	Time
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric agent, atmospheric condition	_____
	Linguistic realizations		tormenta, ventisca, incendio, huracán, terremoto, temporada de huracanes	cinco días, diez días, 31 días
	Phrase type		NP	NP, AVP
	Usage examples	1. Esta tormenta duró cerca de cinco días y se desvió hacia el oeste-suroeste en 2. Por ejemplo, la ventisca de nieve de Saskatchewan de 1947 duró diez días. 3. El incendio duró tres semanas y tuvo un coste total de más de 237 millones de euros. 4. El Huracán/Tifón John duró 31 días, ya que recorrió el noreste y el noroeste de las cuencas del Pacífico.		

<table>
<tr><td></td><td></td><td colspan="3">5. El terremoto de magnitud 6.7 de Northridge duró unos t5 segundos.
6. El terremoto de magnitud 7.8 de San Francisco duró casi dos minutos.
7. Este terremoto en concreto duró unos 30 segundos.
8. El tornado duró 15 minutos y los daños estimados fueron de más de 500 000 dólares.
9. La temporada de huracanes en el Atlántico de 2005 comenzó oficialmente el 1 de junio de 2005 y duró oficialmente hasta el 30 de noviembre de 2005.</td></tr>
<tr><td>2</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_continue_to_exist_(of natural disaster)
Frame definition: [NATURAL DISASTER] continues to exist in [TIME].
Note definition: In this case, the NATURAL DISASTER is usually a lengthy process (e.g. drought or fire). It can also include ATMOSPHERIC AGENT and ATMOSPHERIC CONDITION.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="4">PERDURAR</td><td>Theme</td><td>Time</td></tr>
<tr><td></td><td>Macrorole</td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td>natural disaster, atmospheric condition, atmospheric agent</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td>fuego, sequía, tormenta, viento, precipitaciones</td><td>días, cuatro siglos, toda la noche, unas tres horas</td></tr>
<tr><td></td><td>Phrase type</td><td>NP</td><td>PP (durante), AVP, NP</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. El fuego perduró durante días.
2. Esta sequía perduró alrededor de cuatro siglos y trajo como consecuencia la paulatina destrucción de la actividad agrícola
3. La tormenta perduró toda la noche.
4. El ciclón enfiló a La Habana descargando fuertes vientos y abundantes precipitaciones, que perduraron durante unas tres horas.</td></tr>
<tr><td>3</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_continue_to_exist_(of natural disaster)
Frame definition: [NATURAL DISASTER] continues to exist in [TIME].
Note: The NATURAL DISASTER can also include ATMOSPHERIC AGENT and ATMOSPHERIC CONDITION.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="4">PERSISTIR</td><td>Theme</td><td>Time</td></tr>
<tr><td></td><td>Macrorole</td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td>natural disaster, atmospheric agent, atmospheric condition</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td>huracán, inundación, sequía, incendio forestal</td><td>dos días más, once hasta las tres de la mañana, dos días,</td></tr>
<tr><td></td><td>Phrase type</td><td>NP</td><td>NP, AVP, PP (desde, hasta, durante)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. El huracán persistió dos días más.
2. El huracán persistió con toda su fuerza desde las once hasta las tres de la mañana.
3. El Gran Huracán persistió cerca de Barbados durante aproximadamente dos días.
4. La inundación persistió durante semanas.
5. La inundación persistió durante seis semanas.
6. La agricultura de Somalia es cada vez más afectada ante la intensa sequía que persiste en este país africano.
7. La sequía que sufre Aragón desde el verano pasado no solo no ha remitido, sino que persiste con fuerza.</td></tr>
</table>

		8. Persiste grave incendio forestal en Guadalajara.		
4	Lexical domain: EXISTENCE Frame: to_continue_to_exist_(of natural disaster) Frame definition: [NATURAL DISASTER] continues to exist in [TIME]. Note: The NATURAL DISASTER is usually associated with the effect, impact, season or prediction of the natural disaster.			
	Semantic role	EXTENDERSE$_2$	Theme	Time
	Macrorole		Undergoer	_____
	Labels		natural disaster	_____
	Linguistic realizations		temporada de huracanes, efectos de un terremoto	1 de junio y el 30 de noviembre, años
	Phrase type		NP	PP (entre, durante)
	Usage examples	1. La temporada de huracanes se extiende oficialmente entre el 1 de junio y el 30 de noviembre. 2. Los efectos de un terremoto se pueden extender durante años.		

Table 95. Analysis of the frame *to continue to exist (of natural disaster)* (Spanish)

As shown in Tables 94 and 95, this frame is characterized by a first argument which is a NATURAL DISASTER or something related to a NATURAL DISASTER, such as its consequences or effects. This first argument has the role of THEME and macrorole of UNDERGOER. The second argument has a TIME role. In the EXTREME EVENT, English verbs with these characteristics are 'last', 'persist', 'extend2', and Spansh verbs are 'durar', 'perdurar', 'persistir', and 'extenderse2' (Table 96).

Frame: to continue to exist (of natural disaster)			
Semantic role	Theme	*last* *persist* *extend2* *durar* *perdurar* *persistir* *extenderse2*	Time
Macrorole	Undergoer		_____
Conceptual class	natural disaster		_____
Phrase type	NP		PP (after, ago, at, between, by, during, every, for, from, in, just, on, over, to) (English) NP, PP (en, por, entre, desde, hasta, durante) (Spanish)

Table 96. Template for the frame *to continue to exist (of natural disaster)*

To continue to exist (of people)

The frame *to continue to exist (of people)* includes verbs that collocate with a first argument (HUMAN BEING), who continues to exist in its former state after being involved in a NATURAL DISASTER. As such, the first argument of these verbs is normally a HUMAN BEING or a personified CONSTRUCTION, PLANT or AREA, which is an UNDERGOER and conveys the role of EXPERIENCER. The second argument in this field is a NATURAL

DISASTER whose salient role is that of a SITUATION OR EXPERIENCE and macrorole ACTOR.

The verbs included in *to continue to exist (of people)* are 'survive', 'recover from', for English, (see Table 97), and 'sobrevivir a', 'recuperarse de' for Spanish (Table 98):

<table>
<tr><td>1</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_continue_to_exist_(of people)
Frame definition: [EXPERIENCER] continues to exist in its former state after a [NATURAL DISASTER].
Note: The EXPERIENCER is usually a human being or a personified construction, plant, or area.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="4">SURVIVE</td><td>Experiencer</td><td>Situation/Experience</td></tr>
<tr><td></td><td>Macrorole</td><td>Undergoer</td><td>Actor</td></tr>
<tr><td></td><td>Labels</td><td>human being, plant, construction, area</td><td>natural disaster</td></tr>
<tr><td></td><td>Linguistic realizations</td><td>children, person, people, church, building, Armenia</td><td>drought, landslide, earthquake</td></tr>
<tr><td></td><td>Phrase type</td><td>NP</td><td>PP (from), NP</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. Malnourished children who survive a drought may never overcome their malnutrition.
2. A landslide killed two people in Turkey, while another person survived the landslide with serious injuries.
3. How to survive from a tsunami.
4. Armenia survived the earthquake of 1998.
5. Consequently, the Australian farmer can survive droughts as severe as those that have wracked the Sahel region of Africa in recent years.
6. Only two building survived the earthquake.
7. How many people survived the tsunami in Japan 2011? Many scientists are trying to figure it out.
8. Miracle baby who survived the tornado that killed her family has died.
9. Miraculously the church, reception venue and wedding dress all survived the tornado that devastated the city.
10. Some fall crops survived the drought.
11. Tips to Survive Hurricane Sandy in New York City.
12. I survived hurricane Katrina in 2005.</td></tr>
<tr><td>2</td><td colspan="4">Lexical domain: EXISTENCE
Frame: to_continue_to_exist_(of people)
Frame definition: [EXPERIENCER] continues to exist in its former state after a [NATURAL DISASTER].
Note: The EXPERIENCER is usually a human or a personified plant, or area.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="4">RECOVER FROM</td><td>Experiencer</td><td>Situation/Experience</td></tr>
<tr><td></td><td>Macrorole</td><td>Undergoer</td><td>Actor</td></tr>
<tr><td></td><td>Labels</td><td>human being, plant, construction, area</td><td>natural disaster</td></tr>
<tr><td></td><td>Linguistic realizations</td><td>Haiti, Thailand, crops</td><td>earthquake, tsunami, drought</td></tr>
<tr><td></td><td>Phrase type</td><td>NP</td><td>NP</td></tr>
</table>

	Usage examples	1. Has Haiti recovered from the earthquake? 2. Thailand has well and truly recovered from the tsunami. But it still has to prove it can weather the storm of economic and political volatility. 3. Corn unlikely to recover from drought 4. Wheat crops have recovered from drought. 5. How long did it take Bahamas to recover from the hurricane? 6. Still to date, the region has not recovered from the hurricane with parts of the suburbs submerged in dense, jungle growth. 7. The building has recovered from the. June 2000 flash flood.

Table 97. Analysis of the frame *to continue to exist (of people)* (English)

1	Lexical domain: EXISTENCE Frame: to_continue_to_exist_(of people) Frame definition: [EXPERIENCER] continues to exist in its former state after a [NATURAL DISASTER]. Note: The EXPERIENCER is usually a human being or a personified construction, plant, or area.			
	Semantic role	SOBREVIVIR A	Experiencer	Situation/Experience
	Macrorole		Undergoer	Actor
	Labels		human being, plant, construction, area	natural disaster
	Linguistic realizations		aves, plantas, mamíferos, árboles, viviendas, casa, Miami Beach Resort, alguien, surfistas, montañistas, personas	fuego, huracanes, terremoto, tornado, tsunami, tornado, inundaciones, volcán
	Phrase type		NP	NP
	Usage examples	1. Unas 150 personas sobrevivieron a las inundaciones comiendo plantas y lodo durante una semana en el estado brasileño de Alagoas 2. Algunos árboles cerca del cauce sobrevivieron al fuego. 3. El Miami Beach Resort ha funcionado durante más de 20 años y ha sobrevivido a huracanes importantes en 1979 y 1992. 4. De las 11 viviendas de su zona, sólo la suya sobrevivió al fuego. 5. La casa sobrevivió a un gran terremoto. 6. En la mayoría de los países desarrollados se puede sobrevivir a un gran terremoto. 7. Alguien que haya sobrevivido a un tornado es más probable que perciba el peligro como real que alguien que ha vivido en una región con riesgo de tornados pero que no haya experimentado todavía ninguno. 8. Los surfistas cántabros que sobrevivieron al tsunami en las islas Mentawai, en Indonesia, recaudarán fondos para Mentawai. 9. Otros 10 montañistas sobrevivieron a la avalancha, pero muchos resultaron heridos. 10. Sólo dos o tres personas en St. Pierre sobrevivieron al volcán. .		
2	Lexical domain: EXISTENCE Frame: to_continue_to_exist_(of people) Frame definition: [EXPERIENCER] continues to exist in its former state after a [NATURAL DISASTER]. Note: The EXPERIENCER is usually a human or a personified plant, construction, or area.			
	Semantic role	RECUPERARSE DE	Experiencer	Situation/Experience
	Macrorole		Undergoer	Actor

	Labels		human being, plant, area	natural disaster
	Linguistic realizations		Haití, Nueva Orleans, Indonesia, España, cultivos	terremoto, tsunami, sequía, huracán
	Phrase type		NP	NP
	Usage examples	1. Haití aún no se ha recuperado del terremoto que le afectó hace dos años. 2. Cuando la tierra volvió a temblar ayer, la región de Emilia Romagna aun no se había recuperado del terremoto del 20 de mayo. 3. Nueva Orleans todavía no se ha recuperado del huracán "Katrina" y se ha visto de nuevo amenazada por "Gustav". 4. Desafortunadamente, Indonesia todavía no se ha recuperado del tsunami. 5. Narbona advierte de que España no se ha recuperado de la sequía a pesar de la subida de la reserva. 6. Los cultivos aún no se han recuperado de la sequía.		

Table 98. Analysis of the frame *to continue to exist (of people)* (Spanish)

The template of the frame *to continue to exist (of people)* is shown in Table 99:

Frame: to continue to exist (of people)			
Semantic role	Experiencer	*survive* *recover* *sobrevivir* *recuperarse* *survive* *recover from* *sobrevivir a* *recuperarse de*	Situation/Experience
Macrorole	Undergoer		Actor
Conceptual class	natural disaster, plant, area construction		natural disaster
Phrase type	NP		PP (from) (English) PP (a, de) (Spanish)

Table 99. Template for the frame *to continue to exist (of people)*

To cease to exist

The frame *to cease to exist* includes all verbs that normally have as their first argument a HUMAN BEING who ceases to exist as the result of a NATURAL DISASTER. Occasionally, as happens with the English verb 'disappear1', and the Spanish verb, 'desaparecer1', this argument can be personified. English verbs that belong to this frame are 'die', 'perish', 'drown1' and 'disappear1', whereas verbs include 'morir', 'fallecer', 'perecer', 'ahogarse', and 'desaparecer1'. The slight differences between these predicates are often a question of register (e.g. 'perish' which is more formal than 'die' and 'fallecer' and 'perecer', which are more formal than 'morir) (see Tables 100 and 101):

1	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER]. Note: LOCATION, TIME or MANNER can also be specified.					
	Semantic role	DIE	Experiencer	Situation/ Experience	Geographical location	Time
	Macrorole		Undergoer	_____	_____	_____
	Labels		human being	natural disaster	_____	_____
	Linguistic realizations		people, person, woman	earthquake, bushfires, landslide	northern Turkey, Bhuj (India)	8 January 1969, 1976
	Phrase type		NP	PP (in, under)	PP (in)	PP (in, on)
	Usage examples	1. In Bhuj, Gujarat (India) many thousands of such people died in the earthquake of February 2001. 2. On 8 January 1969, 23 people died in bushfires as temperatures exceeded 40°C and winds topped 3. In Izmit (Turkey) in August 1999, more than 17 000 people died in an earthquake with magnitude 7.6. 4. Fifty-four people died in this tsunami. 5. Only one out of every four people died in the horrific Tangshan earthquake in China in 1976. 6. Two people died in a landslide in northern Turkey caused by heavy rainfall. 7. Thus for every person killed in a volcanic eruption, 19 have died in earthquakes. 8. One old woman died under a landslide. 9. A household whose water buffalo has died in a flood. 10. At least 1,833 people died in the hurricane and subsequent floods, making it the deadliest U.S. hurricane since the 1928 Okeechobee hurricane. 11. More than 1500 died in the hurricane.				
2	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER]. Note: LOCATION, TIME or MANNER can also be specified. Note: Formal register					
	Semantic role	PERISH	Experiencer	Situation/ Experience	Geographical location	Time
	Macrorole		Undergoer	_____	_____	_____
	Labels		human being	natural disaster	_____	_____
	Linguistic realizations		people, inhabitants	earthquake, floods, tsunami	southern Italy, Anda-man Sea	today, last century, 1909, Boxing Day
	Phrase type		NP	PP (in)	PP (in)	PP (during, in, on)
	Usage examples	1. The intent is not to estimate how many people would perish in the same earthquake today, but rather to estimate how many people would have perished in the event had it occurred with today's population. 2. Millions of people have perished in floods during the last century. 3. In southern Italy in 1909 more than 100,000 people perished in an				

		earthquake that struck the region. 4. On Boxing Day, the news went out that about 4,000 people had perished in a tsunami from the Andaman Sea. 5. It was rumored that all of the island's 12,000 inhabitants had perished in the eruption. 6. The initial report that indicated that 18 people perished in the landslide. 7. Allan and I were completely rocked by this news, and feel such sadness for those who perished in the avalanche. 8. The Katrina Memorial in Biloxi is dedicated to the Gulf Coast victims who perished in the hurricane. 9. At least 1,836 people perished in the hurricane and subsequent floods, making it the deadliest U.S. hurricane since 1928.

3 Lexical domain: EXISTENCE
Frame: to_cease_to_exist
Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER].
Note: The NATURAL DISASTER is always a water event. LOCATION, TIME or MANNER can also be specified.

Semantic role	DROWN$_1$	Experiencer	Situation/ Experience	Geographical location	Time
Macrorole		Undergoer	_____	_____	_____
Labels		human being	natural disaster [water]	_____	_____
Linguistic realizations		people	tsunami, storm, floods	Australia	period
Phrase type		NP	PP (in, on)	PP (in)	PP (during)
Usage examples	1. 15,000 expected to have drowned in tsunami. 2. During this period, 40–50 million people drowned on the Chang River floodplain because of tropical cyclone. 3. Six people drowned in the Sydney Hobart storm and five boats sank in Australia.				

4 Lexical domain: EXISTENCE
Frame: to_cease_to_exist
Frame definition: [EXPERIENCER] or [THEME] ceases to exist in [NATURAL DISASTER].
Note: HUMAN BEING can also be extended to CONSTRUCTION or AREA. LOCATION, TIME or MANNER can also be specified.

Semantic role	DISAPPEAR$_1$	Experiencer/ Theme	Situation/ Experience	Geographical Location	Time
Macrorole		Undergoer	_____	_____	____
Labels		human being, construction, area	natural disaster	_____	____
Linguistic realizations		man, people, town, house	storm, earthquake, landslide	Nepal	2004
Phrase type		NP	PP (during, in, after)	PP (in)	PP (in)
Usage examples	1. Body of man who disappeared during storm found in Fort Worth creek. 2. Minami Sanriku: the town that disappeared in the Japan earthquake. 3. Between 25 and 30 houses disappeared in the landslide. 4. A Quebecer was among a group of people who have disappeared after an avalanche in Nepal swept through an encampment of climbers. 5. Wati, a 15 year old girl who disappeared in Indonesia's tsunami 2004				

		found his family who thought she was dead. 6. It was founded on the site that once occupied by the old Coliseum, which disappeared in the hurricane of 1847. 7. They disappeared during the hurricane.

Table 100. Analysis of the frame *to cease to exist* (English)

1	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER]. Note: LOCATION, TIME or MANNER can also be specified.					
	Semantic role	MORIR	Experiencer	Situation/ Experience	Geographical location	Time
	Macrorole		Undergoer	_____	_____	_____
	Labels		human being	natural disaster	_____	_____
	Linguistic realizations		personas, bombero	avalancha, terremoto, fuego, catástrofe, tsunami, inundaciones, huracán	Manaslu, Pakistán	domingo
	Phrase type		NP	PP (en, por, como consecuencia de)	PP (en)	NP
	Usage examples	1. Ocho personas murieron en la avalancha el domingo en Manaslu. 2. Al menos 20 000 personas murieron en el terremoto y los daños fueron de más de 1000 millones de dólares. 3. Se calcula que murieron cerca de 70.000 personas en la catástrofe. 4. Curiosamente, la mayoría de las personas que murieron en las riadas estaban en los coches. 5. El tsunami provocó olas de 15 metros que mataron a 5.388 personas; en la India murieron 10.744 personas y en Sri Lanka, hubo 30.959 víctimas. 6. Tailandia fue golpeada duramente por el tsunami en el que murieron varios miles de turistas. 7. Nueve personas murieron como consecuencia del huracán pero el número de desaparecidos es mayor. 8. Más de 400 personas murieron en las inundaciones en Pakistán. 9. Un bombero y un bombero voluntario de 16 años de edad murieron en el fuego.				
2	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER]. Note: LOCATION, TIME or MANNER can also be specified. Formal register.					
	Semantic role	PERECER	Experiencer	Situation/ Experience	Geographical location	Time
	Macrorole		Undergoer	_____	_____	_____
	Labels		human being	natural disaster	_____	_____
	Linguistic realizations		vecinos, personas, niña, guarda-	incendio, maremoto, tsunami,	Región de San Marcos, isla, Europa	año pasado

			espaldas,	terremoto, inundaciones, avalancha	Central	
	Phrase type		NP	PP (en)	PP (en)	NP
	Usage examples	1. En total, 12 de sus vecinos perecieron en el incendio y de las 11viviendas de su zona, sólo la suya sobrevivió. 2. 10 miembros de una sola familia que perecieron en el terremoto de 7.4 grados Richter en la Región de San Maros, Guatemala. 3. La historia de los tsunamis y el patrimonio cultural se fusionan en la isla de Simeulue, donde siete personas perecieron en el tsunami de 2004. 4. Un guardaespaldas de Michael Schumacher pereció en el maremoto del sudeste asiático. 5. Al menos 10 personas perecieron en las inundaciones provocadas por las lluvias intensas en Europa Central, el año pasado. 6. Siete personas, incluida una niña, perecieron en la avalancha, que afectó a la zona del paso de montaña de Rogers.				
3	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER]. Note: LOCATION, TIME or MANNER can also be specified. Formal register.					
	Semantic role	FALLECER	Experien-cer	Situation/ Experience	Geo. location	Time
	Macrorole		Undergoer	_____	_____	_____
	Labels		human being	natural disaster, atmospheric agent, atmospheric condition	_____	_____
	Linguistic realize-tions		personas	fuertes lluvias y granizo, inundaciones, deslizamiento de tierra, terremoto, volcán, ola de calor	La Paz, Comunidad Valenciana, valle del río Amarillo	19 de febrero de 2002, anualmente, el último siglo, 1969
	Morpho-logy		NP	PP (en, por, a consecuencia de)	PP (en)	PP (durante, en), NP, AVP
	Usage examples	1. El 19 de febrero de 2002 en La Paz, Bolivia, en los episodios de fuertes lluvias y granizo fallecieron 77 personas. 2. Fallecieron cuatro personas en las inundaciones en la Comunidad Valenciana y la Región de Murcia. 3. Casi 10.000 personas fallecen anualmente por terremotos de 8 grados de magnitud. 4. En el valle del río Amarillo de China, donde han ocurrido algunas de las peores inundaciones del planeta, millones de personas han fallecido por esta causa durante el último siglo. 5. La mayoría de los niños fallecieron en el deslizamiento de tierra. 6. En la erupción del Etna, en 1669, fallecieron 20.000 personas por el terremoto y los desprendimientos. 7. A consecuencia de esta ola de calor fallecieron casi 1.000 personas. 8. Entre 60.000 y 100.000 personas fallecieron en uno de los terremotos				

		más destructivos y mortales de la historia.				
4	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [HUMAN BEING] ceases to exist in [NATURAL DISASTER]. Note: The NATURAL DISASTER is always a water event. LOCATION, TIME or MANNER can also be specified.					
	Semantic role	AHOGARSE	Experiencer	Situation/ Experience	Geo. Location	Time
	Macrorole		Undergoer	_____	_____	_____
	Labels		human being	natural disaster [water]	_____	_____
	Linguistic realizations		personas, niños	tsunami, inunda-ciones	Tailandia	noche del domingo, semana pasada
	Phrase type		NP	PP (en)	PP (en)	PP (durante)
	Usage examples	1. Las personas que no se ahogan en un tsunami pueden resultar gravemente heridas ya que la fuerza de la ola las arroja contra objetos estacionarios. 2. Hemos recuperado los cadáveres de 35 personas que se ahogaron en la inundación en Tailandia durante la noche del domingo. 3. Dos niños somalíes se ahogaron en las inundaciones que barrieron el área la semana pasada.				
5	Lexical domain: EXISTENCE Frame: to_cease_to_exist Frame definition: [EXPERIENCER] or [THEME] ceases to exist in [NATURAL DISASTER]. Note: HUMAN BEING can also be extended to CONSTRUCTION or AREA. LOCATION, TIME or MANNER can also be specified.					
	Semantic role	$DESAPARECER_1$	Experiencer/ Theme	Situation/ Experience	Geo. Locat.	Time
	Macro-role		Undergoer	_____	_____	____
	Labels		human being, construction	natural disaster	_____	____
	Linguis-tic realiza-tions		personas, niña	tormenta tropical, terremoto, avalanche, tsunami, inundaciones, deslizamientos	Nepal	madru-gada de ayer
	Morpho-logy		NP	PP (en)	PP (en)	NP
	Usage examples	1. Al menos 19 personas han muerto, 22 han resultado heridas y seis han desaparecido tras el paso de la tormenta tropical 'Isaac' por Haití. 2. Al menos 11 personas murieron y un gran número de pobladores han desaparecido en la avalancha ocurrida en Nepal la madrugada de ayer. 3. La catedral desapareció en el terremoto de Pisco. 4. Una niña que desapareció en el tsunami de 2004, se reunió con su familia. 5. Al menos 422 personas murieron y otras 173 desaparecieron en las inundaciones y deslizamientos de tierra causados por las fuertes lluvias				

Table 101. Analysis of the frame *to_cease_to_exist* (Spanish)

As shown in the template of *to cease to exist* (Table 102), verbs in this frame have as their first argument a HUMAN BEING with the role of EXPERIENCER and macrorole of UNDERGOER. The second argument is a prepositional phrase involving a NATURAL DISASTER which is assigned the role of SITUATION/EXPERIENCE. The use of the asterisk in 'disappear1' and in 'desaparecer1' indicates that the first argument generally realized by a HUMAN BEING can occasionally be metaphorically extended to a CONSTRUCTION with the role of THEME.

Frame: **to cease to exist**			
Semantic role	Experiencer/Theme*	*die* *perish* *drown1* *disappear1** *morir* *perecer* *fallecer* *ahogarse* *desaparecer1**	Situation/Experience
Macrorole	Undergoer		_____
Conceptual class	human being *personified construction		natural disaster
Phrase type	NP		PP (in, under, on) (English) PP (en, por, como consecuencia de) (Spanish)

Table 102. Template of the frame *to cease to exist*

To cease to exist in the perception of others

The frame *to cease to exist in the perception of others* (Tables 103 and 104) is instantiated by predicates involving a NATURAL DISASTER associated with wind events (e.g. 'hurricane', 'typhoon'). Additionally, LOCATION, TIME, or MANNER can also be included. Verbs in this frame are 'disappear2', 'dissipate', and 'blow (itself) out1', for English, and 'desaparecer2' and 'disiparse', for Spanish. The first argument of this frame is restricted to a certain type of NATURAL DISASTER (wind events). Therefore, a construction such as 'The earthquake disappeared immediately*' is erroneous:

1	Lexical domain: EXISTENCE Frame: to_cease_to_exist_in_the_perception_of_others Definition: [NATURAL DISASTER] ceases to exist in the perception of others. Note: The NATURAL DISASTER is usually associated with wind events. LOCATION, TIME or MANNER can also be specified.				
	Semantic role	$DISAPPEAR_2$	Theme	Geographical location	Time
	Macrorole		Undergoer	_____	_____
	Labels		natural disaster [wind]	_____	_____
	Linguistic realizations		tropical storm, tornado, tropical cyclone, hurricane	Bermuda Triangle	Monday morning, mid-May

	Phrase type		NP	PP (into)	PP (on, in)
	Usage examples	1. In mid-May Tropical Storm Alberto appeared and quickly disappeared. 2. After the tornado disappeared, the principal and one of the school janitors went to check on the school. 3. As the tornado disappeared into the dark clouds, the weather chasers were unable to spot the tornado. 4. The tropical cyclone disappeared into the Bermuda Triangle, on Monday morning. 5. The weakened remains of a hurricane disappeared Saturday.			
2	Lexical domain: EXISTENCE Frame: to_cease_to_exist_in_the_perception_of_others Frame definition: [NATURAL DISASTER] ceases to exist in the perception of others. Note: The NATURAL DISASTER is usually associated with wind events. LOCATION, TIME or MANNER can also be specified.				
	Semantic role	DISSIPATE	Theme	Geographical location	Time
	Macrorole		Undergoer	_____	_____
	Labels		natural disaster [wind]	_____	_____
	Linguistic realizations		hurricane, storm, tornado, tropical cyclone	land	afternoon, next day, less than a minute, May 2, 1992
	Phrase type		NP	PP (over)	PP (by, in, on)
	Usage examples	1. The hurricane dissipated the next day over land, causing 19 fatalities. 2. After moving over land, the storm dissipated by the afternoon. 3. The tornado dissipated in less than a minute, and produced no damage. 4. The first tropical cyclone of the season formed on November 13, 1991, while the last tropical cyclone dissipated on May 2, 1992.			
3	Lexical domain: EXISTENCE Frame: to_cease_to exist_in_the_perception_of_others Frame definition: [NATURAL DISASTER] ceases to exist in the perception of others. Note: The NATURAL DISASTER is usually a wind event, though it can also be an atmospheric agent related to wind events. LOCATION, TIME or MANNER can also be specified.				
	Semantic role	BLOW (ITSELF) OUT_1	Theme	Geographical location	Time
	Macrorole		Undergoer	_____	_____
	Labels		natural disaster [wind], atmospheric agent [wind]	_____	_____
	Linguistic realizations		hurricane, tornado, storm, wind	quarter of a mile, from the coast	night, noon
	Phrase type		NP	PP (from)	PP (before, by)
	Usage examples	1. Ultimately the hurricane blew itself out the night before and left us with clear and sunny skies for our ceremony. 2. The tornado blew itself out a quarter of a mile from the coast and there was very little damage and no injures according to local officials. 3. The storm blew itself out, and the next morning the sea blinked			

		innocently in the sunlight. 4. The south wind blew itself out by noon and we had very calm conditions with the hope of sun later on.

Table 103. Analysis of the frame *to cease to exist in the perception of others* (English)

1	Lexical domain: EXISTENCE Frame: to_cease_to_exist_in_the_perception_of_others Definition: [NATURAL DISASTER] ceases to exist in the perception of others. Note: The NATURAL DISASTER is usually associated with wind events. LOCATION, TIME or MANNER can also be specified.				
	Semantic role	DESAPARECER$_2$	Theme	Geographical location	Time
	Macrorole		Undergoer	_____	_____
	Labels		natural disaster [wind]	_____	_____
	Linguistic realizations		tornado, tormenta tropical	río de la Plata	ayer de madrugada
	Phrase type		NP	PP (en)	AVP
	Usage examples	1. Una vez que tocan el suelo, los tornados viajan por lo general de seis a ocho kilómetros y duran sólo unos pocos minutos antes de debilitarse y desaparecer. 2. La tormenta tropical desapareció tan rápidamente como se había originado. 3. El tornado desapareció ayer de madrugada en el Río de la Plata sin que se haya informado de heridos.			
2	Lexical domain: EXISTENCE Frame: to_cease_to_exist_in_the_perception_of_others Frame definition: [NATURAL DISASTER] ceases to exist in the perception of others. Note: The NATURAL DISASTER is usually associated with wind events. LOCATION, TIME or MANNER can also be specified.				
	Semantic role	DISIPARSE	Theme	Geographical location	Time
	Macrorole		Undergoer	_____	_____
	Labels		natural disaster [wind]	_____	_____
	Linguistic realizations		tormenta, huracán, tornado, ciclón	Guatemala	pasado lunes
	Phrase type		NP	PP (sobre)	NP
	Usage examples	1. Incluso después de que se diga que un ciclón tropical es extratropical o se ha disipado, puede tener todavía viento con una fuerza de tormenta tropical (u ocasionalmente fuerza de huracán) y descargar abundante lluvia. 2. El NHC explicó que Iris se había disipado antes de entrar en la región este. 3. Generalmente, la actividad de rayos que van de la nube al suelo decrece cuando un tornado alcanza la superficie y regresa a su nivel normal cuando el tornado se disipa. 4. El Huracán Hattie del Atlántico (octubre-noviembre de 1961) después de disiparse sobre Guatemala el pasado lunes, contribuyó a la formación de la tormenta tropical Simone en el Pacífico Nororiental. 5. Lo que falta saber es si la tormenta se ha disipado completamente o aparecerán réplicas en un futuro.			

Table 104. Analysis of the frame *to cease to exist in the perception of others* (Spanish)

Succinctly put, verbs instantiating the frame *to cease to exist in the perception of others* are verbs whose compulsory argument should be any linguistic realization pointing to a NATURAL WIND DISASTER which fulfils the role of THEME and macrorole of UNDERGOER. Additionally, this frame can include LOCATION, TIME, or MANNER (Table 105):

Frame: **to cease to exist in the perception of others**		
Semantic role	Theme	*disappear2* *dissipate* *blow (itself) out1* *desaparecer2* *disiparse*
Macrorole	Undergoer	
Conceptual class	natural disaster [wind]	
Phrase type	NP	

Table 105. Template of the *to cease to exist in the perception of others*

To cause sb to cease to exist

The frame *to cause sb to cease to exist* (Table 106 and 107) is defined as a NATURAL DISASTER that causes a HUMAN BEING to cease to exist. Verbs instantiating this pattern are 'kill' and 'drown2', for English, and 'matar', for Spanish. As stated in the note section of the verb 'drown2', 'drown2' is constrained to NATURAL DISASTERS involving water (e.g. 'tsunami', 'floods'):

1	Lexical domain: EXISTENCE Frame: to_cause_sb_to_cease_ to_exist Frame definition: [NATURAL DISASTER] causes [HUMAN BEING] to cease to exist. Note: LOCATION, TIME or MANNER can also be specified.					
	Semantic role	KILL	Natural force	Patient	Geographical location	Time
	Macrorole		Actor	Undergoer	_____	_____
	Labels		natural disaster	human being	_____	_____
	Linguistic realizations		avalanche, tsunami, flooding, eruption, earthquake	people	south coast of Newfoundland, Bangladesh, Darwin	last century, Novem-ber 1985
	Phrase type		NP	NP	PP (on, in)	PP (in)
	Usage examples	1. It triggered a number of avalanches that killed over 70 people and caused damage of the order of US$ 1billion. 2. A magnitude 7.2 earthquake in 1929 contributed to a tsunami that killed 51 people on the south coast of Newfoundland. 3. Flooding in Indonesia at the beginning of 2002 killed at least 150 people. 4. The catastrophic 1970 storm in the Bay of Bengal killed at least 300,000 in Bangladesh. 5. There were floods elsewhere in Indonesia, including north and east				

<table>
<tr><td></td><td></td><td colspan="5">Java, where floods and landslides killed 75 people.
6. The resulting flash floods killed 63 people, another 13 were listed as missing.
7. As recently as 2000–2001, avalanches killed 176 people worldwide.
8. Almost two million people killed by earthquakes in the last century.
9. Cyclone Tracy killed 64 people out of a population of 25 000 in Darwin.
10. In November 1985 eruption of the Nevado del Ruiz volcano in Colombia killed most of the people in the path of its subsequent lahar.
11. The hurricane killed more than 400 people, nearly all in the Florida Keys.
12. In October 1935, a Category 1 hurricane killed around 2,150 people in Haiti and Honduras due to extreme rains and flashfloods.</td></tr>
<tr><td>2</td><td colspan="6">Lexical domain: EXISTENCE
Frame: to_cause_sb_to_cease_to_exist
Frame definition: [NATURAL DISASTER] causes [HUMAN BEING] to cease to exist.
Note: The NATURAL DISASTER is always a water event. LOCATION, TIME or MANNER can also be specified.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="5">DROWN$_2$</td><td>Natural force</td><td>Patient</td><td>Geographical location</td><td>Time</td></tr>
<tr><td>Macrorole</td><td>Actor</td><td>Undergoer</td><td>_____</td><td>_____</td></tr>
<tr><td>Labels</td><td>natural disaster [water]</td><td>human being</td><td>_____</td><td>_____</td></tr>
<tr><td>Linguistic realizations</td><td>tsunami, tropical cyclone</td><td>people</td><td>Sunda Strait</td><td>ten years</td></tr>
<tr><td>Phrase type</td><td>NP</td><td>NP</td><td>PP (in)</td><td>PP (ago)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="5">1. The two pre-dawn blasts each generated tsunami that drowned thousands in the Sunda Strait ten years ago.
2. For instance, tropical cyclones, apart from drowning people, usually bring very heavy rainfalls.</td></tr>
</table>

Table 106. Analysis of the frame *to cause sb to cease to exist* (English)

1	Lexical domain: EXISTENCE Frame: to_cause_sb_to_cease_ to_exist Frame definition: [NATURAL DISASTER] causes [HUMAN BEING] to cease to exist. Note: LOCATION, TIME or MANNER can also be specified.				
	Semantic role	MATAR	Natural force	Patient	Geographical location	Time
	Macrorole		Actor	Undergoer	_____	_____
	Labels		natural disaster	human being	_____	_____
	Linguistic realizations		tornado, deslizamiento, suceso, tormenta tropical, huracán, inundaciones	personas	Bangladesh, Filipinas, lago Monoun	26 de abril de 1989, noviembre de 1991
	Phrase type		NP	NP	PP (en)	NP, PP (en)
	Usage examples	1. El tornado más mortífero a nivel mundial fue el tornado de Daultipur-Salturia en Bangladesh el 26 de abril de 1989, que mató aproximadamente a 1.300 personas. 2. El deslizamiento en el Cerro Los Chorros de San Cristóbal Verapaz en Guatemala ocurrido el 4 de enero de 2009 y que mató a casi un centenar de personas. 3. La Tormenta Tropical Thelma, en noviembre de 1991, mató a miles de personas en Filipinas y nunca llegó a ser tifón; 4. Lamentablemente, la lava también provocó una explosión que mató a 60 personas. 5. El Gran Huracán de 1780 mató a 22,000 personas en las Antillas. 6. En 1984, un evento similar mató 37 personas en el lago Monoun. 7. Nuevas inundaciones mataron a 16 personas en Haití. 8. En 1939 otra inundación mató a 500.000 personas.				

Table 107. Analysis of the frame *to cause sb to cease to exist* (Spanish)

The template of *to cause sb to cease to exist* is specified in Table 108. As shown, the first argument is a NATURAL DISASTER, which is a NATURAL FORCE and activates the macrorole of ACTOR. The second argument is a HUMAN BEING which has the role of PATIENT and macrorole UNDERGOER:

Frame: **to cause sb to cease to exist**			
Semantic role	Natural force	*kill* *drown2* *matar*	Patient
Macrorole	Actor		Undergoer
Conceptual class	Natural disaster		human being
Phrase type	NP		NP

Table 108. Template of the frame *to cause sb to cease to exist*

To cause to cease to exist_(of fire disaster)

The frame *to cause to cease to exist (of fire disaster)* includes English verbs, such as 'extinguish', 'douse' (Table 109), and Spanish verbs such as 'apagar', 'extinguir', and 'sofocar' (Table 110). The first argument is either a WATER AGENT (e.g. 'storm'), which has the role of NATURAL FORCE, or a HUMAN BEING with the role of AGENT (e.g. 'firefighter'). The second argument is generally a NATURAL FIRE DISASTER with the role of PATIENT and the macrorole of UNDERGOER. In some usage examples, only the second argument is explicitly stated in the sentence, but the first argument is always implicit. LOCATION, TIME, and/or MANNER can also be specified.

1	Lexical domain: EXISTENCE Frame: to_cause_to_cease_to_exist_(of fire disaster) Frame definition: [NATURAL FORCE] or [AGENT] causes a [NATURAL FIRE DISASTER] to cease to exist. Note: The NATURAL FORCE is usually a water agent (e.g. 'rain'), and the AGENT is usually human being entity. LOCATION, TIME or MANNER can also be specified.						
	Semantic role	EXTINGUISH	Natural force/Agent	Patient	Geogr. location	Time	Man-ner
	Macro-role		Actor	Undergoer	_____	_____	____
	Labels		water agent, human being	natural disaster [fire]	_____	_____	____
	Linguis-tic realiza-tions		rain, firefighter	fire, blaze	the south of the province	early morning	suc-ces-fully
	Morpho-logy		NP	NP	PP (in)	PP (in)	AVP
	Usage examples	1. Fire can be extinguished by removing any one of the elements of the fire tetrahedron. 2. Heavy rain extinguished the fire in the south of the province. 3. The fire can be extinguished by any of the following: turning off the gas supply, covering the flame completely, application of water, application of a retardant chemical such as Halon to the flame. 4. The last remaining fires were extinguished in the early morning of 14 December. 5. Firefighters successfully extinguished St. Petersburg's largest fire in decades after an all-night battle with the inferno. 6. Firefighters managed to extinguish a fire before it reached a store in Poringland.					
2	Lexical domain: EXISTENCE Frame: to_cause_to_cease_to_exist_(of fire disaster) Frame definition: [NATURAL FORCE] or [AGENT] causes a [NATURAL FIRE DISASTER] to cease to exist. Note: The NATURAL FORCE is usually a water agent (e.g. 'rain'), and the AGENT is usually human being entity. LOCATION, TIME or MANNER can also be specified.						
	Semantic role	DOUSE	Natural force/Agent	Patient	Geogr. location	Time	Man-ner

	Macrorole		Actor	Undergoer	_____	_____	____
	Labels		water agent, human being	natural disaster [fire]	_____	_____	____
	Linguistic realizations		fire tenders, snorkel	fire	old Canada Post building	after three hours, three hours	suc-ces-fully
	Phrase type		NP	NP	PP (in)	PP (in)	AVP
	Usage examples	1. Stewards appeared and the fire was doused. 2. The fire was doused and no major damage was reported. 3. Small fire doused at old Canada Post building. 4. The fire was doused after three-four hours of operation and the cooling Theme will continue for hours. 5. Eight fire tenders and two snorkels were employed in the firefighting operation, and the fire was successfully doused in three hours.					

Table 109. Analysis of the frame *to cause to cease to exist (of fire disasters)* (English)

1	Lexical domain: EXISTENCE Frame: to_cause_to_cease_to_exist_(of fire disaster) Frame definition: [NATURAL FORCE] or [AGENT] causes a [NATURAL FIRE DISASTER] to cease to exist. Note: The NATURAL FORCE is usually a water agent (e.g. 'lluvia'), and the AGENT is usually human being entity. LOCATION, TIME or MANNER can also be specified.						
	Semantic role	APAGAR	Natural force/Agent	Patient	Geo. location	Time	Man-ner
	Macrorole		Actor	Undergoer	_____	_____	____
	Labels		water agent, human being	natural disaster [fire]	_____	_____	____
	Linguistic realizations		lluvia, bomberos	fuego, incendio	Los Realejos, Spadina Avenue	trans-curso de dos horas, 2.00 de este domingo	por com-pleto
	Phrase type		NP	NP	PP (en)	NP PP (a, en)	AVP
	Usage examples	1. La lluvia apagó el fuego en Los Realejos. 2. Bomberos de Toronto apagaron el incendio en Spadina Avenue, al norte de Dundas Street West. 3. Según el informe policial, salvo pérdidas materiales, no hubo víctimas que lamentar y el incendio se apagó a las 02:00 de este domingo. 4. Aunque el incendio se apagó pocas horas después, los pisos que quedaron más afectados tuvieron que pasar la revisión de los técnicos. 5. En el transcurso de dos horas, el incendio se apagó por completo.						
2	Lexical domain: EXISTENCE Frame: to_cause_to_cease_to_exist_(of fire disaster) Frame definition: [NATURAL FORCE] or [AGENT] causes a [NATURAL FIRE DISASTER] to cease to exist. Note: The NATURAL FORCE is usually a water agent (e.g. 'lluvia), and the AGENT is usually human being entity. LOCATION, TIME or MANNER can also be specified.						

	Semantic role	EXTINGUIR	Natural force/ Agent	Patient	Geogr. location	Time	Man-ner
	Macro-role		Actor	Undergoer	_____	_____	____
	Labels		water agent, human being	natural disaster [fire]	_____	_____	____
	Linguis-tic realiza-tions		nieve, bomberos, voluntarios, profesionales, dotación	fuego, incendio	Jalisco, Amuay	21.38h	por sí solo
	Morpho-logy		NP	NP	PP (en)	NP	AVP
	Usage examples	1. La nieve de noviembre finalmente extinguió el fuego. 2. Bomberos, voluntarios y profesionales extinguen el fuego en Amuay. 3. Autoridades extinguen el fuego en Jalisco. 4. El fuego se extinguió por sí solo. 5. El incendio se extinguió a las 21.38 horas, tan sólo quemando 0,05 hectáreas de pasto. 6. La primera dotación que empleó más de dos horas y media, extinguió el incendio que ha arrasado unos 4.000 m2 de maleza, olivos y olmos. 7. Los bomberos extinguen el incendio del Empordà.					
3	Lexical domain: EXISTENCE Frame: to_cause_to_cease_to_exist_(of fire disaster) Frame definition: [NATURAL FORCE] or [AGENT] causes a [NATURAL FIRE DISASTER] to cease to exist. Note: The NATURAL FORCE is usually a water agent (e.g. 'lluvia'), and the AGENT is usually human being entity. LOCATION, TIME or MANNER can also be specified.						
	Semantic role	SOFOCAR	Natural force/ Agent	Patient	Geogr. location	Time	Man-ner
	Macrorole		Actor	Undergoer	_____	_____	____
	Labels		water agent, human being	natural disaster [fire]	_____	_____	____
	Linguistic realizations		lluvias, efectivos, unidades de bomberos	incendio, fuego	refinería	12,15h	rápi-da-men-te
	Phrase type		NP	NP	PP (en)	NP	AVP
	Usage examples	1. Al lugar de los hechos se desplazaron efectivos de los Bomberos, que sofocaron el incendio sobre las 12,15 horas. 2. Los efectivos del servicio de extinción de incendios sofocaron el incendio. 3. Sofocan el incendio en la refinería más importante de Venezuela. 4. Cuatro unidades de bomberos sofocaron el fuego rápidamente en Navarra. 5. Las lluvias sofocaron el fuego.					

Table 110. Frame *to cause to cease to exist (of fire disasters)* (Spanish)

Table 111 shows the template for the frame *to cause to cease to exist (of fire disasters)*:

Frame: **to cause to cease to exist (of fire disaster)**			
Semantic role	Natural force/ Agent	*extinguish* *douse* *apagar* *extinguir* *sofocar*	Patient
Macrorole	Actor		Undergoer
Conceptual class	water agent/ human being		natural disaster [fire]
Phrase type	NP		NP

Table 111. Template of the frame *to cause to cease to exist (of fire disaster)*

5.4.3.2 Domain of ACTION

Apart from EXISTENCE, ACTION also plays an important role in the EXTREME EVENT. ACTION is represented in the following frames: (i) *to come against sth with sudden force*; (ii) *to come apart* (especially in reference to construction entities); (iii) *to produce fire.*

To come against sth with sudden force

The frame *to come against sth with sudden force* includes the English verbs, 'hit', 'strike','batter','blast2', and 'crash' (Table 112), and the Spanish verbs, 'golpear', 'impactar', 'batir', 'azotar', 'sacudir', and 'chocar' (Table 113). The complete analysis of each verb is shown in Tables 112 and 113:

1	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE is usually an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION. The PATIENT is usually an AREA, CONSTRUCTION, PLANT, or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, MANNER, and FREQUENCY.							
	Semantic role	HIT	Natural force	Patient	Geo. loc.	Time	Manner	Freq.
	Macro-role		Actor	Undergoer			_____	____
	Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, human being, plant, construc-tion,			_____	____
	Linguis-tic realiza-tions		earthquake, hurricane, tropical cyclone, typhoon, tornado, drought, cold wave, heat wave, tsunami,	El Salvador, land, Chile, Italy, villages, countries, coast, people,	Italy, Eur-ope	January, August 1995, 11 today, last decade	hard, badly, strong force, repeat-edly, strongly	twice

			fire, volcano, wind, reain	communi-tieshouse, schools				
	Morpho-logy		NP	NP	PP (in)	PP (in, on, at, around)	AVP	AVP
	Usage examples	1. Two earthquakes hit El Salvador in January and February. 2. The town hall always offers a refugee whenever a hurricane hits land. 3. The tropical cyclone finally hit land. 4. On August 1995, southern Chile was hit by a cold wave consisting in two successive cold fronts. 5. Italy was hit by the most intense heat wave of the last 70 years. 6. A hundred people died, half of them drowned, as a tsunami hit their villages. 7. In Europe, over the past 30 years, many countries were hit hard by droughts and water scarcity. 8. The Philippines was hit twice by a typhoon. 9. In August 2005, hurricane Katrina hit the coast of the Gulf of Mexico. 10. In 2004–2006 severe droughts hit the south-western part of Europe. 11. Floods hits northern England after torrential rain. 12. At around 11 am today, a rare tornado hit Venice in Italy. 13. Small tsunami waves hit Hawaii after Canada earthquake. 14. Floods and fire hit rail services. 15. Tsunami and volcano hit Indonesia killing hundreds. 14. Droughts, earthquakes, and other disasters do not hit people in the same way. 15. Some 106 million people were hit by floods and 60 million by drought in the last decade in Europe. 16. Every donation, regardless of size, helps to rebuild communities that are hit by natural disasters. 17. The rain hit the roof so hard that he woke up immediately. 18. The wind repeatedly hit the house with a very strong force. 19. The school authorities were aware that a tornado hitting a school full of children would generate a major disaster. 20. The tornado hit strongly three persons in Malaysia. 21. Soldiers dead and several missing as avalanche hits Northern Pakistan. 22. Wheat crops were hit by the drought.						
2	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE can be an ATMOSPHERIC AGENT, NATURAL DISASTER, ATMOSPHERIC CONDITION, but is usually a WATER AGENT The PATIENT is usually an AREA, CONSTRUCTION, PLANT, or HUMAN BEING. This action encoded is generally iterative. This is the reason why it is usually followed by time phrases headed by ‘during’ or ‘over’.							
	Semantic role	BATTER	Natural force	Patient	Geo. location	Time		
	Macrorole		Actor	Undergoer	_____	_____		
	Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, human being, construction	_____	_____		

	Linguistic realizations		hurricanes, cold wave, heat wave, flooding, floods, rain, tsunami, waves	state of Florida, Britain, Bucharest, coast, people, rocks, seawall	capital, Punta Lara, cliff	2004, week-end
	Phrase type		NP	PP (against)	PP (in, at)	PP (during, over)
	Usage examples	1. Four hurricanes battered the state of Florida during 2004. 2. A cold wave battered Bucharest over the weekend. 3. The heat wave battered some of the country's biggest corn producing states. 4. Britain battered by wind and rain causing motoring chaos. 5. The Peruvian coast was battered by high waves. 6. More than 40,000 people were battered by the flooding and pounding rain in the capital and seven provinces, half of whom fled from their inundated homes into government evacuation centers and houses of relatives. 7. These people were battered by very recently floods. 8. Once these people were battered by the tsunami and again they are facing a ferocious threat. 9. Waves battered the seawall at Punta Lara. 10. The waves battered against the rocks at the bottom of the cliff. 11. Crops were battered by intense rain. 12. Deaths as hurricane batters US East Coast. 13. One person killed as Hurricane Sandy batters Jamaica.				

3 Lexical domain: ACTION
Frame: to_come_against_sth_with_sudden_force
Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively.
Note: The NATURAL FORCE is usually an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION. The PATIENT is usually an AREA, CONSTRUCTION, PLANT, or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, MANNER and FREQUENCY.

S. role	STRIKE	Natural force	Patient	Geo. locat.	Time	Manner	Freq
Macro-role		Actor	Undergoer	_____		_____	____
Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, construc-tion, human being, plant	_____		_____	____
Lin-guistic realiza-tions		hurricane, earthquake, extratropical storm, cold wave, rains, flood, avalanche, tornado,	country, Pacific Ocean, Gulf coast, islands, village, school, building,	coast, Af-gha-nistan, Cana-dian Paci-fic	13 January 2001, Mon-day, week-end, March	sudden-ly, violent-ly, simul-tan-eously	re-peat-edly,

			drought, landslide, typhoon, winds	people	Rail-way, beach	1936		
	Mor-pholo-gy		NP	NP	PP (in, on, at)	PP (on, between, at, in)	AVP	AVP
	Usage exam-ples	1. Abrupt cold wave struck the south of the country making temperatures fell to -37 degrees F. 2. On 13 January 2001, a magnitude 7.6 earthquake struck suddenly beneath the Pacific Ocean close to the coast of El Salvador. 3. Hurricane Katrina struck the Gulf coast causing widespread damage. 4. Between 1596 and 1938, the Japanese islands were struck repeatedly by no less than 15 major tsunamis. 5. Many major extratropical storms have struck the eastern North American coastline. 5. China was struck by a magnitude 7.1 earthquake. 6. A series of high floods struck the delta. 7. Villages struck by heavy rains. 8. An avalanche struck two remote villages in northeast Afghanistan on Monday. 9. Heavy drought strikes the USA. 10. Two rare tornadoes struck the outer edges of New York City at the weekend, leading to power cuts and debris strewn across the streets. 11. In March 1936, a severe flood struck Middletown. 12. A large landslide struck a village located on the Canadian Pacific Railway, killing 30 people. 13. There are few studies of what actually happens to people after floods strike them. 14. Strong winds struck a school full of children violently. 15. Our roofs were blown away, when the wind and heavy rain struck simultaneously. 16. Disaster in the form of a cyclone wind struck the building and the logs and records blew in all directions. 17. In August 1994, seven people were struck and injured at Conery Island beach. 18. Crops were struck by a severe drought. 19. The hurricane struck Kingston on Wednesday evening and headed north. 20. If a hurricane struck your town tomorrow, would you be prepared? 21. Deadly hurricane strikes Escuminac. 22. Don't wait until Hurricane strikes to begin preparing.						

4 Lexical domain: ACTION
Frame: to_come_against_sth_with_sudden_force
Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively.
Note: The NATURAL FORCE can be an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION, but is usually a wind event (e.g. hurricane, tornado). The PATIENT is usually an AREA, or CONSTRUCTION.
Although not compulsory, it can also include LOCATION, TIME, MANNER and FREQUENCY.

Semantic role	$BLAST_3$	Natural force	Patient	Geo. location	Time	Man-ner
Macrorole		Actor	Undergoer	_____	_____	____
Labels		natural disaster [wind], atmospheric	area, construction	_____	_____	____

			agent [wind]				
	Linguistic realiza-tions		hurricane-force winds, earthquake, storm	islands, Costa Rica, West, levee, airport	Gulf Coast	Christmas day, 5 hours, overnight	down
	Morpho-logy		NP	NP	PP (on)	PP (on, for)	AVP
	Usage examples	1. Hurricane-force winds blasted the islands on Christmas Day. 5. Costa Rica blasted by 7.6 earthquake. 2. Hurricane-like storm blasted S.A for 5 hours. 4. Hurricane Isaac blasted down on the Gulf Coast late Tuesday night. 5. Fierce winds blasted the West overnight and Thursday, knocking down trees in California and forcing some schools to close. 6. Hurricane Isaac blasts the New Orleans levee. 7. Kennedy Space Center was blasted by the hurricane. 8. Strong thunderstorm winds blasted the Jefferson County Airport near Broomfield, causing an estimated 3 million dollars in damage.					
5	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE can be an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION, usually involving a water event (e.g. water). The PATIENT is usually an AREA, CONSTRUCTION or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, MANNER and FREQUENCY.						
	Semantic role	CRASH$_1$	Natural force	Patient			
	Macrorole		Actor	Undergoer			
	Labels		natural disaster, water agent, atmospheric agent	area, construction, human being			
	Linguistic realizations		waves, tsunami, wind	cliffs, rocks, house, shoreline			
	Phrase type		NP	PP (against)			
	Usage examples	1. In a gale huge storm, waves crash against the cliffs. 2. The wind howled and the waves crashed against the rocks. 3. The waves crashed against the house. 4. The tsunami crashed against a concrete wall fortifying the hillside. 5. The tsunami crashed against Thailand's shoreline on December 26, 2004.					

Table 112. Analysis of the frame *to come against sth with sudden force* (English)

1	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE is usually an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION. The PATIENT is usually an AREA, CONSTRUCTION, PLANT or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, and MANNER.						
	Seman-tic role	GOLPEAR	Natural force	Patient	Geo. loc.	Time	Ma-nner
	Macro-role		Actor	Undergoer	_____	_____	_____

	Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, human being, construction	_____	_____	_____
	Linguistic realizations		sequía, olas, tsunami, tornado, avalancha, huracán, temblor, ciclón, terremoto	Amazonia brasileña, Francia, museo, campamento, comunidad, Las Bahamas, estados de Florida y Luisiana, Bangladesh	Camelle Japón, isla indonesia de Sumatra	1926-1960, enero de 1995, mes más tarde, mañana del domingo	duramente, con fuerza, forma frecuente
	Morphology		NP	NP	PP (en, a)	PP (entre,en) NP	AVP
	Usage examples	1. La sequía golpea el corazón de la Amazonia brasileña. 2. La sequía golpea duramente a Francia. 3. Un tornado golpea con fuerza Italia. 4. Las olas golpean con dureza el museo de Man, en Camelle. 5. Tsunami golpea Hawai con olas de baja intensidad. 6. Siete desaparecidos tras la avalancha que golpeó un campamento de Nepal. 7. ¿Cuánto tiempo hace desde que un huracán o un huracán intenso golpeó una comunidad determinada de los Estados Unidos? 8. El Huracán Katrina (2005), golpeó a las Bahamas y los estados de la Florida y Luisiana. 9. Los huracanes más destructivos golpearon de forma frecuente entre 1926-1960. 10. Un temblor de magnitud 9.0, golpeó la mañana del domingo 26 de diciembre de 2004 a poca distancia de la isla indonesia de Sumatra. 11. Otro ciclón que golpeó Bangladesh en 1991 se cobró 145 000 vidas. 12. El terremoto que golpeó Kobe, en Japón, en enero de 1995 tuvo un efecto devastador en la población de la ciudad. 13. Exactamente un mes más tarde, un terremoto de magnitud 6.1 golpeó la misma zona, con el resultado de más de 3 500 personas muertas o heridas. 14. En 1996, uno de los terremotos más mortíferos registrados golpeó cerca de la ciudad minera de Tanshan, en China, provocando la muerte de más de 240.000 personas. 15. El Huracán Iniki en 1992 fue la tormenta más poderosa que golpeó Hawai en los registros históricos. 16. El Ciclón Bhola, el más mortífero registrado, golpeó la zona altamente poblada del Delta del Ganges en Pakistán. 17. El Ciclón Tropical Mahina golpeó la bahía Bathurst, en Australia en 1899. 18. El Ciclón Tropical Tracy, con vientos huracanados extendiéndose a 50 kilómetros [30 millas] de su ojo, golpeó Darwin, en Australia, el 24 de diciembre del 1974. 19. Las intensas heladas golpearon a las cosechas.					
2	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force,						

affecting it negatively.
Note: The NATURAL FORCE is usually an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION. The PATIENT is usually an AREA, CONSTRUCTION, PLANT or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, and MANNER.

Semantic role	IMPACTAR	Natural force	Patient	Geo. loc.	Time	Manner
Macro-role		Actor	Undergo-er	_____	_____	_____
Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, human being, plant, construc-tion	_____	_____	_____
Lin-guistic realiza-tions		avalancha, sequía, terremoto, tsunami, huracán, ciclón, olas	carretera, viviendas, Hawai, costa, cultivo	Paraguay, ciudades, costas, oeste de Estados Unidos	este domingo, día 3 de septiem-bre	conside-rable-mente, con intensi-dad, con vientos de más de 190km/h
Mor-phology		NP	NP	PP (en)	NP	AVP

Usage exam-ples

1. La última avalancha impactó sobre la carretera de montaña que conecta las ciudades de San Cristóbal y Chicaman.
2. La sequía impactó en el cultivo de sorjo en Paraguay.
3. Un fuerte terremoto impactó en Japón y Rusia.
4. El terremoto impactó considerablemente en varias ciudades y destruyó Guano, Patate, Pelileo, Pillaro, y un tercio de Ambato.
5. Un tsunami impactó este domingo en las costas de Hawai, en el oeste de Estados Unidos, después de un sismo de 7.7 en la escala Richter.
6. El huracán impactó sus viviendas.
7. Cuando el huracán impactó en México (Guaymas, Sonora) el día 3 de septiembre, produjo US$ 59,8 millones en daños y dejó un saldo de tres muertos y dos heridos.
8. El huracán impactó con intensidad en Santiago, que permaneció sin suministro eléctrico durante la noche y donde el viento derribó centenares de árboles e hizo volar cubiertas.
9. El ciclón impactó con vientos de más de 190 km por hora la costa del estado de Veracruz poco antes del mediodía.
10. Grandes olas impactan Hawai tras un sismo de 7,7 grados en Canadá.

3 Lexical domain: ACTION
Frame: to_come_against_sth_with_sudden_force
Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively.
Note: The NATURAL FORCE can be an ATMOSPHERIC AGENT, NATURAL DISASTER, ATMOSPHERIC CONDITION, but is usually a WATER AGENT The PATIENT is usually an AREA, CONSTRUCTION.

Semantic role	BATIR	Natural force	Patient	Geo. location	Time	Manner

	Macro-role		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster [water], atmospheric condition [water], atmospheric agent, water agent	area, construction	_____	_____	_____
	Linguis-tic realiza-tions		olas, tsunami, tormenta	paseo marítimo, playas, cultivos y frutales, rocas	distrito	a última hora	con furioso ímpetu, con furia
	Morpho-logy		NP		PP (en)	NP ,PP (a)	AVP
	Usage examples	1. Fuertes olas batieron ayer el paseo marítimo de Carril. 2. Las olas batieron, durante toda la noche, con furioso ímpetu. 3. El tsunami batió las playas de Colachel, en el distrito de Kanyakumari. 4. A última hora de la tarde de ayer una impresionante tormenta batió, con granizo de gran tamaño, los cultivos y frutales de los términos de Dehesas y Villaverde. 5. Las olas baten contra las rocas con furia y el viento agota sin piedad toda la estructura del faro.					
4	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE is usually an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION. The PATIENT is usually an AREA, CONSTRUCTION, PLANT. Although not compulsory, it can include LOCATION, TIME, and MANNER.						
	Seman-tic role	AZOTAR	Natural force	Patient	Geo. loc.	Time	Man-ner
	Macro-role		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, plant, construction	_____	_____	_____
	Lin-guistic realiza-tions		ola de frío, huracán, tormentas, inundaciones, tornado, temporal, avalancha, deslizamiento, terremoto,	Rusia, isla, Inglaterra, Mississipi, ciudad, Cuerno de África, dique de contención	sur monta-ñoso de Kosovo, Nueva Orleans, Cuba	hace 10 días, ayer por la tarde, miér-coles, semana pasada, última	con fuerza

			tsunami, tifón, sequía, incendio, olas			semana, ayer	
	Morphology		NP	NP	PP (en)	NP, PP (desde)	AVP
	Usage examples	1. Una intensa ola de frío azota Rusia desde hace diez días. 2. En Cuba, el huracán azotó el extremo occidental de la isla. 3. Fuertes tormentas e inundaciones azotaron la región. 4. Intensas tormentas acompañadas de inundaciones azotaron el sur de Inglaterra. 5. Tremendas inundaciones azotaron Mississippi. 6. Un potente tornado azotó ayer por la tarde las inmediaciones de la ciudad de Pergamino. 7. Un violento tornado azotó el miércoles la ciudad de Taranto. 8. Un nuevo temporal de nieve y viento azota España. 9. Un terremoto azota el noreste de Italia y deja al menos 6 muertos. 10. Una avalancha azotó un poblado en el sur montañoso de Kosovo. 11. El año pasado, más de 900 personas murieron en inundaciones y deslizamientos que azotaron Teresópolis y dos ciudades cercanas. 12. El tifón 'Bopha' que azotó la semana pasada el sur de Filipinas, ya está considerado el peor de este año en ese país. 13. La sequía azota al Cuerno de África de nuevo. 14. Controlado el incendio que azotó La Gomera. 15. El incendio que azotó la última semana al municipio sureño de Tavira, el más grave registrado este año en Portugal. 16. Un tsunami de 21 metros azotó Fukushima el 11 de marzo de 2011. 17. Grandes olas azotan uno de los diques de contención, ayer en Nueva Orleans. 18. Las olas azotan con fuerza el oeste de EEUU. 19. El huracán azotó a Galveston, Texas, matando a más de 6.000 personas. 20. La sequía azota severamente a las cosechas de trigo.					
5	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE is usually an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION. The PATIENT is usually an AREA, CONSTRUCTION or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, and MANNER.						
	Semantic role	SACUDIR	Natural force	Patient	Geo. loc.	Time	Manner
	Macro-role		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster, atmospheric condition, atmospheric agent, water agent	area, human being, construction	_____	_____	_____
	Linguistic realizations		huracán, tornado, sequía, explosión, incendio,	islas, archipiélago, Filipinas, agricultores, refinería,	departamento, Pakistán, Medi-	lunes, este jueves	con poderosos vientos, con fuertes

			tifón, tsunami, olas, avalancha, terremoto	campo	terráneo		lluvias
	Morpho-logy		NP	NP	PP (en)	NP	AVP
	Usage examples	1. El huracán sacudió con poderosos vientos de unos 150 km/h e intensas lluvias a la isla de Santa Lucía. 2. A su paso, el huracán sacudió el lunes el extremo norte de las islas Barlovento con lluvias, vientos y marejadas. 3. Un tornado sacudió este jueves a la ciudad de Dolores, en el departamento de Soriano. 4. Un potente terremoto de magnitud 6,8 sacudió este sábado el archipiélago de Vanuatu. 5. Un fuerte tifón sacudió el sur de Filipinas dejando a más de 50.000 personas refugiadas y sin hogares. 6. La sequía sacude fuerte a los campos argentinos. 7. Una intensa sequía sacude a pequeños agricultores de Acoyapa. 8. Una explosión seguida de un incendio sacudió ayer la mayor refinería de petróleo de Kuwait. 9. Un tsunami sacudió a Samoa tras un fuerte maremoto. 10. Un crucero y sus pasajeros han sido sacudido por fuertes vientos y olas en el Mediterráneo. 11. Fuertes olas sacuden la costa de Chile. 12. Una nueva avalancha sacude el pueblo de Talis, en Pakistán. 13. Una intensa sequía sacudió las cosechas.					

6	Lexical domain: ACTION Frame: to_come_against_sth_with_sudden_force Frame definition: [NATURAL FORCE] comes against [PATIENT] with sudden force, affecting it negatively. Note: The NATURAL FORCE can be an ATMOSPHERIC AGENT, WATER AGENT, NATURAL DISASTER or ATMOSPHERIC CONDITION, usually involving a water event (e.g. water). The PATIENT is usually an AREA, CONSTRUCTION or HUMAN BEING. Although not compulsory, it can include LOCATION, TIME, MANNER and FREQUENCY.			
	Semantic role	CHOCAR	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, water agent, atmospheric agent	area, construction, human being
	Linguistic realizations		tsunami, fuerte viento, olas	planta nuclear, viviendas, diques de protección
	Phrase type		NP	PP (contra)
	Usage examples	1. El tsunami chocó contra la planta nuclear. 2. El fuerte viento ocasionado por el huracán chocó contra todas las viviendas del municipio con brusquedad. 3. Las olas chocaron contra los diques de protección y crearon potentes corrientes que despedazaron los muelles.		

Table 113. Analysis of the frame *to come against sth with sudden force* (Spanish)

As shown in Tables 112 and 113, the frame *to come against sth with sudden force* is generally has two arguments: (i) a NATURAL DISASTER with the role of NATURAL

FORCE and macrorole of ACTOR; (ii) an AREA or CONSTRUCTION affected by the NATURAL FORCE, which has the role of PATIENT and macrorole of UNDERGOER. Table 114 shows the template *to come against sth with sudden force:*

Frame: **to_ come_against_sth_with_sudden_force**			
Semantic role	Natural force	*hit* *strike* *batter** *blast2*** *crash** *golpear* *impactar* *batir** *azotar* *sacudir* *chocar**	Patient
Macrorole	Actor		Undergoer
Conceptual class	natural disaster [water]* [wind]**		area, construction, human being
Phrase type	NP		NP

Table 114. Template of the frame *to come against sth with sudden force*

As shown (Table 114), not all the verbs that conform to this template are synonyms and can be used in the entire set of contexts. In other words, some of the verbs impose certain restrictions on their arguments. As such, the first argument of 'batter', 'crash' 'batir' and 'chocar' is a water-based NATURAL DISASTER (e.g. 'A **cold wave** battered Bucharest over the weekend'). In line with this, the first argument of 'blast2' is restricted to wind-related NATURAL DISASTERS (e.g. '**Hurricane Isaac** blasts the New Orleans levee'). In addition, the second argument of verbs such as 'hit', 'strike', 'batter', 'golpear', 'azotar', and 'sacudir' is a PATIENT, which can be a person or a construction (e.g. 'These **people** were battered by very recently floods', 'Some 106 million **people** were hit by floods and 60 million by drought in the last decade in Europe'). Finally, even though the verbs in this frame generally have two arguments, sometimes the NATURAL FORCE appears by itself with the verb since the second argument is implicit (e.g. Please, do not wait until a **hurricane** strikes to ask for help). As with the other verbs, LOCATION, TIME, and MANNER can always be included.

To (cause) to come apart (of construction artefacts)

The frame *to (cause) to come apart (of construction artefacts)* is instantiated by the English verbs, 'burst', 'collapse', and 'explode', and the Spanish verbs, 'explotar', 'explosionar', 'reventar', 'estallar', and 'desplomarse'. The verbs, 'burst', 'explode' and 'reventar' can also have causative meaning. The complete analysis is shown in Tables 115 and 116. The usage examples for the Spanish verbs 'explotar', 'estallar', and

'explosionar' are the same since they are synonyms. As such, contexts were extracted for the verb 'explotar', and for the rest of verbs, contexts were maintained, and only the verb was changed.

<table>
<tr><td>1</td><td colspan="4">Lexical domain: ACTION
Frame: to come apart (of construction artefacts)
Frame definition: [CONSTRUCTION] comes apart.
Note: SITUATION/ EXPERIENCE can also be included.</td></tr>
<tr><td rowspan="6"></td><td>Semantic role</td><td rowspan="5">BURST$_3$</td><td>Patient</td><td>Situation/Experience</td></tr>
<tr><td>Macrorole</td><td>Undergoer</td><td>_____</td></tr>
<tr><td>Labels</td><td>construction</td><td>_____</td></tr>
<tr><td>Linguistic realizations</td><td>dam, dike</td><td>torrential rain</td></tr>
<tr><td>Phrase type</td><td>NP</td><td>PP (in)</td></tr>
<tr><td>Usage examples</td><td colspan="3">1. The dam burst after torrential rain.
2. The dike burst, forcing the evacuation of hundreds of people.
3. The dam burst a few months later and the lake overflowed, sweeping away many villages.
4. It occurred when a dam burst in torrential rain.</td></tr>
<tr><td rowspan="6"></td><td colspan="4">Lexical domain: ACTION
Frame: to cause to come apart (of construction artefacts)
Frame definition: [NATURAL DISASTER] causes a [CONSTRUCTION] to come apart suddenly and violently.
Note: The NATURAL DISASTER can also be an ATMOSPHERIC AGENT, but is usually associated with wind events.</td></tr>
<tr><td>Semantic role</td><td rowspan="4">BURST$_3$</td><td>Natural force</td><td>Patient</td></tr>
<tr><td>Macrorole</td><td>Actor</td><td>Undergoer</td></tr>
<tr><td>Labels</td><td>natural disaster [wind], atmospheric agent [wind]</td><td>construction</td></tr>
<tr><td>Linguistic realizations</td><td>typhoon, rains, earthquake, winds</td><td>dike, dam, mill-dam</td></tr>
<tr><td>Phrase type</td><td>NP</td><td>NP</td></tr>
<tr><td>Usage examples</td><td colspan="3">1. The typhoon burst a dike in Kaohsiung and another in Tainan that had been damaged by the earthquake.
2. Heavy rains burst a dam just outside this desert zone.
3. The earthquake burst the mill-dam.
4. The powerful winds from the storm burst a dike and sent a twenty-foot wall of water through three towns, killing over two thousand people</td></tr>
<tr><td>2</td><td colspan="4">Lexical domain: ACTION
Frame: to come apart (of construction artefacts)
Frame definition: [CONSTRUCTION] comes apart.
Note: It includes the notion of coming apart, by moving downwards. It can include SITUATION/ EXPERIENCE, and GEOGRAPHICAL LOCATION, TIME and MANNER.</td></tr>
<tr><td rowspan="4"></td><td>Semantic role</td><td rowspan="3">COLLAPSE</td><td>Patient</td><td>Situation/Experience</td></tr>
<tr><td>Macrorole</td><td>Undergoer</td><td>_____</td></tr>
<tr><td>Labels</td><td>construction</td><td>natural disaster, atmospheric agent. Atmospheric condition</td></tr>
<tr><td>Linguistic realizations</td><td>buildings, road, schools, infrastructure,</td><td>landslide, earthquake, storm events, extreme seismic</td></tr>
</table>

			hospital	conditions
	Phrase type		NP	PP (in, under, during)
	Usage examples	1. Unfortunately, the main buildings collapsed. 2. Part of major road collapses in landslide. 3. A large number of these buildings could collapse in any future large earthquake. 4. Infrastructures continue to collapse during storm events. 5. This will help to determine which buildings will survive or collapse in an earthquake and why. 6. 992 primary schools collapsed completely. 7. Many dangerous structures could collapse under extreme seismic conditions. 8. Five hospitals collapses and 22 were seriously damaged.		
3	Lexical domain: ACTION Frame: to come apart (of construction artefacts) Frame definition: [CONSTRUCTION] comes apart. Note: It is normally as a result of rapid combustion. SITUATION/EXPERIENCE can also be included.			
	Semantic role	EXPLODE$_1$	Patient	Situation/Experience
	Macrorole		Undergoer	_____
	Labels		construction, explosive	_____
	Linguistic realizations		petrol station, gas bottle, Fukushima reactor, bombs	Earthquake
	Phrase type		NP	PP (after)
	Usage examples	1. The petrol station exploded violently as people looted petrol. 2. A GAS bottle exploded in a fire that destroyed a shed at Altona Meadows this morning. 3. Fukushima reactor exploded after earthquake.		
	Lexical domain: ACTION Frame: to (cause) to come apart (of construction artefacts) Frame definition: [NATURAL DISASTER] causes a [CONSTRUCTION] to come apart suddenly and violently. Note: It is normally as a result of rapid combustion.			
	Semantic role	EXPLODE$_1$	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster, explosive	Construction
	Linguistic realizations		hot blast	rum distillery
	Phrase type		NP	NP
	Usage examples	1. The hot blast exploded a rum distillery and ignited rum.		

Table 115. Analysis of the frame *to (cause) to come apart (of construction artefacts)* (English)

1	Lexical domain: ACTION Frame: to come apart (of construction artefacts) Frame definition: [CONSTRUCTION] comes apart. Note: SITUATION/ EXPERIENCE can also be included.			
	Semantic role	REVENTAR	Patient	Situation/Experience
	Macrorole		Undergoer	_____

	Labels		construction	_____
	Linguistic realizations		cañerías, dique, presa	frío, lluvia, vendaval
	Phrase type		NP	PP (por, debido a)
	Usage examples	1. Las cañerías reventaron por el frío. 2. El dique reventó en la madrugada debido a la intensa lluvia y al vendaval. 3. La presa reventó e inundó el valle.		
	Lexical domain: ACTION Frame: to (cause) to come apart (of construction artefacts) Frame definition: [NATURAL DISASTER] causes a [CONSTRUCTION] to come apart suddenly and violently. Note: The NATURAL DISASTER can also be an atmospheric agent, but is usually associated with wind events.			
	Semantic role	REVENTAR	Natural force	Patient
	Macrorole		Actor	Undergoer
	Labels		natural disaster [wind], atmospheric agent [wind]	Construction
	Linguistic realizations		huracán, seísmo, 458uertes vientos	dique, calles, línea de alta tensión, presas
	Phrase type		NP	NP
	Usage examples	1. El huracán Katrina reventó el dique del Canal Industrial provocando una riada que elevó el nivel de las aguas por encima de los seis metros. 2. El huracán reventó calles, inundó y destrozó casas, pero también desnudó oportunismo político, actitudes mezquinas y proselitismos infames. 3. Fuertes vientos reventaron línea principal de alta tensión. 4. El seísmo reventó presas, agrietó carreteras y derrumbó edificios.		
2	Lexical domain: ACTION Frame: to come apart (of construction artefacts) Frame definition: [CONSTRUCTION] comes apart. Note: It is normally as a result of rapid combustion.			
	Semantic role	EXPLOTAR	Patient	Situation/Experience
	Macrorole		Undergoer	_____
	Labels		construction, explosive	_____
	Linguistic realizations		depósito de petróleo, central nuclear, bombona	tsunami, incendio
	Phrase type		NP	PP (tras, durante)
	Usage examples	1. Un depósito de petróleo de una refinería explotó en Urumqi. 2. Una central nuclear explota en Japón tras el tsunami. 3. Una bombona explotó en la fábrica riojana, en Carrizal. 4. Una bombona de gas explotó durante el incendio.		
3	Lexical domain: ACTION Frame: to come apart (of construction artefacts) Frame definition: [CONSTRUCTION] comes apart. Note: It is normally as a result of rapid combustion. Note: SITUATION/ EXPERIENCE can also be included.			
	Semantic role	EXPLOSIONAR	Patient	Situation/Experience
	Macrorole		Undergoer	_____

	Labels		construction, explosive	_____
	Linguistic realizations		depósito de petróleo, central nuclear, bombona	tsunami, incendio
	Phrase type		NP	PP (tras, durante)
	Usage examples	1. Un depósito de petróleo de una refinería explosionó en Urumqi. 2. Una central nuclear explosionó en Japón tras el tsunami. 3. Una bombona explosionó en la fábrica riojana, en Carrizal. 4. Una bombona de gas explosionó durante el incendio.		
4	Lexical domain: ACTION Frame: to come apart (of construction artefacts) Frame definition: [CONSTRUCTION] comes apart. Note: It is normally as a result of rapid combustion.			
	Semantic role	ESTALLAR	Patient	Situation/Experience
	Macrorole		Undergoer	_____
	Labels		construction, explosive	_____
	Linguistic realizations		depósito de petróleo, central nuclear, bombona	tsunami, incendio
	Phrase type		NP	PP (tras, durante)
	Usage examples	1. Un depósito de petróleo de una refinería estalló en Urumqi. 2. Una central nuclear estalló en Japón tras el tsunami. 3. Una bombona estalló en la fábrica riojana, en Carrizal. 4. Una bombona de gas estalló durante el incendio.		
5	Lexical domain: ACTION Frame: to come apart (of construction artefacts) Frame definition: [CONSTRUCTION] comes apart. Note: It includes the notion of coming apart, by moving downwards. It can include SITUATION/EXPERIENCE and GEOGRAPHICAL LOCATION, TIME and MANNER.			
	Semantic role	DESPLOMARSE	Patient	Situation/Exper-ience	Manner
	Macrorole		Undergoer	_____	_____
	Labels		construction	natural disaster, atmospheric agent. Atmospheric condition	_____
	Linguistic realizations		edificio, estructura	temblor, vientos, lluvias	parcialmente
	Phrase type		NP	PP (por, debido a)	AVP
	Usage examples	1. Un edificio se desplomó en La Paz sin dejar víctimas, otras seis viviendas están en riesgo. 2. Un edificio se desploma en España por temblor de 5,2 grados. 3. Un edificio de viviendas se desploma parcialmente en pleno centro de Madrid. 4. Tres muertos y 12 heridos al desplomarse una estructura en construcción por fuertes vientos en Guatemala. 5. Una estructura de 7 metros de altura y 23 de longitud del sitio arqueológico maya El Tazumal, en el oeste de El Salvador, se desplomó el lunes debido a las fuertes lluvias que azotaron a la zona en los últimos días.			

Table 116. Analysis of the frame *to (cause) to come apart (of construction artefacts)* (Spanish)

The frame *to come apart (of construction artefacts)* is activated by an argument that is conceived generally as a CONSTRUCTION but which can also be extended to EXPLOSIVE entities (this is the case of 'explode', 'explotar', 'explosionar' or 'estallar', which are easy identifiable in the template in Table 117 by the incorporation of an asterisk). These categories are PATIENTS and UNDERGOERS. They can include very frequently the SITUATION/EXPERIENCE in which the action takes place (e.g. It occurred when a dam burst in **torrential rain**) (see Table 117):

Frame: **to come apart (of construction artefacts)**		
Semantic role	Patient	*burst* *collapse* *explode** *reventar* *explotar** *explosionar** *estallar** *desplomarse*
Macrorole	Undergoer	
Conceptual class	construction, explosive*	
Morpholoy	NP	

Table 117. Template of the frame *to come apart (of construction artefacts)*

The causative construction is realized normally by a two argument clause, being one of the arguments a NATURAL DISASTER, usually involving wind events, and another argument, which is the affected entity or PATIENT, which normally refers to a CONSTRUCTION in this specific frame (Table 118):

Frame: **to (cause) to come apart (of construction artefacts)**			
Semantic role	Natural force	*burst**	Patient
Macrorole	Actor	*explode***	Undergoer
Conceptual class	natural disaster [wind]* explosive**	*reventar**	Construction
Phrase type	NP		NP

Table 118. Template of the frame *to (cause) to come apart (of construction artefacts)*

To produce fire

Finally, within the domain of ACTION, the frame *to produce fire* was underlined. This frame is instantiated by verbs such as 'burn', 'blaze', 'smolder/smoulder', 'flare', in English, and 'arder', in Spanish (Tables 119 and 120):

1	Lexical domain: ACTION Frame: to_produce_fire Frame definition: [THEME] produces flames. Note: The THEME is always a fire or extreme heat entity. PATH, GEOGRAPHICAL LOCATION, TIME or MANNER can also be specified.						
	Semantic role	$BURN_2$	Theme	Path	Geog. location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____	_____
	Labels		natural disaster [fire]	_____	_____	_____	_____
	Linguistic realizations		fire, fire-storm	km2 hectares	Yoko-hama, areas of the country	half an hour, mid-August	out of control, rates as high as 20 km day, more slowly
	Phrase type		NP	PP (across, through)	PP (in)	PP (within, by)	AVP
	Usage examples	1. Within half-an hour, over 200 small fires were burning in Yokohama and 136 were raging in Tokyo. 7. This caused simultaneous fires in large areas of the country that burnt out of control. 8. By mid-August, at least eight separate fires were burning at rates as high as 20 km day-1 with temperatures exceeding 100 degrees. 12. The fires were extinguished only because they began to burn more slowly as they reached flatter terrain. 6. 122 827 fires burnt across 34 000 km2 – more than double the annual average. 9. After burning for three days, the firestorms had burnt through 6.4 km2 hectares.					
2	Lexical domain: ACTION Frame: to_produce_fire Frame definition: [THEME] produces flames. Note: The THEME is always a fire or extreme heat entity. PATH, GEOGRAPHICAL LOCATION, TIME or MANNER can also be specified.						
	Semantic role	BLAZE	Theme	Path	Geog. location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____	_____
	Labels		natural disaster [fire]	_____	_____	_____	_____
	Linguistic realizations		fire	square meters, area	north of the province	day and into the night, another three days, more than 4 hours	unchecked, merrily, up, with great fury
	Phrase type		NP	PP (across, over)	PP (in)	PP (through, into, for)	AVP
	Usage	1. The fire blazed through the day and into the night and was one of the					

	examples	largest fires in Glasgow for years. 2. The fire blazed across 2,000 square meters (2,400 square yards), burning several other vehicles 3. The fire blazed over an area of 2,000 square-meters, igniting several other vehicles and eventually spreading to a nearby timber mill and work sheds. 4. The fire blazed unchecked for another three days. 5. The fire blazed for more than four hours in the north of the province and destroyed the gym area, much of the interior of the building, the roof and swimming pool. 6. The fire blazed merrily. 7. The fire blazed up with great fury, and burned about two thousand men.

3 Lexical domain: ACTION
Frame: to_produce_fire
Frame definition: [THEME] produces flames.
Note: The THEME is always a fire or extreme heat entity. PATH, GEOGRAPHICAL LOCATION, TIME or MANNER can also be specified.

Semantic role	SMOLDER/ SMOULDER	Theme	Path	Geog. location	Time	Manner
Macrorole		Undergoer	_____	_____	_____	_____
Labels		natural disaster [fire]	_____	_____	_____	_____
Linguistic realizations		fire	under-ground, valley	south of Alaba-ma	all night, four months	quietly
Phrase type		NP	PP (across)	PP (in)	PP (for)	AVP
Usage examples	1. The fire smoldered all night in the south of Alabama. 2 The fire smoldered for four months until a series of coastal storms in the fall finally put an end to it. 3. Fire smoldered for three weeks before igniting. 4. Whatever the cause, once ignited, the fire smoldered underground. 5. The fire smoldered for hours before smoke prompted an evacuation. 6. For most of November, the fire smoldered quietly, being monitored by about a dozen firefighters. 7. The fire smouldered for 6 days. 8. Apparently the fire smouldered and firemen were again called to the scene at 11:40 a.m. Monday. 9. The fire smoldered across the valley.					

4 Lexical domain: ACTION
Frame: to_produce_fire
Frame definition: [THEME] produces flames.
Note: The THEME is always a fire or extreme heat entity. PATH, GEOGRAPHICAL LOCATION, TIME or MANNER can also be specified.

Semantic role	FLARE	Theme	Path	Geog. location	Time	Manner
Macrorole		Undergoer	_____	_____	_____	_____
Labels		natural disaster [fire]	_____	_____	_____	_____
Linguistic realizations		fire, blaze, flame	up, dark sky	100 feet from a pump jacket	last week, again, Thanks-giving	suddenly

						day, Saturday	
	Phrase type		NP	PP (up, into)	PP (from)	PP (on)	AVP
	Usage examples	1. The York Town fire flared up last week, but has been contained. 2. Then the fire flared up suddenly. 3. After a period of unseasonably dry weather, the Fern Lake Fire flared up again on Thanksgiving Day and remained active throughout the weekend. 4. On Saturday the blaze flared up just 100 feet from a pump jack. 5. The flame above the oil well flared into the dark sky.					

Table 119. Analysis of the frame *to produce fire* (English)

1	Lexical domain: ACTION Frame: to_produce_fire Frame definition: [THEME] produces flames. Note: The THEME is always a fire or extreme heat entity. PATH, GEOGRAPHICAL LOCATION, TIME or MANNER can also be specified.						
	Semantic role	ARDER$_2$	Theme	Path	Geog. location	Time	Manner
	Macrorole		Undergoer	_____	_____	_____	_____
	Labels		natural disaster [fire]	_____	_____	_____	_____
	Linguistic realizations		fuego, incendio	_____	región, islas	semanas, once de la noche, seis días, días	fuera de control, sin control, vivamente, de forma continua
	Phrase type		NP	_____	PP (en)	PP (durante), NP	AVP
	Usage examples	1. Indonesia captó la atención internacional por los incendios arrasadores severos que ardieron fuera de control en las islas de Borneo y Sumatra. 2. Los incendios ardieron durante varias semanas. 3. Este fuego ardía a las once de la noche sin control tras haber arrasado buena parte de un perímetro de unas 13.000 hectáreas. 4. El fuego ardía vivamente en la región. 5. El fuego ardió durante seis días y siete noches en total. 6. El fuego ardió de forma continua durante horas.						

Table 120. Analysis of the frame *to produce fire* (Spanish)

As shown in the template below (Table 121), the first argument of *to produce fire* is a NATURAL FIRE DISASTER, which is a THEME with the macrorole of UNDERGOER:

Frame: **to produce fire**		
Semantic role	Theme	*burn2*
Macrorole	Undergoer	*blaze*
Conceptual class	natural disaster [fire]	*smolder/smoulder*
Phrase type	NP	*flare*
		arder2

Table 121. Template of the frame *to produce fire*

5.4.3.2 Domain of CHANGE

Additionally, after contact takes place, there is a negative change (damage). For this reason, some of the most frequent verbs in the EXTREME EVENT belong to the lexical domain of CHANGE. More specifically, they activate the frame *to cause to change for the worse*.

To cause to change for the worse

The frame *to cause to change for the worse* includes English verbs, such as 'affect', 'damage', 'destroy', 'devastate', 'ravage', 'demolish', 'wreck', 'sweep away', 'burn1', and 'injure'. Spanish verbs include 'afectar', 'castigar', 'dañar', 'destruir', 'destrozar', 'asolar', 'arrasar', 'devastar', 'demoler', 'derrumbar', 'derribar', 'derruir', 'quemar', 'arder1', and 'calcinar'.

As shown in Tables 122 and 123, these verbs generally have two arguments: (i) a NATURAL DISASTER with the role of NATURAL FORCE and the macrorole of ACTOR; (ii) a PATIENT, which is the entity affected by the NATURAL FORCE, and which can be a CONSTRUCTION, HUMAN BEING, AREA, or PLANT. SITUATION/EXPERIENCE can also be specified, as well as LOCATION, TIME or MANNER. There are times when the verbs only occur with the PATIENT since the NATURAL FORCE is implicit (e.g. '**the house** was completely devastated').

In addition, different verbs activate different types of PATIENT For example, the PATIENT argument of 'damage' is very general, and as such, can be a CONSTRUCTION (e.g. 'the **factory** was damaged in an earthquake in San Felice'), a GEOGRAPHIC AREA (e.g. 'the hurricane damaged the **coast'**), a PLANT (e.g. '**crops** were seriously damaged by drought'), or a HUMAN BEING (e.g. 'How many **people** were damaged by hurricanes?'). In contrast, the PATIENT arguments of verbs such as 'demolish' or 'wreck' can only be CONSTRUCTION or GEOGRAPHIC AREA entities.

This is totally in consonance with the LGM. Namely, as the hierarchy becomes more specific, the information becomes more focalized, and the contexts of use are thus

more restricted. In this sense, 'demolish' and 'wreck' are more focalized than 'damage', as shown in Appendix 1. The analysis of the verbs in *to cause to change for the worse* is shown in Table 122 (English) and Table 123 (Spanish).

1	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse. Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included. LOCATION, TIME or MANNER can also be specified.							
	Semantic role	AFFECT	Natural force	Patient	Sit. loc.	Geo. Locat.	Time	Man-ner
	Macrorole		Actor	Under-goer	_____	_____	_____	_____
	Labels		natural disaster	cons-truction, human being, area, plant	_____	_____	_____	_____
	Linguistic realiza-tions		landslide, storms, earth-quake, floods, natural disaster, eruption, tsunami, tornado, hurricane	people, Tonga, urban areas, Hondu-ras, build-ings, families, homes, crops	disas-ter	Bolivia, Santa Cruz, some 150 km from La Paz, area	2001, 5 March, Novem-ber 1995	badly, again
	Morpho-logy		NP	NP	PP (in)	PP (in, at, from)	PP (during, on, in)	AVP
	Usage examples	1. Some Chinese scientists have expressed grave concern in a public warning about the dangers of landslides affecting people. 2. A timely warning and response saved many lives in cyclonic storms affecting Tonga during 2001. 3. The earthquake also affected ten other Southern California counties. 4. Urban areas seem to be most affected by landslides. 5. Honduras was again badly affected. 6. Those usually most affected by natural and other disasters are the poor and socially disadvantaged groups. 7. Paradoxically, floods are also likely to affect areas that at other times are prone to drought. 8. The floods in Bolivia, at Santa Cruz and Cochabamba affected 1,300 families. 9. Some 12,700 buildings were affected, 65 per cent of which were residential. 10. On 5 March, some 150 km from La Paz, flooding affected 100 families and displaced 225 people. 11. This is when the last mega-tsunami affected the New South Wales							

<table>
<tr><td></td><td></td><td colspan="7">coastline.
12. On 14–15 October 2000, major flooding accompanied by landslides affected the Italian Alps.
13. Over 1.2 million homes were affected in the disaster.
14. The Latur earthquake of 1993 affected a very large area.
15. The eruption affected farmers in the nearby area.
16. The tornado affected Great Bend, Kansas, in November 1915.
17. Sandy Hurricane affected residents of Union Township, NJ.
18. Atlantic city area was affected by the hurricane.
19. Crops were badly affected by the droughts.</td></tr>
<tr><td>2</td><td colspan="8">Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/ EXPERIENCE can also be included. LOCATION, TIME and MANNER can also be specified.</td></tr>
<tr><td></td><td>S. Role</td><td>DA-MAGE</td><td>Natural force</td><td>Patient</td><td>Situa-tion/ Exp.</td><td>Geo. locat.</td><td>Time</td><td>Manner</td></tr>
<tr><td></td><td>Macro-role</td><td></td><td>Actor</td><td>Under-goer</td><td>_____</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, part_of natural disaster (material), water agent, atmosphe-ric agent, atmosphe-ric condition</td><td>cons-truction, human being, area, plant</td><td>_____</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Lin-guistic reali-zations</td><td></td><td>tornado, cyclone, earthquake, lava, hurricane, flooding, drought, rockfall, water, ice, wind</td><td>people, boat, homes, houses, dam, building, the coast, crops</td><td>earth-quake, floods</td><td>region, Fossa and Stiffe, Indone-sia, Texas to Louisia-na</td><td>1985, the begin-ning of 2002</td><td>serious-ly, severely, badly</td></tr>
<tr><td></td><td>Mor-phology</td><td></td><td>NP</td><td>NP</td><td>PP (in)</td><td>PP (in, at, from…to)</td><td>PP (in, at)</td><td>AVP</td></tr>
<tr><td></td><td>Usage exam-ples</td><td colspan="7">1. Have there been instances of ships or boats being capsized or badly damaged by tornadoes?
2. In 1985 another cyclone killed 10,000 people, destroyed 17,000 homes and damaged a further 122,000.
3. The dam has already been damaged by frequent earthquakes in the region.
4. Huge rockfalls severely damaged some buildings at Fossa and Stiffe.
5. Hurricane Gilbert damaged more than 100,000 low-income homes in 1988,</td></tr>
</table>

		producing costs of $558 million. 6. Lava spilling out of the mouth of the volcano damaged several houses. 7. This is done to protect the dam from being damaged by high water or overspill. 8. Flooding in Indonesia at the beginning of 2002 killed at least 150 people and destroyed or damaged $177 million worth of infrastructure: roads, school buildings, dikes and drainage works. 9. Over 7000 homes were damaged by wind. 10. Farm buildings and electricity poles were destroyed or damaged, mainly due to the avalanche. 11. About 12 landslides destroyed more than 100 homes and damaged railways and roads. 12. This hurricane damaged the coast from Texas to eastern Louisiana. 12. Crops were seriously damaged by drought. 13. Firefighters arrive at a factory which was damaged in an earthquake, in San Felice. 14. 200 cars damaged in floods.

3 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse so that it does not exist anymore.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/ EXPERIENCE can also be included. It can also include LOCATION, TIME and MANNER.

Semantic role	DESTROY	Natural force	Patient	Sit./ Exp.	Geo. locat.	Time	Manner
Macro-role		Actor	Undergo-er	_____		_____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent, atmospheric condition	area, construction, human being, plant	_____		_____	_____
Linguistic realizations		pyroclastic flow, lava, flooding, fires, tornado, hurricane, tsunami, strong wind, waves	island, city, town, coastal regions, coastline, homes, schools, buildings, house	eruption, fire, earthquake	St. Pierre (Martinique)	2.00pm, 7 May, 1928, June 1968, 2 September 1923, a week ago	regularly, completely
Morphology		NP	NP	PP (in)	PP (at)	PP (at, on, in, ago)	AVP

	Usage examples	1. At 2.00 pm on 7 May, the eruption of Soufrière generated a pyroclastic flow that destroyed most of the northern part of the island and killed 1565 people. 2. The pyroclastic flow destroyed the city. 3. In 1928, the town of Mascali was also destroyed by lava from Mt Etna. 4. 5 million homes were destroyed by severe flooding of the Yangtze. 5. A further 7,061 buildings were destroyed by post-earthquake fires. 6. The earthquake-induced fires that destroyed Tokyo on 2 September 1923. 7. Severe flooding regularly destroys coastal regions. 8. The worst tsunami recorded destroyed Port Roya, Jamaica, in June 1692. 9. A new airport in the north of Montserrat was being built to replace the airport destroyed in the eruption. 10. How can a tornado destroy one house and leave the next one almost unscratched? 11. In 1996, some historic buildings were destroyed by hurricane Lili, but no one died. 12. A tsunami can destroy coastlines on opposite sides of an ocean basin. 13. Katrina destroyed 44 oil drilling platforms and damaged 299 pipelines. 14. 18 April 1906 was the worst American earthquake, killing 498 people, with much of the city being destroyed by ensuing fires. 15. Many of these schools were destroyed in the November earthquake. 16. Fires raged for three days, destroying over 50 per cent of both cities. 17. The cyclone destroyed two of three cashew Themeing factories. 18. Hundreds of villages were destroyed because of the tsunami. 19. The left part of the village was destroyed by the 1995 avalanche. 20. Switzerland landslide destroyed several buildings. 21. An earthquake destroys building by shaking them. 22. The whole block was completely destroyed just over a week ago in a fire. 23. The rest of the building was destroyed by a typhoon. 24. The strong wind destroyed most of the houses and buildings within a 1.5 miles radius. 25. The waves destroyed many small boats anchored in the harbor. 26. 29,000 people were destroyed by pyroclastic surges at St.Pierre (Martinique) in 1900. 27. Crops were entirely destroyed.
4	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it almost does not exist anymore. Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/ EXPERIENCE can also be included. It can also include LOCATION, TIME and MANNER.	

Semantic role	DEVASTATE	Natural force	Patient	Sit./ Exp.	Geog. locat.	Time	Manner
Macro-role		Actor	Undergoer	_____	____	____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent, atmospheric condition	area, construction, human being, plant	_____	____	____	_____

	Linguistic realizations		fire, tsunami, flooding, eruption, earthquake, wave, windstorm, hurricane	coastal towns, area, coastlines, villages, homes, houses, people, communities	floods, earthquake	Chile, southern part of the country, from Winnipeg, Japan	year, April	completely, badly
	Morphology		NP	NP	PP (in)	PP (in, on, up)	PP (last, in)	AVP
	Usage examples	1. Several houses had been devastated by a fire in April last year. 2. Three large waves devastated coastal towns in Chile. 3. Expected tsunami devastated the Aitape area of Papua New Guinea. 4. It generated a tsunami over 2 m in height that can devastate coastlines within a 1000 km radius. 5. The south of the province was devastated, first by the flooding, and then by swarms of insects. 6. The village was devastated by a huge wave. 7. The eruptions devastated populated areas on the southern part of the country. 8. Windstorm devastated southern England. 9. Two cities were completely devastated. 10. The United States sent aid to Cuba after it was devastated by a hurricane. 11. An earthquake had devastated some of the most deprived areas of Cairo. 12. 12. 350 people were killed and 390,000 fled for emergency shelter as the pyroclastic flow and lava devastated hundreds of homes and several villages. 13. The 1997 flood devastated many communities upriver from Winnipeg. 14. In 2010, 263 million people were devastated by natural disasters. 15. It's heartbreaking to see how many people were devastated by the earthquake and tsunami in Japan. 16. The Fukushima plant was devastated in an earthquake in March 2011. 17. The ancient city was badly devastated in an earthquake. 18. They are building homes to replace those devastated in the floods after Hurricane Katrina. 19. Hurricane devastated New York City in 1821. 20. The hurricane devastated the country, killing several thousand people. 21. A category 4 hurricane devastated Miami in 1926. 22. Crops were completely devastated.						
5	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore. Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included. It can also include LOCATION, TIME and MANNER.							
	Semantic role	RAVAGE	Natural force	Patient	Sit./ Exp.	Geog. locat.	Time	Manner
	Macro-role		Actor	Undergoer	_____	____	____	____
	Labels		natural disaster,	area, construc-	_____	____	____	____

			part_of natural disaster (material), water agent, atmospheric agent, atmospheric condition	tion, human being, plant				
	Linguistic realizations		hurricane, drought, earthquake, tsunami, landslide, fire	Florida, areas, state of Lousiana, town, Philippines, school, crops, people	earth-quake	East Coast, mountainous area	day	furiously
	Morphology		NP	NP	PP (in)	PP (along, in)	PP (during)	AVP
	Usage examples	1. The hurricane ravaged southern Florida. 2. The area has been ravaged by intense drought. 3. The hurricane ravaged areas along the East Coast, including New York. 4. The hurricane ravaged the state of Louisiana. 5. The fishery town was ravaged by a tsunami during the day and suffered intense blazes at night. 6. Philippines ravaged by floods. 7. New Jersey town ravaged by fire is once again engulfed in flames. 8. At least 21 people were missing and three confirmed dead after a night-time landslide ravaged a gold mining village in a mountainous area of the Philippines. 9. The earthquake ravaged Haiti. 10. Port-au-Prince was furiously ravaged in an earthquake. 11. The school was ravaged by fire during the early 1990s. 12. Crops were ravaged by golf ball size hail and 70 mph winds. 13. Twenty million people were ravaged by a mountain of flood water overflowing riverbanks and bursting dams.						
6	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore. Note: The PATIENT is normally a construction, or area. SITUATION/EXPERIENCE can also be included. It can also include LOCATION, TIME and MANNER.							
	Semantic role	DEMOLISH	Natural force	Patient	Sit. /Exp.	Geo. location	Time	Manner
	Macro-role		Actor	Undergoer	____	____	____	____
	Labels		natural disaster, part_of natural disaster (material),	construction, area	____	____	____	____

			water agent, atmos-pheric agent					
	Linguis-tic realizati-ons		rains, tsunami, landslide, earthquake, waves, hurricane	houses, district, school buildings, barrier, resorts, coast	floods, earth-quake	city of Kama-ishi, rural areas	Octo-ber, 1995	com-pletely
	Morpho-logy		NP	NP	PP (in)	PP (in, on)	PP (in)	AVP
	Usage examples	1. Heavy rains demolished 13,000 houses. 2. 150,000 houses were subsequently demolished. 3. The Ribera district was completely demolished. 4. The school buildings were demolished in 1995. 5. The tsunami demolished much of the coast. 6. The tsunami demolished the massive concrete barrier on the city of Kamaishi. 7. The landslide demolished nearly 300-tin-roofed mud huts in October. 8. The hurricane demolished nearly everything. 9. The earthquake demolished hundred of buildings in rural areas. 10. Waves demolished dozens of new beachside luxury resorts in Khao Lak. 11. The IHH also launched projects for the construction of houses for those whose houses were demolished in the floods. 12. The bridge was demolished in an earthquake. 13. Hurricane demolished roofs in Berezny district. 14. The hurricane demolished three-quarters of the structures on Grand Turk Island, while the remaining intact buildings were damaged. 15. The hurricane demolished two hotels, several churches and schoolhouses. 16. In 1900, a hurricane demolished Galveston, Texas, racking up to over $809 million dollars in damage and over 8000 lives lost.						
7	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it almost does not exist anymore. Note: The PATIENT normally refers to a construction or area. SITUATION/EXPERIENCE can also be included. It can also include LOCATION, TIME and MANNER.							
	Semantic role	WRECK	Natural force	Patient	Sit./ Exp.	Geo. loc.	Time	Manner
	Macrorole		Actor	Undergoer	___	___	____	_____
	Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent, atmospheric condition	construc-tion, area	___	___	____	_____

	Linguistic realizations		hurricane, earthquake, tsunami, landslide, waves	resorts, hotels, blocks, waterfront, building, coast	earthquake	90 km out to sea, Troughton Island	earlier in the year	pieces
	Morphology		NP	NP	PP (in)	PP (about, on)	PP (in)	PP (into)
	Usage examples	1. The hurricane wrecked resorts and hotels. 2. The earthquake wrecked 400 blocks. 3. The waterfront had been wrecked by a tsunami. 4. Building wrecked by a landslide. 5. Many more building were flattened and wrecked into pieces. 6. Strong waves wrecked the wooden boat about 90 kilometres out to sea. 7. Strong winds wrecked all installations on Troughton Island. 8. A crazy cyclone wrecked the coast earlier in the year. 9. My house was wrecked in the Northridge earthquake.						

8	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes [PATIENT] to change for the worse, so that it does not exist anymore. Note: The PATIENT is usually an area, construction, or human being. SITUATION/EXPERIENCE can also be included. It can also include LOCATION, TIME and MANNER.					
	Semantic role	SWEEP AWAY	Natural force	Patient	Sit./Exp.	Geo. location
	Macrorole		Actor	Undergoer	_____	_____
	Labels		natural disaster, part_of natural disaster (material)	construction, area, human being	_____	_____
	Linguistic realizations		hurricane, tsunami, debris flows	forest, bridge, building, town, coast, car, boy	storm	near the coast, Japan
	Phrase type		NP	NP	PP (in)	PP (in)
	Usage examples	1. About 25 hectares of forest swept away. 2. Two hours later, at 10:00 pm, the debris-laden water swept away a section of the Tangiwai Rail Bridge. 3. Debris flows swept away hundreds of buildings and killed 160 people. 4. In 1944, the Great Atlantic Hurricane swept away the south end of the town. 5. The powerful tsunami swept away cars near the coast. 6. Car swept away by tsunami in Japan. 7. Many people died when floods swept their homes away. 8. Bodies of boys swept away in storm found in NYC.				
9	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes [PATIENT] to change for the worse, so that it does not exist anymore. Note: The NATURAL FORCE is always a fire or extreme heat entity. The PATIENT is					

	normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, DESTINATION, TIME and MANNER.								
	Semantic role	BURN$_1$	Natural force	Patient	Sit./ Exp.	Geogr. location	Dest.	Time	Man ner
	Macro-role		Actor	Undergo-er	_____	_____	_____	_____	_____
	Labels		natural disaster [fire]	construc-tion, area, plant, human being	_____	_____	_____	_____	_____
	Linguis-tic realiza-tions		fire	structu-res, houses, building, school, hectares village, people, crops	fire	North Ameri-ca, the Danden-ongs, southern Califor-nia	gro-und	single day, half an hour, mid-August, Octo-ber 2003	out, to de-ath
	Morpho-logy		NP	NP	PP (in)	PP (in)	PP (to)	PP (in, within, by)	AVP
	Usage examples	1. Over 500 000 hectares of urban forest and pasture burnt in a single day. 2. In North America at least 13 fires have burned more than 400 000 hectares. 3. Over 200 structures were burnt and 16 lives lost. 10. Fires in the Dandenongs in 1962 killed 14 people and burnt 450 houses. 1. The village of Lidice was burnt to the ground in a suspicious fire. 2. The building was burnt to the ground. 3. After the school burnt to the ground in 2007, students are taught in the street. 1. In October 2003, fifteen fires in southern California burnt out 3000 km2, destroying 3640 homes and 1174 other structures. 11. Hundreds of lives were lost and an untold number of farms, houses and sawmills burnt out. 12. At least 314 people were burnt to death in a fire which engulfed factories in Pakistani cities of Karachi and Lahore on Wednesday. 13. Many homes had been destroyed, and crops were burnt.								
10	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse. Note: The AGENT is usually a natural disaster of some sort. The PATIENT is always a human being or a personified plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME and MANNER.								
	Semantic role	INJU-RE	Natural force	Patient	Sit./ Exp.	Geo. locat.	Time	Manner	
	Macrorole		Actor	Under-goer	_____	_____	_____	_____	
	Labels		natural disaster, part_of natural disaster (material),	human being, plant	_____	_____	_____	_____	

			water agent, atmospheric agent, atmospheric condition					
	Linguistic realize-tions		earthquake, avalanche, drought, hurricane, tsunami, landslide, tornado, flood, fire, hot ash	people, guest, many, villager, Iranians	floods, earth-quake, blaze	coast, island, Central Indone-sia	Fri-day	seriously
	Morpho-logy		NP,	NP	PP (in, du-ring)	PP (along, in, on)	PP (on)	AVP
	Usage examples	1. Up to 87,000 people are killed or missing and as many as 370,000 injured by an earthquake in just one county in China's south-western. 2. Thirty-thousand individuals died and around 50,000 were seriously injured by an earthquake measuring 7.9 on the Richter scale. 3. A week after Prince John Friso, 43, was seriously injured by an avalanche after skiing in Lech, Austria. 4. Hardened plants are less injured by drought. 5. This latter crop was injured by drought. 6. Many injured by hurricane Gustav. 7. Hotels and resorts know very well that having guests killed or injured by a hurricane is not the kind of publicity they want. 8. Along the southern Indian coast, 74 people were killed and many injured by a tsunami. 9. At least 25 people have been killed and dozens more injured by a landslide on a southern Philippine island. 10. Forty-two people are killed and more than 1,700 injured by a tornado in Wichita Falls. 11. The flood injured three women and took the life of another as well as those of five children, 12. The fire injured three personnel, torched the pumper and damaged a variety of equipment during the blaze. 13. A government official says one villager has been injured by an erupting volcano in central Indonesia. 14. Indonesian volcano erupts, 20 injured by hot ash. 15. At least 64 people were killed and more than a 700 injured in an earthquake that struck south-western China on Friday. 16. Iranians who were injured in an earthquake receive medical treatments in the town of Varzaqan. 17. 26 die, 17 injured in floods in Tanzania.						

Table 122. Analysis of the frame *to cause to change for the worse* (English)

1	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse. Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME or MANNER.						
Semantic role	AFECTAR	Natural force	Patient	Sit./ Exp.	Geo. Locat.	Time	Manner
Macro-role		Actor	Under-goer	_____	_____	_____	_____
Labels		natural disaster, atmospheric agent, atmospheric condition	construction, human being, area, plant	_____	_____	_____	_____
Linguistic realizations		lluvias torrenciales, ola de calor, inundaciones, terremoto, ciclones tropicales, temporal, aludes, incendio, tornado, tsunami	costas, cuenca, Mediterráneo, EE.UU	tsunami	Pacífico	período de 5 semanas, 16 junio y 31 julio, 1997, 22 de mayo de 2004	principalmente
Morphology		NP	NP	PP (en)	PP (en)	NP, PP (en, entre)	AVP
Usage examples	1. Las lluvias torrenciales afectaron también a la cuenca del Júcar. 2. El estado mexicano de Tabasco sufrió graves inundaciones que afectaron al 80% del territorio. 3. La ola de calor de julio de 1995 afectó al cuadrante noreste de los Estados Unidos y sureste de Canadá. 4. El terremoto de 1522 con epicentro en Almería afectó a amplias zonas del Mediterráneos occidental. 5. Cinco ciclones tropicales afectaron a las Islas Cook en un período de cinco semanas en el año 2005. 6. El temporal afectó incluso al norte del Caribe, causando numerosos destrozos. 7. Los aludes afectaron a las zonas alpinas, debido al cierre temporal de los grandes túneles de San Gotardo, San Bemardino, Tauren, Alrberg y Gran San Bernardo. 8. El episodio de lluvias intensas y granizo afectaron a las provincias de Anhui, Zhejiang, Jiangxi y Qinghai entre el 16 de junio y el 21 de julio de 1999. 9. Los incendios de Indonesia y Malasia en 1997 y 1998 afectaron a personas de muchos otros países de la región. 10. El conocido como terremoto de Lisboa afectó a todo el suroeste peninsular y llegó a originar un gran tsunami. 11. El huracán afectó a los estados de Alabama, Louisiana, Mississippi, Georgia y Florida.						

		12. Las inundaciones afectaron a una superficie de alrededor de un millón de kilómetros cuadrados. 13. Un fuerte tornado afectó a Hallam, Nebraska, el 22 de mayo de 2004. 14. Las inundaciones europeas de 2002 afectaron principalmente a las cuencas del Elba y del Danubio. 15. ¿Qué trayectoria han tenido los huracanes intensos que afectaron a los EE. UU. el año pasado? 16. El terremoto produjo un gran maremoto que afectó a todas las costas atlánticas. 17. ¿Qué países fueron afectados en el tsunami de 2004 en el Pacífico? 18. Huracán afectó a más de 11 mil especies de animales de Yucatán. 19. El huracán afectó a más de 213 mil personas, destruyó casas y dejó atrapados a cientos. 20. Las tormentas de ayer afectaron a las cosechas y causaron algunas inundaciones.

2 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME and MANNER.

Semantic role	CASTIGAR	Natural force	Patient	Sit./ Exp.	Geo. locat.	Time	Manner
Macro-role		Actor	Undergoer	_____	_____	_____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	human being, construction, area, plant	_____	_____	_____	_____
Linguistic realizations		terremoto, tsunami, ciclón, tifón, huracán, olas, avalancha, incendio, inundación, tornado	Jamaica, costas, archipiélago, noreste de Japón, islas Fiji, Valencia, muelle	inundaciones	El Bierzo	pasado jueves, este año, de nuevo	severamente, con especial dureza, con especial virulencia
Morphology		NP	NP, PP (a)	PP (en)	PP (en)	NP	AVP

Usage examples	1. Un potente terremoto castigó el noreste de Japón. 2. Un tsunami castigó las costas de las regiones de El Maule y el archipiélago Juan Fernández, donde causó grandes daños materiales, fallecidos y personas desaparecidas. 3. Antes de llegar a las Islas Caimán, el ciclón castigó a Jamaica, dejando al menos dieciséis muertos, miles de damnificados y cuantiosos daños. 4. Un poderoso ciclón castigó la costa este de Australia. 5. El huracán castigó la parte sur del país el pasado jueves. 6. Huracán castiga las Bahamas y amenaza el noreste de EU. 7. 'Sandy' castiga la costa este de EE.UU.

		8. Huracán Irene castiga a las Bahamas en su ruta hacia los Estados Unidos. 9. Tifón castiga severamente a China. 10. Un tifón castiga a las islas Fiji. 11. El fuego castiga de nuevo a Valencia. 12. El fuego castiga con especial dureza a Galicia este año. 13. Un incendio castiga los montes de Benuza, en El Bierzo. 14. Las inundaciones castigan a 2'5 millones de paquístanís. 15. Inundaciones castigan a Bolivia, Ecuador y Argentina. 16. Severas inundaciones castigan el norte de Brasil. 17. Otro devastador tornado castiga al estado de Oklahoma. 18. Poderoso tornado castiga a Texas. 19. Otra avalancha castiga la misma zona y sepulta a decenas de personas. 20. Las olas castigaron la costa con especial virulencia. 21. Fuertes olas castigaron el antiguo muelle comercial sin que ocurriera ningún accidente marítimo. 22. El norte de la provincia fue castigada profundamente en las inundaciones. 23. Fuertes granizadas castigaron las cosechas y un temblor de tierra puso a la región en situación desesperada.

3 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME and MANNER.

Semantic role	DAÑAR	Natural force	Patient	Sit./ Exp.	Geo. locat.	Time	Manner
Macro-role		Actor	Under-goer	_____	_____	_____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmosphe-ric agent	human being, construc-tion, area, plant	_____	_____	_____	_____
Lin-guistic realiza-tions		inunda-ciones, incendios, movimien-tos de ladera, volcán, desliza-mientos, terremoto, huracán	red de carreteras, localida-des, edificios, casas, campa-mentos, instala-ciones	terre-moto, hura-cán	parque nacional de Yose-mite, base de la ladera	1991	_____
Mor-phology		NP	NP	PP (en, por)	PP (en)	PP (en)	AVP
Usage exam-ples	1. Las inundaciones de 1994 dañaron campamentos y otras instalaciones en el parque nacional de Yosemite. 2. Los incendios también dañaron la población de orangutanes de la zona. 3. Los abundantes episodios de movimientos de ladera dañaron la red de carreteras de las zonas montañosas y numerosas localidades. 4. La inundación dañó carreteras, servicios y edificios.						

<table>
<tr><td></td><td></td><td colspan="7">5. El volcán dañó muchas casas en Shimabara, Japón, en 1991.
6. Este deslizamiento sepultó, derrumbó o dañó casas en la base de la ladera.
7. Más de edificios 1.000.000 fueron dañados en el terremoto de Japón.
8. Sólo en la ciudad de Nueva York, 61 centros de voto no podrán abrir mañana debido a que los edificios fueron dañados por el huracán están siendo empleados como albergues para desplazados o para almacenamiento y distribución de ayuda.
9. Más de 10 postes, seis transformadores y miles de metros de cable fueron dañados por el incendio.
10. El huracán dañó cosechas en Cuba, Haití y Jamaica.
11. El huracán dañó algunas viviendas en la zona del municipio Banes y provocó cortes eléctricos y telefónicos en las localidades de Cueto y Mayarí.</td></tr>
<tr><td>4</td><td colspan="8">Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can also be included as well as LOCATION, TIME and MANNER.</td></tr>
<tr><td></td><td>Seman-tic role</td><td>DESTRUIR</td><td>Natural force</td><td>Patient</td><td>Sit./ Exp.</td><td>Geo. locat.</td><td>Time</td><td>Manner</td></tr>
<tr><td></td><td>Macro-role</td><td></td><td>Actor</td><td>Undergo-er</td><td>_____</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, part_of natural disaster (material), water agent, atmosphe-ric agent</td><td>area, construc-tion, human being, plant</td><td>_____</td><td>_____</td><td>_____</td><td>_____</td></tr>
<tr><td></td><td>Lin-guistic realiza-tions</td><td></td><td>avalanchas, incendios, terre-motos, olas, riadas, huracán, sequía, tsunami, desliza-miento, tornado</td><td>casas, viviendas, escuelas, puentes, parque, ciudad</td><td>inun-dacio-nes, terre-moto, volcán</td><td>Rusia, Ma-drid, Perú</td><td>2012, hace un año, la noche del martes al miér-coles</td><td>com-pleta-mente, por com-pleto</td></tr>
<tr><td></td><td>Mor-phology</td><td></td><td>NP</td><td>NP</td><td>PP (en)</td><td>PP (en)</td><td>PP (en), NP</td><td>AVP</td></tr>
<tr><td></td><td>Usage exam-ples</td><td colspan="7">1. La avalancha de detritos, la descarga lateral, y los flujos de lodo provocaron la muerte de 57 personas y destruyeron más de 100 viviendas.
2. Los incendios de 1988 no destruyeron el parque.
3. Los terremotos casi destruyeron la ciudad de Nuevo Madrid.
4. El desastre se cobró la vida de 50 personas y las riadas destruyeron 43 puentes.
5. Las olas que generó la explosión de la isla volcánica de Santorini destruyeron al completo la ciudad de Teras.
6. El huracán destruyó cientos de casas, dejando a 30.000 personas sin hogar.</td></tr>
</table>

		7. Las inundaciones destruyeron miles de escuelas en Pakistán. 8. La avalancha destruyó 17 viviendas en Neiva. 9. La sequía destruyó en 2012 el 7,3% de los cultivos en Rusia. 10. Hace un año, un tsunami destruyó varias ciudades. 11. El terremoto destruyó la Ciudad de San Juan. 12. El deslizamiento de tierra destruyó un orfanato en Malasia. 13. Un fuerte tornado destruyó casi por completo una ciudad en Bolivia. 14. Un incendio destruyó completamente en la noche del martes al miércoles la planta séptima del hotel Mayorazgo, en Madrid. 15. En esta erupción el volcán destruyó dos ciudades importantes de la zona de Nápoles. 16. Cerca de 9.400 casas fueron destruidas en las inundaciones. 17. Más de 500 personas murieron y más de 90.000 casas fueron destruidas en el terremoto de 2007 en Perú. 18. Una vivienda y dos estructuras fueron destruidas en el incendio. 19. Las heladas destruyeron plantaciones enteras de citronella. 20. Cientos de personas están destruidas tras el paso del huracán.

5 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

	DESTRO-ZAR						
Semantic role		Natural force	Patient	Sit./ Exp.	Geo. locat.	Time	Manner
Macro-role		Actor	Undergoer	_____	_____	_____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	area, construction, human being	_____	_____	_____	_____
Linguistic realizations		incendio, avalancha, sequía, huracán, tsunami, corrimiento de tierra, tornado, erupción volcánica, inundación, lluvias	ciudad, costa, puente, casa, residencias, viviendas	tsunami, huracán	localidad asturiana, Japón, sur de China		completamente, totalmente, prácticamente por completo
Morphology		NP	NP	PP (por)	PP (en)	PP (en), NP	AVP
Usage examples	1. Un incendio iniciado el 4 de mayo ardió furiosamente fuera de control y destrozó 80 kilómetros cuadrados. 2. La avalancha destrozó el Campo del distrito de Gorkha, a unos 250 km al noroeste de la capital nepalí. 3. La avalancha destrozó dos viviendas en la localidad asturiana y tras un día						

		de búsqueda sigue sin aparecer una mujer de 77 años. 4. La sequía destrozó las cosechas, los bosques ardían sin control por todo el país y el mítico río Mississipi bajaba tan seco que ni se podía navegar. 5. El terremoto destrozó la ciudad y dejó cientos de damnificados sin techo. 6. Un terremoto destroza veinte mil viviendas en China. 7. El huracán destroza la costa este de Estados Unidos: evacuaciones masivas, millones de personas sin luz, carreteras y túneles inundados y sistemas de transporte público cerrados en Boston, Filadelfia, Nueva York y Washington. 8. El huracán destrozó completamente unas 1980 viviendas. 9. Se estima que son 1,000 las residencias totalmente destrozadas por el huracán Sandy y 80,000 las que han sufrido daños por inundaciones. 10. Cientos de muertos, barcos y casas destrozadas por el tsunami en Japón. 11. Un corrimiento de tierra destrozó prácticamente por completo el templo, del que sólo se han conservado la campana y la puerta. 12. Un tornado destrozó un estadio deportivo en Portugal. 13. Las erupciones volcánicas pueden destruir ciudades enteras y destrozar bosques. 14. Las inundaciones destrozan miles de casas en el sur de China. 15. Las intensas lluvias destrozan un puente de la A-7. 16. El intenso terremoto destroza a cientos de personas.

6 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

	ASOLAR	Natural force	Patient	Sit./ Exp.	Geog. locat.	Time	Manner
Semantic role							
Macro-role		Actor	Undergoer	_____	_____	_____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	area, construction, plant, huma being	_____	_____	_____	_____
Linguistic realizations		tormentas de granizo, huracán, inundación, avalancha, sequía, incendio, tornado, fuego, deslizamiento de tierra	campo, países, ciudad, Filipinas, hectáreas, autopista, personas	inundación, terremoto	costa del Pacífico, India, ciudad	1998, pasado martes, 29 de agosto	repentinamente
Morphology		NP	NP	PP (por)	PP (en)	NP	AVP

	Usage examples	1. El año 1988 fue extraordinario por el número de tormentas de granizo que asolaron el campo español (junio y julio), mientras que 1989 se saldaba con sequía. 2. El huracán Mitch asoló varios países de Centroamérica en 1998. 3. En noviembre de 1970, un ciclón tropical asoló el territorio de Bangladesh. 4. La avalancha asoló el pasado martes el pueblo de Wama, perteneciente al distrito de Longyang, habitado por 82 personas, de las que 71 quedaron atrapadas. 5. Una tenaz sequía asoló este lugar de Uganda. 6. El tsunami asoló más de 1.300 kilómetros en la costa del Pacífico. 7. El 29 de agosto de 2005, el huracán asoló la ciudad, provocando una de las peores catástrofes naturales en la historia de Estados Unidos. 8. Un deslizamiento de tierra asoló repentinamente una sección de la autopista ubicada en la ciudad de Keelung, en el extremo norte de Taiwán, enterrando tres vehículos y a sus cuatro ocupantes. 9. Un intenso tornado asola el estado de Kansas. 10. Las inundaciones asolan Filipinas. 11. Los pueblos de regiones como Karnataka y Andhra en la India se vieron asolados por las inundaciones en octubre de 2009. 12. Indonesia, Malasia, Tailandia, Sri Lanka, La India, Islas Maldivas y Bangladesh fueron asolados por el terremoto. 13. El fuego asoló un total de 12.560 hectáreas en el Parque Nacional Torres del Paine. 14. Decenas de personas asoladas por el terremoto de anoche.

7 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The PATIENT normally refers to an area, construction, or human being. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

Semantic role	ARRASAR	Natural force	Patient	Sit./ Exp.	Geog. locat.	Time	Manner
Macro-role		Actor	Undergoer	_____	____	____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	area, construction, human being	_____	____	____	_____
Linguistic realizations		huracán, ola, inundación, maremoto, avalancha, sequía, incendio, tornado	Nueva Orleans, costas, territorio, viviendas, templo, personas	inundación, terremoto, tsunami	Sicilia, Niyagi	martes pasado, 1647	completamente
Morphology		NP	NP	PP (por)	PP (en)	NP, PP (en)	AVP

	Usage examples	1. El Huracán Katrina arrasó entre el 27 y el 31 de agosto del 2005 Nueva Orleans. 2. Tras un terremoto de magnitud 9,2 en la escala de Richter con epicentro al oeste de Indonesia se produjo una ola de gran altura (más de 30 m en algunas zonas) que arrasó las costas de Tailandia, Indonesia, Bangladesh, Sri Lanka, India, e incluso países tan lejanos como Somalia. 3. La inundación arrasó el Valle del Guadalentín, Huerta de Murcia y Vega Baja. 4. Como consecuencia del terremoto se originaron maremotos que arrasaron las costas del Japón. 5. Todo ello fue acompañado de un maremoto que arrasó completamente la ciudad de Mesina, en Sicilia. 6. La avalancha arrasó 15 viviendas en Santa Rosa de Cabal. 7. La sequía arrasó una cuarta parte del cultivo de cereal. 8. Así fue el terremoto que arrasó Fukushima. 9. Trece días después de que un tornado arrasara parte de la ciudad de Gandía, el herido más grave sigue ingresado en el Hospital Universitario. 10. El incendio arrasó unas unas 39 hectáreas de monte el martes pasado. 11. En España en 1996, el Camping Las Nieves fue arrasado por una inesperada inundación. 12. Este templo fue arrasado por el terremoto del año 1647. 13. Con vientos de 240 kilómetros por hora, el huracán arrasó extensas áreas de Louisiana, Mississippi y Alabama. 14. En la localidad de Higashi-Matsushima, en Miyagi, el 63% de su territorio fue arrasado por el tsunami. 15. El huracán arrasó con todo a su paso por Estados Unidos. 16. Al parecer hay más de 10.000 personas arrasadas por el tsunami.

8 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it almost does not exist anymore.
Note: The PATIENT normally refers to an area, construction, or human being. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

Semantic role	DEVASTAR	Natural force	Patient	Sit./ Exp.	Geog. locat.	Time	Manner
Macro-role		Actor	Under-goer	_____	____	____	_____
Labels		natural disaster, part_of natural disaster (material), water agent, atmos-pheric agent	area, construc-tion, human being, plant	_____	____	____	_____
Linguis-tic realiza-tions		maremoto, ciclón, inundación, incendio, huracán, sequía, lluvias	templo, América Central, campiña, costas, Indonesia, aldea,	llu-vias to-rren-ciales, incen-dio	Ban-gla-desh, 10000 km del epi-	agos-to de 2010, 1991	por comple-to

			torrenciales avalancha, erupción, tsunami, terremoto	vecin-dario		centro		
	Morpho-logy		NP	NP	PP (por, en)	PP (en, a)	PP (en)	AVP
	Usage examples	1. El sismo causó un maremoto que se propagó por el océano Pacífico y devastó Hilo a 10.000 km del epicentro. 2. Ciclón devasta estado federal de Orissa India, en Bangladesh. 3. Las inundaciones fueron causadas por el Río Pánuco y devastaron vastas hectáreas de tierra. 4. El incendio arrasador en las colinas de Oakland, California, en 1991 devastó todo un vecindario, provocando la muerte de 25 personas. 5. El huracán Mitch devastó América Central y causó aproximadamente 11000 muertos. 6. La sequía devastó y despobló aquellas hermosas campiñas. 7. La ciudad fue devastada por las lluvias torrenciales. 8. El templo fue devastado en un incendio que lo destruyó por completo. 9. El tsunami resultante del terremoto devastó las costas de Indonesia. 10. Una avalancha devastó el distrito de Zhouqu por completo en agosto de 2010, dejando 1.434 muertos. 11. La erupción de un volcán devastó siete aldeas. 12. El terremoto devastó Puerto Príncipe. 13. Los campos fueron devastados por el terremoto. 14. Cientos de personas devastadas por las graves inundaciones.						
9	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore. Note: The PATIENT is normally a construction, or area. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.							
	Semantic role	DEMOLER	Natural force	Patient	Sit./ Exp.	Geo. loca-tion	Time	Man-ner
	Macro-role		Actor	Undergoer	____	____	____	____
	Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	construc-tion, area	____	____	____	____
	Linguis-tic realizati-ons		terremoto, fuertes lluvias, sismo, tsunami, tifón,	capital Puerto Príncipe, escuela, casa, vivienda,	terre-moto, incen-dio	norte de la India, calle Ato-cha	12 de enero, 2006	por com-pleto

			tornado, incendio	iglesia, pared				
	Morpho-logy		NP	NP	PP (por, en)	PP (en)	NP, PP (en)	AVP
	Usage examples	1. Un gran terremoto demolió gran parte de la capital Puerto Príncipe el 12 de enero. 2. La escuela fue demolida en el año 2006. 3. Las fuertes lluvias demolieron una escuela en el norte de la India. 4. El sismo y los tsunamis demolieron casas, derrumbaron puentes y dejaron dos millones de damnificados. 5. El tifón demolió Tokyo. 6. Más de 200 casas fueron demolidas. 7. Ocho estudiantes murieron cuando un tornado demolió las paredes de la escuela. 8. 50 viviendas de un territorio fueron demolidas por un terremoto. 9. La iglesia Santo Tomás en la calle Atocha fue demolida por completo por un incendio. 10. Miles de cubanos regresaron ayer a casas demolidas por el huracán Paloma.						
10	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [AGENT] causes a [PATIENT] to change for the worse, so that it does not exist anymore. Note: The PATIENT is normally a construction. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.							
	Seman-tic role	DERRUM-BAR	Natural force	Patient	Sit./ Exp.	Geo. loca-tion	Time	Man-ner
	Macro-role		Actor	Undergoer	____	____	____	____
	Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	construc-tion	____	____	____	____
	Lin-guistic realize-tions		movimiento del suelo, desliza-miento, avalancha, tromba de agua, olas	autovía, edificios, muros de viviendas, puentes, verjas, casas	terre-moto, tem-blor sís-mico	Versa-lles, base de la ladera, cole-gio	dos horas	____
	Mor-pho-logy		NP	NP	PP (por, duran-te)	PP (en, frente a)	PP (en)	____
	Usage exam-ples	1. El intenso movimiento del suelo derrumbó el cuarto piso superior de la autovía. 2. Un edificio de cuatro pisos de una escuela en Ahmedabad, India, se						

		derrumbó durante el terremoto. 3. Muchos edificios modernos se derrumbaron por el intenso temblor sísmico. 4. La parte de autovía que se derrumbó estaba construida sobre sedimento blando de la bahía de San Francisco. 5. En dos horas el huracán derrumbó muros de viviendas y causó el desbordamiento de arroyos, cuyas aguas arrastraron automóviles. 5. La avalancha derrumbó dos puentes en Versalles. 6. La tromba de agua derrumbó también el techo del almacén de Alcampo. 7. Del mismo modo las potentes olas derrumbaron una verja frente al colegio San Antonio, en ese mismo municipio. 8. Este deslizamiento derrumbó cientos de casas en la base de la ladera.

1 1

Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The PATIENT is normally a construction or a plant. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

Semantic role	DERRIBAR	Natural force	Patient	Sit./ Exp	Geo. location	Time	Manner
Macro-role		Actor	Undergoer	____	____	____	____
Labels		natural disaster, part_of natural disaster (material), water agent, atmos-pheric agent	construc-tion, plant	____	____	____	____
Linguistic realizations		huracán, tornado, avalancha, terremoto, tormenta, inundación, olas	viviendas, escuelas, pared, columnas, dique, árboles	terre-moto, ava-lancha	Para-guay, toda la ciu-dad	2 sema-nas des-pués, año 749	____
Morphology		NP	NP	PP (por, duran-te)	PP (en)	AVP, PP (en)	____

Usage examples	1. El huracán derribó árboles de gran tamaño en toda la ciudad. 2. Un tornado derriba unas 200 viviendas en Paraguay. 3. Anoche, una fuerte avalancha derribó a 52 viviendas, y 25 más están en grave peligro de venirse al suelo. 4. Más de dos semanas después de que un devastador terremoto derribara escuelas, destruyera ciudades enteras y causara unas 67.000 víctimas en la Provincia de Sichuan, los niños y niñas de las zonas más castigadas por el desastre sufren aún efectos psicosociales que dificultan su vuelta a la normalidad. 5. Las tormentas derribaron un gran número de árboles. 6. Otros tres operarios lograron salir al exterior después de que una inundación derribara una pared y bloqueara la vía de salida. 7. El terremoto derribó casas, templos, gasolineras, postes eléctricos y causó

		daños en las carreteras. 8. Las columnas derribadas durante el terremoto que se produjo en el año 749 d. C. 9. Los habitantes regresaron pronto a sus viviendas, excepto siete familias cuyas casas fueron inundadas y derribadas por la avalancha. 10. Las olas derribaron parte del dique.

12 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes a [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The PATIENT is normally a construction. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

Semantic role	DERRUIR	Natural force	Patient	Sit./ Exp.	Geo. location	Time	Manner
Macro-role		Actor	Undergoer	____	____	____	____
Labels		natural disaster, part_of natural disaster (material), water agent, atmospheric agent	construction	____	____	____	____
Linguistic realizations		terremoto, seísmo, temblor, olas	edificios, diques	temblor, seísmo	ciudad de Ahmedabad	martes pasado	por completo
Morphology		NP	NP	PP (a consecuencia, por)	PP (en)	NP	AVP

Usage examples
1. El terremoto derruyó cientos de edificios y causó numerosos incendios.
2. El tsunami derruyó un gran número de diques.
3. Más de 500 edificios se han derruido por completo en la ciudad de Ahmedabad, de ocho millones de habitantes por el seísmo.
4. Numerosos edificios se han derruido a consecuencia del fuerte temblor.
5. La marejada derribó edificios sacándolos de sus cimientos y las olas los derruyeron el martes pasado.

13 Lexical domain: CHANGE
Frame: to_cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The AGENT is always a fire or extreme heat entity. The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.

Semantic role	QUEMAR	Natural force	Patient	Sit./ Exp.	Geogr. location	Time	Manner
Macro-role		Actor	Undergoer	_____	_____	____	____

	Labels		natural disaster [fire]	construc-tion, area, plant, human being	_____	_____	____	____
	Lin-guistic realiza-tions		incendio fuego	terreno, cuencas, vegeta-ción, viviendas	incendio	cuenca hidrográ-fica	mayo de 2006	____
	Mor-phology		NP	NP	PP (en)	PP (en)	PP (en)	____
	Usage exam-ples	1. El incendio de Hayman, al sudoeste de Denver, Colorado, quemó aproximadamente 560 kilómetros cuadrados de bosque. 2. El fuego quemó entre 80.000 y 100.000 kilómetros cuadrados de terreno. 3. El incendio de Cedar se cobró 14 vidas y quemó unos 1.110 kilómetros cuadrados y destruyó casi 2.300 residencias. 4. En mayo de 1996, un incendio de 48 kilómetros cuadrados quemó dos de las cuencas que alimentan el embalse de agua potable. 5. El fuego quemó unos 20 kilómetros cuadrados. 6. El incendio quemó la vegetación en la cuenca hidrográfica. 7. Grandes zonas del hábitat preferido por los orangutanes se quemaron obligándoles a adaptarse a nuevas condiciones ambientales. 8. Los flujos de ceniza provocaron la muerte de 41 personas y quemaron más de 125 viviendas. 9. Los incendios arrasadores en el sur de California quemaron unas 300.000 mil hectáreas, destruyeron varios miles de viviendas. 10. Varias personas se quemaron vivas en el devastador incendio.						
14	Lexical domain: CHANGE Frame: to_cause_to_change_for_the_worse Frame definition: [NATURAL FORCE] causes [PATIENT] to change for the worse, so that it does not exist anymore. Note: The NATURAL FORCE is always a fire or extreme heat entity. The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.							
	Seman-tic role	ARDER$_1$	Natural force	Patient	Sit./Exp.	Geogr. location	Time	Man-ner
	Macro-role		Actor	Undergo-er	_____	_____	____	____
	Labels		natural disaster [fire]	construc-tion, area, plant, human being	_____	_____	____	____
	Lin-guistic realiza-tions		incendio	edificios, hectáreas, personas, vegetación	incendio	Grand Forks, zona norte	ayer	con viru-lencia
	Mor-phology		NP	NP	PP (en)	PP (en)	ADV	ADV
	Usage exam-ples	1. Los edificios ardieron en Grand Forks, Dakota del norte, en un incendio provocado por la inundación del rio Rojo del norte en 1997. 2. En estos incendios ardieron entre 70 y 80 hectáreas en la zona norte y otras 13 en el suroeste.						

<table>
<tr><td></td><td></td><td colspan="7">3. 32 hectáreas ardieron ayer en la zona del Montermoso en un devastador incendio.
4. Otras 20 hectáreas ardieron con virulencia el año pasado en esta misma región.
5. Varias personas arden en el incendio detectado en el día de ayer.
6. La vegetación ardió abundantemente y las llamas llegaron a afectar al cierre de la finca de una vivienda</td></tr>
<tr><td>15</td><td colspan="8">Lexical domain: CHANGE
Frame: to_ cause_to_change_for_the_worse
Frame definition: [NATURAL FORCE] causes [PATIENT] to change for the worse, so that it does not exist anymore.
Note: The NATURAL FORCE is always a fire or extreme heat entity. The PATIENT is normally a construction, human being, area, or plant. SITUATION/EXPERIENCE can be included as well as LOCATION, TIME and MANNER.</td></tr>
<tr><td></td><td>Semantic role</td><td>CALCINAR</td><td>Natural force</td><td>Patient</td><td>Sit./Exp. locat.</td><td>Geogr. location</td><td>Time</td><td>Manner</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Actor</td><td>Undergoer</td><td>_____</td><td>_____</td><td>____</td><td>____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster [fire]</td><td>construction, area, human being, plant</td><td>_____</td><td>_____</td><td>____</td><td>____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>incendio, fuego</td><td>casa, vivienda, naves industriales, hectáreas, superficie, personas</td><td>fuego, incendio</td><td>Granada, Anero</td><td>este verano</td><td>____</td></tr>
<tr><td></td><td>Morphology</td><td></td><td>NP</td><td>NP</td><td>PP (en)</td><td>PP (en)</td><td>PP (durante)</td><td>____</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="7">1. Un incendio calcina una casa en Anero.
2. Un incendio calcina parte de una vivienda de Villaviciosa.
3. Un incendio calcina el Balneario de Lodos de Torrevieja.
4. El fuego calcina al menos siete naves industriales de Azuqueca.
5. El fuego calcina en Granada una superficie igual a la de 500 campos de fútbol durante este verano.
6. 200 hectáreas calcinadas en el incendio.
7. Todavía falta la confirmación definitiva del número de hectáreas calcinadas en el fuego que ha afectado a cinco poblaciones de la provincia de Sevilla.
8. Varias personas calcinadas en el incendio.
9. Controlado el fuego que ha calcinado varias plantas protegidas.</td></tr>
</table>

Table 123. Analysis of the frame *to cause to change for the worse* (Spanish)

As shown in Tables 122 and 123, when the verbs in the frame *to cause to change for the worse* have one argument, it is a CONSTRUCTION, GEOGRAPHICAL AREA, HUMAN BEING or PLANT. The argument has the role of PATIENT and activates the macrorole of UNDERGOER. The natural disaster then is either implicit in the action of the verb or appears as a SITUATION or EXPERIENCE in which the action occurs.

When the verb has two arguments, the other one will generally be any kind of NATURAL DISASTER, which fulfils the role of NATURAL FORCE and acts as an ACTOR.

However, this does not mean that these verbs are interchangeable in all contexts. For this reason, the following template (Table 124) is subdivided into three columns.

Frame: **to_cause_to_change_for_the_worse**						
Semantic role	Natural force	*affect* *damage* *destroy* *devastate* *ravage* *sweep away* *afectar* *castigar* *dañar* *destruir* *destrozar* *asolar* *arrasar* *devastar*	*burn1** *quemar** *arder1** *calcinar**	*demolish* *wreck* *demoler* *derrumbar* *derribar* *derruir*	*injure*	Patient
Macrorole	Actor					Undergoer
Conceptual class	natural disaster [fire]*					construc-tion, area, human being, plant
Phrase type	NP					NP

Table 124. Template of the frame *to cause to change for the worse*

In this regard, English verbs such as 'burn1' and Spanish verbs such 'quemar', 'arder1', and 'calcinar' (Spanish), tend to appear with NATURAL DISASTER entities that are fire events (e.g. 'The **fire** burnt the house'). Moreover, the verbs in the first column on the left ('affect', 'damage', 'destroy', 'devastate', 'ravage', 'sweep away' and 'afectar', 'castigar', 'dañar', 'destruir', 'destrozar', 'asolar', 'arrasar', 'devastar') tend to have a PATIENT which can be a CONSTRUCTION, AREA, PLANT, or HUMAN BEING, i.e. the entire set of categories identified for PATIENTS in the analysis. Verbs in the third column, i.e. 'demolish', 'wreck' 'demoler' 'derrumbar' 'derribar' 'derruir' tend to have PATIENTS that are CONSTRUCTION or GEOGRAPHICAL AREA entities. Finally, 'injure' has PATIENTS that are HUMAN BEINGS and PLANTS.

As can be observed, this template is almost the same as that corresponding to the frame *to come against sth with sudden force*. The difference resides in the definitions of the verbs contained in each frame, which clearly differ from each other. Another difference between the two frames lies in the obligatoriness of certain arguments. As such, in *to change sth for the worse* the only strictly obligatory argument is PATIENT (e.g. '**The coast** was devastated') since NATURAL FORCE is implicit. In contrast, in *to come against sth with sudden force*, the obligatory argument is NATURAL DISASTER '**the hurricane** struck with force' and the PATIENT is implicit.

5.4.3.4 Domain of MOVEMENT

The domain of MOVEMENT is relevant to the EXTREME EVENT frame, as reflected in the large number of verbs of movement verbs extracted. Verbs involving MOVEMENT within the EXTREME EVENT generally activate the following frames: *to move forcefully; to move slowly; to (cause) to move vertically; to move horizontally over a large space*, and *to cause motion*.

To move forcefully

The frame *to move forcefully* is instantiated in English by 'surge', 'sweep', 'blast', and 'burst'. These verbs generally have two arguments: (i) a NATURAL DISASTER; (ii) PATH followed by the NATURAL DISASTER. The nuance of *forcefully* is part of the meaning of the verbs. The usage examples in Table 125 indicate that these NATURAL DISASTERS are associated with wind and water events (e.g. 'hurricane', 'floods').

This frame does not have Spanish verbs that transmit this meaning, given the differences in the way that movement is lexicalized in the two languages. In Spanish, this type of meaning is conveyed by combining verbs from the frame *to move in a certain direction* (e.g. 'ascender', 'elevar', 'emerger', 'transportar', 'penetrar', 'adentrar', 'entrar', 'cruzar', 'atravesar'), or simply *to move* (e.g 'mover', 'desplazar', 'circular') with adverbial modification indicating intensity. For this reason, both in this frame and in the following one, *to move slowly* (Table 127), there are no verbs in Spanish.

<table>
<tr><td>1</td><td colspan="4">Lexical domain: MOVEMENT
Frame: to_move_forcefully
Frame definition: [NATURAL DISASTER] moves suddenly and forcefully in a specified direction [PATH].
Note: The NATURAL DISASTER can also include ATMOSPHERIC and WATER AGENTS, but they should always be associated with water or wind. PATH usually indicates forward or upward movement.</td></tr>
<tr><td></td><td>Semantic role</td><td>SURGE</td><td>Theme</td><td>Path</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster [water] [wind], atmospheric agent, water agent</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>tsunami, waters, monsoon, wall of water, tide, wind, storm, hurricane</td><td>height 14m above sea level, streets, shoreline, breakwater, ground, east coast of the United States</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>PP (to, through, up, over)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. The tsunami had surged to a height 14 m above sea level.
2. Although the storm lasted just 45 minutes, flood waters up to 2 m high surged through the streets.
3. The southwest monsoon may have surged on June 20.
4. The resulting wall of water surged 524 m up the shoreline on the opposite side
5. The tide surged over the breakwater.
6. A hot wind surged over the ground, laden with dust and smoke;
7. In 1993, an early March storm surged up the east coast of the United States.
8. The hurricane surged through New York in the early hours of Sunday morning, flooding waterfronts and low-lying areas.
9. The waters of the gulf tossed and roiled as the hurricane surged toward the shore.</td></tr>
<tr><td>2</td><td colspan="4">Lexical domain: MOVEMENT
Frame: to_move_forcefully
Frame definition: [NATURAL DISASTER] moves quickly and forcefully in a specified direction [PATH].
Note: It is usually over a large area</td></tr>
<tr><td></td><td>Semantic role</td><td>SWEEP$_1$</td><td>Theme</td><td>Path</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster, atmospheric agent, water agent</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>fire, typhoon, hurricane, tsunami, tornado, landslide, floods, rains, thunderstorms, waves</td><td>north-east, Tokyo area, the city, Gulf of Mexico, coastline, Senegal, Pacific Ocean, towns and villages, uphill, all major streams, English channel, Whangaehu River</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>PP (through, into, onto, across, over, up, down)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. Disastrous fires swept through the north-east in 1880.
2. A typhoon swept through the Tokyo area on 1 September.
3. Shortly after the earthquake, a tsunami swept into the city.
4. The hurricane then swept into the Gulf of Mexico.
5. Canada is seismically active and has produced a tsunami that has</td></tr>
</table>

<table>
<tr><td></td><td></td><td colspan="3">swept onto that coastline.
6. Hurricane Hazel swept through southern Ontario, Canada, in October 1954.
7. A devastating fire swept through Canberra 20 years ago.
8. Heavy off-season rains swept through Senegal between 9–11 January.
9. Since 1980, large fires have swept through the tropics of Indonesia and Amazonia.
10. Waves swept through the city and killed hundreds of people.
11. The tornado swept through Oklahoma and Kansas.
12. Tsunami exceeded 10 m in height along the Alaskan coast and swept across the Pacific Ocean at intervals of one hour.
13. This resulted in large landslides that swept as flow failures over numerous towns and villages.
14. On 21 January 1991, a thunderstorm swept uphill through the northern suburbs of the city.
15. Flash floods swept down all major streams including Rapid Creek.
16. The storm swept up the English channel.
17. The floodwater swept down the Whangaehu River picking up mud and boulders.
18. In 2003, Cyclone Zoe swept the outer islands of Tikopia, Anuta, and Fatutaka.
19. About once per century, killer tsunamis have swept this coastline.
20. Four hurricanes swept these 361 islands over two months.</td></tr>
<tr><td>3</td><td colspan="4">Lexical domain: MOVEMENT
Frame: to_move_forcefully
Frame definition: [NATURAL DISASTER] moves quickly and forcefully in a specified direction [PATH].
Note: The NATURAL DISASTER or ATMOSPHERIC AGENT is usually a wind event (e.g. hurricane, tornado).</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="5">BLAST$_2$</td><td>Theme</td><td>Path</td></tr>
<tr><td></td><td>Macrorole</td><td>Undergoer</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td>natural disaster [wind], atmospheric agent [wind]</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td>tornado, hurricane</td><td>its way, southern Missouri, fields</td></tr>
<tr><td></td><td>Phrase type</td><td>NP</td><td>PP (across, through)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="3">1. A massive tornado blasted its way across southwestern Missouri yesterday.
2. He filmed the drama as the tornado blasted through fields, trees and buildings in its path, spewing debris skyward.
3. A hurricane blasted through southern Ohio and caused a blackout throughout most of the city.
4. A powerful typhoon blasted across Japan yesterday, killing at least 31 people, causing deadly mudslides and flash floods.
5. Bitterly cold air and high winds blasted across the northern Plains on Sunday.</td></tr>
<tr><td>4</td><td colspan="4">Lexical domain: MOVEMENT
Frame: to_move_forcefully.
Frame definition: [NATURAL DISASTER] moves suddenly and forcefully in a specified direction [PATH].
Note: The NATURAL DISASTER is usually a wind event.</td></tr>
<tr><td></td><td>Semantic role</td><td rowspan="3">BURST$_1$</td><td>Theme</td><td>Path</td></tr>
<tr><td></td><td>Macrorole</td><td>Undergoer</td><td></td></tr>
<tr><td></td><td>Labels</td><td>natural disaster</td><td>_____</td></tr>
</table>

			[wind]	
	Linguistic realizations		tsunami, hurricane, tornado	northern coast of Japan, us, the city's levees, the city
	Phrase type		NP	PP (across, upon, through)
	Usage examples	1. A tsunami burst across the northern coast of Japan. 2. The hurricane burst through the city's levees and left much of the city flooded and without power. 3. The tornado burst upon the city, dashing diagonally through it.		

Table 125. Template of the frame *to move forcefully* (English)

The template of *to move forcefully* is displayed in Table 126. As shown, it has two arguments: (i) NATURAL DISASTER, ATMOSPHERIC AGENT or WATER AGENT, mostly associated with water and wind events, which has the role of THEME and the macrorole of UNDERGOER; (ii) PATH, lexicalized by prepositional phrases.

Frame: to move forcefully			
Semantic role	Theme	*surge* *sweep1* *blast2* *burst1*	Path
Macrorole	Undergoer		_____
Conceptual class	natural disaster, atmospheric agent, water agent		_____
Phrase type	NP		PP (across, along, from…to, in, into, out, over, through, to, up, upon)

Table 126. Template of the frame *to move forcefully*

To move slowly

The frame *to move slowly* is only represented in our corpus by the verb *drift* (Table 127). In line with the previous frame, it involves a NATURAL DISASTER, mostly associated with water and wind events and a second argument indicating PATH. The notion of *slowly* is part of the meaning of the verb.

1	Lexical domain: MOVEMENT Frame: to_move_slowly Frame definition: [NATURAL DISASTER] moves slowly in an underspecified direction [PATH]. Note: The NATURAL DISASTER is usually a water or wind event.			
	Semantic role	DRIFT	Theme	Path
	Macrorole		Undergoer	Undergoer
	Labels		natural disaster [water, wind]	_____
	Linguistic realizations		hurricane, tornado, storm	Honduran coastline, the west-southwest into the central South Atlantic

	Phrase type		NP	PP (along, toward, southward, into)
	Usage examples	1. The hurricane drifted southward along the Honduran coastline. 2. The tornado drifted southward as it moved eastward. 3. This storm lasted about five days and drifted toward the west-southwest into the central South Atlantic.		

Table 127. Analysis of the frame *to move slowly* (English)

Table 128 shows the template of *to move slowly*:

Frame: to move slowly			
Semantic role	Natural force	*drift*	Path
Macrorole	Undergoer		_____
Conceptual class	natural disaster [water, wind]		_____
Phrase type	NP		PP (across, along, from…to, in, into, out, over, through, to, up, upon)

Table 128. Template of the frame *to move slowly*

To (cause) to move vertically

The frame *to (cause) to move vertically* includes verbs with the meaning of moving or the causative meaning of *to cause to move*. In English, 'shake' belongs to both dimensions (see Table 129), whereas in Spanish, one dimension or the other is activated, depending on the verb. In this sense, 'temblar' conveys motion, whereas 'agitar' and 'remover', causation of movement (see Table 130):

1	Lexical domain: MOVEMENT Frame: to_move_vertically Frame definition: [CONSTRUCTION] or [AREA] moves up and down in quick short movements. Note: It can include SITUATION/ EXPERIENCE. LOCATION, TIME, or MANNER can also be specified.				
	Semantic role	SHAKE	Patient	Situation/Exp.	Manner
	Macrorole		Undergoer	_____	_____
	Labels		construction, area	_____	_____
	Linguistic realizations		building, Haiti	quake, earthquake	violently
	Phrase type		NP	PP (in, during)	AVP
	Usage examples	1. Building shakes violently in 7.5 magnitude quake. 2. Haiti shakes again. 3. Building shakes during Japan earthquake.			
	Lexical domain: MOVEMENT Frame: to_cause_to_move_vertically Frame definition: [NATURAL DISASTER] cause an [AREA] or [CONSTRUCTION] to move up and down in quick short movements. Note: The NATURAL DISASTER is usually an earth event (i.e. earthquake).				
	Semantic role	SHAKE	Natural force	Patient	

	Macrorole		Actor	Undergoer
	Labels		natural disaster [earth]	area, construction
	Linguistic realizations		earthquake, tsunami	towns, Japan, Costa Rica, region, building
	Phrase type		NP	NP
	Usage examples	1. A magnitude 6.0 earthquake shook several small towns in northeast Italy Sunday, killing four people. 2. A strong 6.8-magnitude earthquake shook northeastern Japan on Wednesday. 3. Powerful, magnitude-7.6 earthquake shook Costa Rica and a wide swath of Central America on Wednesday, collapsing some houses, blocking highways and causing panic. 4. It has only been 11 months after this region was shaken by the tsunami, following the Tohoku earthquake. 5. The information provided by the broadcaster NHK said the earthquake shook a building in the capital, but with no victims or other consequences.		

Table 129. Analysis of the frame *to (cause) to move vertically* (English)

1	Lexical domain: MOVEMENT Frame: to_cause_to_move_vertically Frame definition: [NATURAL DISASTER] cause an [AREA] or [CONSTRUCTION] to move up and down in quick short movements. Note: The NATURAL DISASTER is usually an earth event (i.e. earthquake).				
	Semantic role	AGITAR	Natural force	Patient	
	Macrorole		Actor	Undergoer	
	Labels		natural disaster [earth]	area, construction	
	Linguistic realizations		terremoto, tsunami	edificaciones, ciudad, localidad Pacífico, Sumatra, edificios	
	Phrase type		NP	NP	
	Usage examples	1. El terremoto agitó edificaciones y desplomó vehículos fuera de los puentes hacia el agua debajo. 2. La mañana del 18 de abril de 1906 un terremoto agitó la ciudad de San Francisco. 3. El pasado 15 de julio un leve terremoto agitó la localidad de Ciempozuelos. 4. Un tsunami agita el Pacífico y ya se han evacuado varias zonas costeras. 5. El seísmo agitó Sumatra derribando numerosos edificios, entre ellos hospitales y colegios. 6. El terremoto de magnitud 7 agitó el miércoles por la tarde los edificios en la capital, Yakarta, y derrumbó casas en localidades cercanas.			
2	Lexical domain: MOVEMENT Frame: to_move_vertically Frame definition: [CONSTRUCTION] or [AREA] moves up and down in quick short movements. Note: SITUATION/EXPERIENCE can also be included.				
	Semantic role	TEMBLAR	Patient	Sit./Exp.	Manner
	Macrorole		Undergoer	_____	_____
	Labels		construction, area	_____	_____
	Linguistic realizations		tierra, Indonesia, edificios, centro	seísmo, terremoto	ligeramente

			fincanciero		
	Phrase type		NP	PP (con, por, debido a)	AVP
	Usage examples	1. La tierra tembló en el sur de España. 2. La tierra tiembla dos veces ayer en Galicia. 3. El centro financiero de la capital indonesia tiembla con un seísmo de 6,4 grados. 4. Indonesia tiembla por un fuerte seísmo. 5. Los edificios temblaron ligeramente debido al terremoto.			
3	Lexical domain: MOVEMENT Frame: to_cause_to_move_vertically Frame definition: [NATURAL DISASTER] cause an [AREA] or [CONSTRUCTION] to move up and down in quick short movements. Note: The NATURAL DISASTER is usually an earth event (i.e. earthquake).				
	Semantic role	REMOVER	Natural force	Patient	
	Macrorole		Actor	Undergoer	
	Labels		natural disaster [earth]	area, construction	
	Linguistic realizations		terremoto, maremoto	Haití, cimientos, suelo, fondo marino	
	Phrase type		NP	NP	
	Usage examples	1. En enero, un terremoto removió el sur de Haití, derrumbó desde el palacio de gobierno hasta casitas en barrios marginados. 2. De 8.8 grados en la escala de Richter, el terremoto removió los cimientos de la capital y del centro y el sur del país. 3. El 12 de enero de este año un seísmo removió parte del suelo haitiano. 4. El trágico maremoto removió el fondo marino y permitió hallar un poblado que había sido devorado por el océano hace 1.200 años.			

Table 130. Analysis of the frame *to (cause) to move vertically* (Spanish)

Although the meaning conveyed by both constructions is the same, it appears in two different templates because of the difference in argument structure. As such, the frame *to move vertically* has one argument that is a CONSTRUCTION or an AREA with the role of PATIENT and macrorole of UNDERGOER. Additionally the SITUATION/EXPERIENCE can also be included in the sentence (e.g. 'Building shakes **during Japan earthquake**'):

Frame: to move vertically		
Semantic role	Patient	*shake* *temblar*
Macrorole	Undergoer	
Conceptual class	area, construction	
Phrase type	NP	

Table 131. Template of the frame *to move vertically*

The causative counterpart of the frame is instantiated by a NATURAL DISASTER associated with earth events (e.g. earthquake) which acts as a NATURAL FORCE and

ACTOR. This NATURAL FORCE causes an AREA or CONSTRUCTION, which is the PATIENT or UNDERGOER to move:

Frame: to cause to move vertically			
Semantic role	Natural force	*shake* *remover* *agitar*	Patient
Macrorole	Actor		Undergoer
Conceptual class	natural disaster [earth]		area, construction
Phrase type	NP		NP

Table 132. Template of the frame *to cause to move vertically*

To move horizontally over a large space

The frame *to move horizontally over a large space* is characterized by a THEME that can be a NATURAL DISASTER, PART_OF_NATURAL DISASTER (e.g. 'wave'), ATMOSPHERIC AGENT, WATER AGENT, ATMOSPHERIC CONDITION, AREA, MATERIAL ENTITY, LANDFORM, or ENERGY. This frame includes English verbs such as 'spread, 'extend', and 'propagate' (Table 133), and Spanish verbs such as 'extenderse', and 'propagarse' (Table 134):

1	Lexical domain: MOVEMENT Frame: to_move_horizontally_over_a_large_space Frame definition: [THEME] continues to move in a certain direction [PATH] over a large space. Note: The THEME is usually a natural disaster, atmospheric agent, water agent, atmospheric condition, area, material entity, landform, or energy.					
	Semantic role	SPREAD$_1$	Theme	Path	Origin	Manner
	Macrorole		Undergoer	_____	_____	_____
	Labels		natural disaster, atmospheric agent, water agent, atmospheric condition, area, material entity, landform, energy	_____	_____	_____
	Linguistic realizations		wave, snow, flood waters, fire, storm, wave energy , waves, tsunami waves, flows, fine ash, pyroclastic deposits,	out, outwards, crest of a wave, the surface, the entire valley of Bucheben, the land, southward, long distance, up, surrounding villages, a depth	epicenter, Santorini	quickly

			sediments	of 250m		
	Morpholo-gy		NP	PP (out, along, outwards, over, across, up, into)	PP (from)	AVP
	Usage examples	2. The wave will spread out to affect the entire coastline inside a harbor. 3. The tendency for wave energy to spread out along the crest of a wave. 4. These waves do not transmit through the mantle or core, but spread slowly outwards from the epicenter along the surface. 5. The initial tsunami waves may have been 60 m in height as they spread out from Santorini. 6. The 300 000 m³ of snow spread over the entire valley of Bucheben. 7. Generally though, as flood waters spread out across the land, they slow down. 8. Fires in Sweden spread quickly under strong wind conditions. 9. As the drought continued into 1971, large dust storms spread southward. 10. From these locations, the flows spread across the interfluve area. 11. Excess sediment must move offshore from the beach, to be spread over a long distance. 13. Fine ash spread over an area of 106 km2. 14. About 18–21 km3 of pyroclastic deposits spread out over 300 km2 to an average depth of 40 m. 15. Pyroclastic flows exceeded 5–6 km3 and spread up to 16 km from the volcano over an area of 400 km2. 16. A cloud of carbon dioxide gas bubbled up from Lake Nyos in Cameroon, spread out into the surrounding villages and killed 1,700 people. 17. The pyroclastic flow from this eruption flooded into a valley 20 km long, leaving an estimated 11 km3 of hot, fused ash spread to a depth of 250 m.				

2 Lexical domain: MOVEMENT
Frame: to_move_horizontally_over_a_large_space
Frame definition: [THEME] continues to move in a certain direction [PATH] over a largerspace.
Note: The THEME is usually natural disaster, atmospheric agent, water agent, atmospheric condition, area, material entity, landform.

Semantic role	EXTEND$_1$	Theme	Path
Macrorole		Undergoer	_____
Labels		natural disaster, atmospheric agent, water agent, atmospheric condition, area, material entity, landform	_____
Linguistic realizations		floods, flooded area, area, snow, wind, fire, landslide, storm conditions, mountains	Serbia and Romania, regions, over 600km2, Portugal to Spain, 1500 km, up to western Mindanao
Phrase type		NP	PP (into, over, from…to, along, across)
Usage examples	1. Floods also extended into Serbia and Romania where they caused the largest human toll associated with this event. 2. The flooded area may extend over regions of more than one country. 3. Droughts affected the area extending from Portugal and Spain to the Czech Republic.		

		4. The Mt St Helens pyroclastic flows extended over 600 km2. 5 These collapses have triggered landslides extending along the Sea. 7. Heavy snow extended across 1500 km, from New York to Chicago. 8. Tropical Storm Marco had gale force wind that only extended out 19 km [12 miles] from its center of circulation. 9. The extent of heavy rainfall covered a wide area, extending up to western Mindano. 12. Storm conditions extended along the coast from North Carolina to Nova Scotia. 13. The fire extended along the coast for several kilometers outside the area. 14. The Zagros Mountains extend to the south-east from Turkey and Azerbaijan along the country's western border to the Persian Gulf. 15. The Rocky Mountains extend from Alaska to the US-Mexico border.

3 Lexical domain: MOVEMENT
Frame: to_move_in_horizontally_over_a_large_space
Frame definition: [THEME] continues to move in a certain direction [PATH], over a large space.
Note: The THEME is usually a natural disaster, part of natural disaster (wave), material or water agent. It can also include manner.

Semantic role	PROPAGATE	Theme	Path	Manner
Macrorole		Undergoer	_____	_____
Labels		natural disaster, part of natural disaster (wave), material, water agent	_____	_____
Linguistic realizations		wave, water, flow, tsunami, earthquake, rupture, fire	away, offshore, alongshore, west to east, outwards and upwards, Pacific Ocean, further inland, shallow water, fault surface, south-western direction	faster, velocities, at speeds
Morphology		NP	PP (away, from, outwards, upwards, into, along, in)	AVP

Usage examples

1. A nuclear explosive produces a shock wave, or pulse of high pressure, that propagates away from the site of the explosion somewhat faster.
3. On the west coast, the reflected waves propagate offshore as well as alongshore.
4. Rossby waves move a few hundred kilometres an hour faster than the Earth rotates and thus appear to propagate from west to east.
5. A drawback occurs because the water propagates outwards with the trough of the wave at its front.
6. The fireball expands as a powerful blast wave that propagates outward and upwards.
7. The Antarctic earthquake was not conducive to tsunami and no major wave propagated into the Pacific Ocean.
8. This allowed the tsunami flows to propagate further inland and collide with the flow from the north.
9. As tsunamis propagate into shallow water, the forces acting on the water

		column must be considered to accurately model the inundation. 10. Once the rupture has initiated it begins to propagate along the fault surface. 11. Earthquake ruptures typically propagate at velocities that are in the range 70–90% of the S-wave velocity. 12. A small subset of earthquake ruptures appear to have propagated at speeds greater than the S-wave velocity. 13. Fires can move forward or 'propagate' faster than this, creating a firestorm that can travel at 60 km per hour. 14. The fire propagated mainly in south-western direction

Table 133. Analysis of the frame *to move horizontally over a large space* (English)

1	Lexical domain: MOVEMENT Frame: to_move_horizontally_over_a_large_space Frame definition: [THEME] continues to move in a certain direction [PATH], covering a large space. Note: The THEME is usually a natural disaster, atmospheric agent, water agent, atmospheric condition, area, material entity, landform			
	Semantic role	EXTENDERSE	Theme	Path
	Macrorole		Undergoer	_____
	Labels		natural disaster, atmospheric agent, water agent, atmospheric condition, area, material entity, landform	_____
	Linguistic realizations		inundaciones, fuego, área, avalanchas, magma, temporal de nieve	estado de Plateau, mayor parte de la ciudad, zona oeste, miles de kilómetros, superficie
	Phrase type		NP	PP (por, a lo largo de, cerca de, entre…y)
	Usage examples	1. Más de una treintena de aldeas en el centro de Nigeria quedaron sumergidas después de que las inundaciones se extendieran por el estado de Plateau. 2. El fuego se extendió rápidamente por la mayor parte de la ciudad. 3. El área afectada se extendió por toda la zona oeste de la ciudad. 4. Podemos hacernos idea de lo que pueden implicar megadeslizamientos cuyas avalanchas se extienden a lo largo de miles de km2. 5. Cuando las condiciones de enfriamiento son más rápidas porque el magma se extiende cerca de la superficie, a los materiales no le da tiempo de configurarse. 6. El fuerte temporal de nieve se extiende por toda la Península. 7. El Sistema de los Himalayas se extiende entre Afganistán y el noroeste de Pakistán.		
2	Lexical domain: MOVEMENT Frame: to_move_ horizontally_over_a_large_space Frame definition: [THEME] continues to move in a certain direction [PATH], over large space. Note: The THEME is usually a natural disaster, part of natural disaster (wave), material or water agent. It can also include manner.			
	Semantic role	PROPAGARSE	Theme	Path

	Macro-role		Undergoer	_____
	Labels		natural disaster, part of natural disaster (wave), material, water agent	_____
	Linguistic realiza-tions		incendio, inundaciones, ondas, tsunami	exterior, arriba y abajo, Mississipi, lugar a otro, oceáno Pacífico
	Morpho-logy		NP	PP (hacia, a lo largo, de, por, a través, de…a)
	Usage examples	1. El incendio se propagó hacia el exterior. 2. Las inundaciones se propagaron a lo largo del Mississippi. 3. Este tipo de ondas se propagan en el suelo hacia arriba y abajo, de un lugar a otro, de la misma manera que las olas marinas lo hacen al océano. 4. Cuando la energía de un terremoto llega al fondo del mar, se crean ondas que, al propagarse, llegan a la costa. 5. El tsunami se propagó por toda la cuenca del océano Pacífico causando daños y víctimas en Hawái, Oceanía y Japón. 6. El tsunami se propagó a través del Pacífico, matando a 61 personas en Hawái y 138 en Japón.		

Table 134. Analysis of the frame *to move horizontally over a large space* (Spanish)

The template of the frame *to move horizontally over a large space* is displayed in Table 135. The only compulsory argument is that of THEME, which can be instantiated by various kinds of semantic categories. In general, the THEME of all these verbs can be any kind of NATURAL DISASTER, ATMOSPHERIC AGENT, WATER AGENT, ATMOSPHERIC CONDITION, or AREA. In addition, depending on the verbs, the additional categories of MATERIAL ENTITY, LANDFORM, or ENERGY can also be instantiated. As such, 'propagate' and 'propagar' are the only ones whose THEME can be a part of natural disaster (e.g. 'wave') (e.g. 'A small subset of **earthquake ruptures** appear to have propagated at speeds greater than the S-wave velocity'). In contrast, 'extend' or 'extenderse' can also activate a LANDFORM, in which landform is described as occupying a larger space (e.g. '**The Zagros Mountains** extend to the south-east from Turkey and Azerbaijan along the country's western border to the Persian Gulf').

The verb 'spread' is also used with ENERGY ENTITIES (e.g 'The tendency for **wave energy** to spread out along the crest of a wave'). Finally, it is worth mentioning that this frame is particularly relevant for fire events. This is the reason why 'fire' comes in combination with all the verbs included here (e.g. '**Fires** in Sweden spread quickly under strong wind conditions', 'The **fire** extended along the coast for several kilometers outside the area', '**Fires** can move forward or 'propagate' faster than this, creating a firestorm that can travel at 60 km per hour').

The second optional argument is normally that of PATH. However sometimes PATH is implicit in the sentence (e.g. 'Earthquake ruptures typically propagate at velocities that are in the range 70–90% of the S-wave velocity'). Here MANNER is added, but PATH is implicit.

Frame: to move in a certain direction covering a large space			
Semantic role	Theme	*spread* *extend** *propagate*** *extenderse1** *propagarse***	Path
Macrorole	Undergoer		_____
Conceptual class	natural disaster, part of natural disaster (wave)**, atmospheric agent, water agent, atmospheric condition, area, material entity, landform*, energy		_____
Semantic role	NP		PP (across, along, from…to, in, into, out, over, through, to, up, upon) (English) PP (por, hacia, a lo largo de, a través de, a, entre…y) (Spanish)

Table 135. Template of the frame *to move horizontally over a large space*

To cause motion

The frame *to cause motion* is also extremely important in the EXTREME EVENT frame. Evidence of this is the number of verbs that belong to this frame: 'loft', 'release', 'eject', 'erupt', 'blast', 'emit', 'expel', 'blow out', 'spew', 'spit', for English (Table 136), and 'echar', 'expulsar', 'desprender', 'liberar', 'lanzar', 'despedir', 'soltar', 'emitir', 'emanar', and 'arrojar', for Spanish (Table 137). With the exception of 'loft', the rest of the verbs only appear with NATURAL DISASTERS involving volcanic events, and have two or three arguments. When there are two arguments, PATIENT is compulsory since it refers to the MATERIAL that is moved out or away (e.g. **'Very hot ash** can be blasted across the lands') and the NATURAL VOLCANIC DISASTER is implicit. However, in the majority of sentences, the NATURAL VOLCANIC DISASTER is also made explicit (e.g. '**The volcano** ejected sufficient gas and dust into the atmosphere to cool the climate'). In addition, very often, PATH is specified as well (e.g. 'At the time of the final eruption, ash was ejected into the **atmosphere'**). The movement encoded by these verbs is carried out by the MATERIAL which follows a certain PATH. SITUATION/EXPERIENCE can also be included in some usage contexts (e.g. 'It was found

that most of the gas was released **during the eruption**'), as well as origin (e.g. 'Volcanoes are essentially vents on the Earth's surface where molten rock, debris, and gases from **the planet's interior** are emitted') or destination (e.g. 'At the time of the final eruption, ash was ejected into the atmosphere **towards the north-east'**).

1	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a specified direction [PATH]. Note: SITUATION/EXPERIENCE can also be specified.					
	Semantic role	LOFT	Natural force	Patient	Path	Situation/ Experience
	Macrorole		Actor	Undergoer	_____	_____
	Labels		natural disaster	material entity	_____	_____
	Linguistic realization		volcano, eruption, tornado	gases, debris	stratosphere, air, up to 5 km	eruption of Mt. Tambora
	Phrase type		NP	NP	PP (into, up)	PP (in)
	Usage examples	1. Large, predominantly effusive eruptions, such as that at Laki (Iceland) in 1783 may also be capable of lofting sufficiently large volumes of gases into the stratosphere. 2. Large amounts of windborne debris will be lofted into the air. 3. Debris from a tornado can be lofted into the parent storm and carried a very long distance. 4. Numerous tornadoes have lofted (mainly light) debris many miles into the sky. 5. The eruption lofted clouds of ash up to 5 km (more than 16,000 feet) high. 6. The ash and volcanic bombs lofted into the atmosphere in the 1815 eruption of Mt. Tambora, in Indonesia, caused the destruction of nearly 100 villages in the neighboring area.				

2	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. It can also include SOURCE. SITUATION/EXPERIENCE can also be specified.						
	Semantic role	RELEASE	Natural force	Patient	Path	Sit./Exp.	Source
	Macrorole		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster [earth]	material entity	_____	_____	_____
	Linguistic realiza-tions		volcano, eruption	sulfur dioxide, cyanide, ash	atmosphere, stratosphere	eruption	bottom sedi-ments of Lake Nios
	Morpho-logy		NP	NP	PP (into)	PP (during)	PP (from)

	Usage examples	1. The eruption of El Chichon, Mexico, in 1982 released 40 000 tonnes of HCl into the stratosphere. 2. How much ash has the Icelandic volcano released into the atmosphere? 3. On 21 August 1986, these gases, plus sulfur dioxide and cyanide, were released by a landslide from the bottom sediments of Lake Nios. 4. For example, the Mt Pinatubo eruption of 15 June 1991 released 20.106 tonnes of sulfur dioxide in the space of a few weeks. 5. It was found that most of the gas (66%) was released during the eruption.

3 Lexical domain: MOVEMENT
Frame: to_cause_motion
Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH].
Note: The NATURAL DISASTER event is usually a volcanic event, but not an earthquake. Destination can also be included. SITUATION/EXPERIENCE can also be specified.

Semantic role	EJECT	Natural force	Patient	Path	Sit. /Exp.	Dest.
Macrorole		Actor	Undergoer	_____	_____	_____
Labels		natural disaster [earth]	material entity	_____	_____	_____
Linguistic realizations		volcano, eruption	lava, material, magma, pumice, ash	atmosphere	eruption	north east
Phrase type		NP	NP	PP (into)	PP (during)	PP (towards)

Usage examples	1. There are three types of basaltic lava that can be ejected from a volcano: pahoehoe, aa, and block lava. 2. This eruption ejected ten times the material of the 1902 Mt Pelée event. 3. The eruption of Mt Hekla, Iceland, in 1947 ejected 100 000 m3 s-1 of material, dropping ash as far away as Finland. 4. About 4.5 km3 of dense magma was ejected, producing 10 km3 of ash. 5. Large rafts of pumice ejected by the volcano blocked the Sunda Straits. 6. At the time of the final eruption, ash was ejected into the atmosphere towards the north-east. 7. By this time, over 70 million m3 of tephra had been ejected. 8. The volcano ejected sufficient gas and dust into the atmosphere to cool the climate. 9. It resulted in about 0.25 km3 of ejected tephra (fragmental material ejected by the volcano) and an ash plume to a height of about 9 km. 10. In addition to the ash, a large amount of sulfur dioxide gas and aerosols were ejected in the eruption.

4 Lexical domain: MOVEMENT
Frame: to_cause_motion
Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH].
Note: The NATURAL DISASTER is a volcanic event, but not an earthquake.

Semantic role	$ERUPT_2$	Natural force	Patient	Path	Manner
Macrorole		Actor	Undergoer	_____	_____
Labels		natural	material	_____	_____

			disaster [earth]	entity		
	Linguistic realizations		volcano, eruption	volcanic ash, gas	stratosphere, air	high
	Phrase type		NP	NP	PP (into)	AVP
	Usage examples	1. Volcanic ash that is light enough to be erupted high into the Earth's atmosphere may travel many kilometres before it falls back to ground. 2. Hot ash was erupted from the volcano and a lahar destroyed La Contiega village. 3. About 65 million years ago, an enormous amount of ash and gas was erupted into the atmosphere by volcanic activity in the Deccan Trapps, in India.				
5	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.						
	Semantic role	$BLAST_1$	Natural force	Patient	Path	Sit/Exp.	Manner
	Macrorole		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster [earth]	material entity	_____	_____	_____
	Linguistic realiza-tions		eruption, volcano	material, ash, rock and smoke	atmos-phere, lands	eruption	very quickly, away, as high as, away, with massive explosion
	Morpho-logy		NP	NP	PP (into, across, out)	PP (during)	AVP
	Usage examples	1. The eruption blasted 151 km3 of material into the atmosphere. 2. Very hot ash can be blasted across the lands. 3. The volcano blasted out rock and smoke at a foreboding rate. 4. The volcano blasted steam and ash as high as 15,000 metres. 5. The volcano blasted out with a massive explosion 6. The volcano blasted away with a massive explosion. 7. 150 km3 of rock and magma were blasted out during the eruption.					
6	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake. It can also include SOURCE. SITUATION/EXPERIENCE can also be specified.						
	Semantic role	EMIT	Natural force	Patient	Path	Sit./Exp.	Origin
	Macrorole		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster [earth]	material entity	_____	_____	_____
	Linguistic		volcano,	hot gases,	atmos-	eruption,	planet's

	realizations		eruption	steam, ash cloud, lava and ash	phere	volcanic eruption	interior
	Phrase type		NP	NP	PP (to/ into)	PP (in, during)	PP (from)
	Usage examples	1. The Vatnajökull glacier caps the 40 km2 Grimsvötn caldera, which continually emits hot gases. 2. The volcano started emitting steam. 3. The ash cloud emitted by the recent eruption of the Eyjafjallajökull volcano caused the grounding of air traffic across northern and central Europe. 4. By the end of May, the volcano had stopped emitting lava and ash but was emitting sulphurous gases. 5. Conversely, carbon dioxide emitted during volcanic eruptions is a greenhouse gas and contributes to global warming. 6. Volcanoes are essentially vents on the Earth's surface where molten rock, debris, and gases from the planet's interior are emitted. 7. Layers of ash and lava were emitted by the volcano 8. Teide's summit is emitting 400 tons m2 day-1 of CO2 to the atmosphere. 9. The acid is mainly sulphuric acid produced as an aerosol from sulphur dioxide emitted in the eruption.					
7	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. It can also include SOURCE. SITUATION/EXPERIENCE can also be specified.						
	Semantic role	EXPEL	Natural force	Patient	Path	Sit./Exp.	Origin
	Macrorole		Actor	Undergoer	____	_____	_____
	Labels		natural disaster [earth]	material entity	____	_____	_____
	Linguistic realizations		volcano, eruption	water, toxic gases, lava, rock	air	eruption, volcanic events	crater, seabed
	Phrase type		NP	NP	PP (into)	PP (in, during)	PP (from)
	Usage examples	1. The 1919 eruption of Mt Kelat on Java expelled water from a crater lake, covering 200 km2 of farmland. 2. Water was simply expelled from the seabed by the pyroclastic flows. 3. It began to flow down slope at great velocity suspended by the gases expelled in the eruption. 4. These latter events suggest that water may be expelled from the water table by escaping gases during some volcanic events. 5. Some landslides include toxic gases from deep in the Earth expelled by volcanoes. 6. Eruptions can be explosive, expelling lava, rocks and ash into the air. 7. The rate at which lava is expelled also controls the speed of movement.					
8	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and						

	forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	BLOW OUT$_2$	Natural force	Patient	Path	Sit/Exp.
	Macrorole		Actor	Undergoer	_____	_____
	Labels		natural disaster [earth]	material entity	_____	_____
	Linguistic realizations		volcano	lava, volcanic ejecta, solid rock	atmosphere	eruption
	Phrase type		NP	NP	PP (into)	PP (during)
	Usage examples	1. A type of lava that consists of angular, bouldersized pieces of solid rock blown out of an erupting volcano. 2. Pyroclastic material is volcanic ejecta violently blown out of the volcano into the atmosphere during an eruption. 3. They corresponded to boulder-sized pieces of liquid lava blown out of an erupting volcano. 4. 9–10 km3 of solid rock was blown out of the volcano.				
9	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. It can also include SOURCE. SITUATION/ EXPERIENCE can also be specified.						
	Semantic role	SPEW	Natural force	Patient	Path	Situation/ Experience	Source
	Macrorole		Actor	Undergoer	_____	_____	_____
	Labels		natural disaster [earth]	material entity	_____	_____	_____
	Linguistic realizations		volcano, Mount Nyira-gongo	lava flows, lava, gases, material	several directions	eruption	crater
	Phrase type		NP	NP	PP (in)	PP (during)	PP (from)
	Usage examples	1. Mount Nyiragongo erupted with lava flows spewing in several directions. 2. What remained after the material spewed during the eruption? 3. The most common perception of a volcano is of a conical mountain, spewing lava and poisonous gases from a crater at its summit.					
10	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake.					
	Semantic role	SPIT	Natural force	Patient	Path	
	Macrorole		Actor	Undergoer	_____	
	Labels		natural disaster [earth]	material entity	_____	
	Linguistic		volcano,	rock, ash,	more than 1600 km,	

	realizations		eruption, fire	gas, embers	mile, sea, out, over 1.5 km, air
	Phrase type		NP	NP	PP (out, across, up to, into, for)
	Usage examples	1. Iceland volcano spat out 1000-ton rock. 2. The volcano spat rocks for over 1.5 km (0.93 mi). 3. That volcano spat ash across more than 1600 square kilometers. 4. On Saturday morning 3 explosions were recorded and the volcano spat out ash and gas up to a mile into the air. 5. The eruption spat 1m wide rocks into the sea. 6. The crackling fire spat embers up into the air.			

Table 136. Analsyis of the frame *to_cause_motion* (English)

1	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	ECHAR	Natural force	Patient	Path	Sit./Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		volcán, erupción	gases, cenizas, arena, piedras	nivel del mar, tres mil cuatrocientos metros	_____
	Phrase type		NP	NP	PP (sobre, a)	_____
	Usage examples	1. En la víspera, el volcán echó gases y cenizas a tres mil cuatrocientos metros sobre el nivel del mar. 2. El volcán echó piedras que cayeron sobre las casas. 3. La erupción echó a la atmósfera toneladas de arena y ceniza.				
2	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	EXPULSAR	Natural force	Patient	Path	Sit. /Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		erupciones fisurales, ondas expansivas, erupciones volcánicas	lava basáltica, fragmentos, ceniza volcánica, dióxido de azufre, ácido	atmósfera	erupción

				sulfúrico, ácido fluorhídrico, dióxido de azufre, magma, ceniza		
	Phrase type		NP	NP	PP (a, hasta)	PP (durante)
	Usage examples	1. Las erupciones fisurales expulsaron lava basáltica muy líquida. 2. Las ondas expansivas machacaron rápidamente las rocas de debajo y llenaron todas las grietas, fundieron parcialmente las rocas, produjeron la brecha y expulsaron fragmentos a gran altura de la atmósfera. 3. Una erupción del monte Pinatubo en 1991 expulsó grandes cantidades de ceniza volcánica y dióxido de azufre hasta unos 30 kilómetros en la atmósfera. 4. Se trató de erupciones volcánicas extraordinariamente grandes que expulsaron un volumen enorme de ceniza volcánica a la atmósfera, cambiando el clima. 5. La erupción del lago Toba sumió a la Tierra en un invierno volcánico, expulsando ácido sulfúrico a la atmósfera y originando así la denominada Edad de Hielo. 6. Está erupción expulsó unos 14 kilómetros cúbicos de lava basáltica y nubes tóxicas de ácido fluorhídrico y dióxido de azufre que acabaron con el 20% de la población islandesa y más del 50% del ganado de la isla. 7. Durante la erupción, el volcán expulsó de forma poco violenta muchos fragmentos de magma que se acumularon hasta formar su cono volcánico de 160 metros de altura. 8. El volcán expulsa ceniza a cuatro kilómetros de altura en Rusia. 9. El volcán expulsa fragmentos viscosos de lava durante la erupción.				
3	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	DESPREN-DER	Natural force	Patient	Path	Sit/Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		volcán, cráter del volcán, volcán en erupción	columna de humo, gases de azufre, lava, CO2, gas explosivo	exterior	erupciones
	Phrase type		NP	NP	PP (a)	PP (durante)
	Usage examples	1. El volcán desprendió una columna de humo de 15.000 metros de alto. 2. En lo alto de la colina resulta complicado respirar ya que el cráter del volcán desprende gases de azufre. 3. Durante las erupciones, el volcán desprende lava al exterior formando coladas. 4. Hay que tener el cuenta que un volcán en erupción desprende más CO2				

<table>
<tr><td></td><td></td><td colspan="5">que todas las fábricas y coches del planeta.
5. Las erupciones hawaianas, cuando entran en erupción desprenden un gas explosivo.</td></tr>
<tr><td>4</td><td colspan="6">Lexical domain: MOVEMENT
Frame: to_cause_motion
Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH].
Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.</td></tr>
<tr><td></td><td>Semantic role</td><td>LIBERAR</td><td>Natural force</td><td>Patient</td><td>Path</td><td>Sit./Exp.</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Actor</td><td>Undergoer</td><td>____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster [earth]</td><td>material entity</td><td>____</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>erupción del volcán, volcán, erupción</td><td>nube de ceniza volcánica, cenizas, magma, gases, pluma de ceniza volcánica</td><td>Europa, océano</td><td>erupción</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>NP</td><td>PP (sobre, a)</td><td>PP (en)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="5">1. La erupción del volcán liberó una nube de ceniza volcánica sobre casi toda Europa.
2. Sólo el 23 de mayo el volcán liberó casi dos mil toneladas de cenizas por segundo y un total de 120 millones de toneladas durante las primeras 48 horas.
3. El volcán liberó magma y gases que afectaron a la fauna marina.
4. La erupción volcánica submarina de El Hierro liberó magma y gases al océano.
5. Este es el mapa de la posición aproximada de la pluma de ceniza volcánica durante la erupción del Eyjafjallajökull, Islandia. 4 de mayo.</td></tr>
<tr><td>5</td><td colspan="6">Lexical domain: MOVEMENT
Frame: to_cause_motion
Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH].
Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.</td></tr>
<tr><td></td><td>Semantic role</td><td>LANZAR</td><td>Natural force</td><td>Patient</td><td>Path</td><td>Sit. /Exp.</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Actor</td><td>Undergoer</td><td>____</td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster [earth]</td><td>material entity</td><td>____</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>volcán, erupción</td><td>emisiones de ceniza, gases, cenizas, lava, cenizas volcánicas, materiales, columna de vapor, lava incandescente, pluma de gases</td><td>atmósfera, cráter, altura, ladera</td><td>_____</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>NP</td><td>PP (a, por)</td><td>_____</td></tr>
</table>

	Usage examples	1. El pasado 8 de septiembre de este año 2012, el volcán lanzó tres emisiones de cenizas que afectaron a la visibilidad. 2. En la víspera, el volcán lanzó gases y cenizas a la atmósfera. 3. La erupción lanzó enormes cantidades de lava y cenizas volcánicas que cayeron sobre el principal poblado de Catania y otras aldeas. 4. La erupción lanzó grandes cantidades de materiales a la atmósfera y el derrumbe de la isla originó grandes tsunamis. 5. Según informes locales, la erupción lanzó una columna de vapor y ceniza volcánica a una altitud de 6 y 7 kilómetros de altura. 6. El Instituto Nacional de Geofísica ha informado que anoche el volcán lanzó lava incandescente a 1 kilómetro por encima de su cráter y una pluma de gases y cenizas volcánicos. 7. El volcán lanzó espesas nubes de ceniza a casi tres kilómetros de altura y ríos de lava por las laderas.				
6	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	DESPEDIR	Natural force	Patient	Path	Sit./Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		erupción, volcán	piroclastos, lava, cenizas, material incandescente, gas, roca	atmósfera	erupción
	Phrase type			NP	PP (a)	PP (en)
	Usage examples	1. La erupción despidió a la atmósfera muchos más piroclastos y material incandescente que el Krakatoa. 2. El volcán despidió gran cantidad de lava y provocó una nube impresionante de ceniza que, por la acción de los vientos del sudoeste, cubrió la provincia de Río Negro. 3. El volcán despidió cenizas y lava que afectaron a los pueblos cercanos. 4. A lo largo del mes de julio el volcán despidió lava en cuatro ocasiones. 5. El inestable volcán despidió el viernes cerca de 50 millones de metros cúbicos de gas, rocas y cenizas en la erupción.				
7	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	SOLTAR	Natural force	Patient	Path	Sit./Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		volcán, erupción	rocas volcánicas, material volcánico,	aires	erupción

				materiales, cenizas		
	Phrase type		NP	NP	PP (por)	PP (en)
	Usage examples	1. Al entrar en erupción el volcán soltó gran cantidad de rocas volcánicas. 2. En aquella ocasión, el volcán soltó ríos de material volcánico que arrasaron con todo a su paso. 3. La erupción soltó sobre la Isla 60.500.000 metros cúbicos de materiales. 4. Cuando el volcán entró en erupción soltó por los aires miles de toneladas de cenizas.				
8	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is usually a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	EMITIR	Natural force	Patient	Path	Sit./Exp.
	Macrorole		Actor	Undergoer	_____	_____
	Labels		natural disaster [earth]	material entity	_____	_____
	Linguistic realizations		explosion, fuego, volcán, erupción,	aerosoles, humo, piroclastos humeantes, CO2, columna de ceniza, material, vapor de agua, derivados del nitrógeno	estratosfera, atmósfera	erupción
	Phrase type		NP	NP	PP (a)	PP (durante)
	Usage examples	1. La explosión emitió a la estratosfera gran cantidad de aerosoles. 2. El fuego emitió tanto humo que Indonesia, Malasia y Singapur se cubrieron de lo que de manera eufemística, se denominó «neblina». 3. El volcán ha vuelto a emitir piroclastos humeantes de grandes dimensiones, así como chorros de vapor, durante la mañana de este domingo. 4. El Involcan estima que el volcán emite 332 toneladas diarias de CO2 a la atmósfera. 5. El volcán emitó una columna de ceniza de 2.5 kilómetros de altura durante la erupción. 6. La erupción emitió tanto material a la atmósfera que el clima se enfrió y hubo una Edad Glacial, que a punto estuvo de terminar con la especie humana. 7. La erupción emite aún grandes cantidades de ceniza a la atmósfera, y los expertos estiman que la situación de caos podría alargarse al menos dos días más. 8. Cuando un volcán entra en erupción emite grandes cantidades de vapor de agua, CO2 y derivados de nitrógeno.					
9	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH].					

	Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	EMANAR	Natural force	Patient	Path	Sit. /Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		volcán	ceniza, gases, arena, columna de vapor	altura	explosión
	Phrase type		NP	NP	PP (a)	PP (durante)
	Usage examples	1. El volcán emanó aproximadamente 1.600.000 metros cúbicos de ceniza que se repartieron en cinco provincias. 2. Desde las doce de la noche el volcán emanó gases, ceniza y arena. 3. Cerca de las 9:25 de la mañana, el volcán emanó una columna de vapor y un elevado contenido de ceniza. 4. Durante la explosión el volcán emanó gases a una altura de 600 metros. 5. El volcán Telica sigue emanando gases.				
10	Lexical domain: MOVEMENT Frame: to_cause_motion Frame definition: [NATURAL DISASTER] causes [MATERIAL] to move quickly and forcefully in a certain direction [PATH]. Note: The NATURAL DISASTER is a volcanic event, but not an earthquake. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	ARROJAR	Natural force	Patient	Path	Sit./Exp.
	Macrorole		Actor	Undergoer	____	_____
	Labels		natural disaster [earth]	material entity	____	_____
	Linguistic realizations		fisura, volcán, erupción	flujo de lava basáltica, piedras candentes, ceniza volcánica, material volcánico	atmósfera, cráter, ciudad de Pompeya	erupción del monte
	Phrase type		NP	NP	PP (a, alrededor, sobre)	PP (durante)
	Usage examples	1. En junio de 1944, una fisura que se había abierto en el volcán arrojó un flujo de lava basáltica. 2. El volcán arrojó piedras candentes a la atmósfera. 3. El volcán arrojó una enorme columna de cenizas a 5 km de altura. 4. El volcán arrojó rocas incandescentes alrededor del cráter. 5. Esta segunda erupción arrojó ceniza volcánica a la atmósfera, llegando a una altura de varios kilómetros y extendiéndose por un área de miles de kilómetros. 6. La erupción arrojó ceniza caliente, piedras y carbonilla sobre la ciudad de Pompeya, que quedó cubierta por una capa de 4 m de espesor. 7. Las inusuales formaciones de cono se formaron de ceniza volcánica y escombros arrojados durante la erupción del monte Sahand.				

Table 137. Analsyis of the frame *to cause motion* (Spanish)

The template underlying the frame *to cause motion* is in Table 138:

Frame: **to_cause_motion**				
Semantic role	Natural force	*loft* *release* *eject* *erupt2* *blast1* *emit* *expel* *blow out2* *spew* *spit* *echar* *expulsar* *desprender* *liberar* *lanzar* *despedir* *soltar* *emitir* *emanar* *arrojar*	Patient	Path
Macrorole	Actor		Undergoer	____
Conceptual class	natural disaster [earth]		material entity	____
Phrase type	NP		NP	PP (across, along, from…to, in, into, out, over, through, to, up, upon) (English) PP (por, hacia, a lo largo de, a través de, a, entre…y, sobre) (Spanish)

Table 138. Template of the frame *to cause motion*

5.4.3.5 Domain of POSITION

Within the domain of POSITION, the frames relevant in our study are those of *to put sth on top of or over sth* and *to put water over/in a space*

To put sth on top of or over sth

The frame *to put sth on top of or over sth* is defined as a NATURAL DISASTER which puts sth (the THEME) on top or over sth (PATIENT). Verbs in our corpus that instantiate this frame are 'cover', 'engulf', 'bury', for English (Table 139), and '(re)cubrir', 'sepultar', 'enterrar', 'envolver', for Spanish (Table 140).

<table>
<tr><td>1</td><td colspan="6">Lexical domain: POSITION
Frame: to_put_sth_on_top_of_or_over_sth
Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT].
Note: The PATIENT is usually a construction or area. Very often the THEME is implicit. SITUATION/EXPERIENCE can also be specified.</td></tr>
<tr><td></td><td>Semantic role</td><td>COVER</td><td>Natural force</td><td>Patient</td><td>Theme</td><td>Sit./Exp.</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Actor</td><td>Undergoer</td><td></td><td>_____</td></tr>
<tr><td></td><td>Labels</td><td></td><td>natural disaster or part of natural disaster</td><td>construction, area</td><td>material</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realizations</td><td></td><td>landslide, avalanche, floods</td><td>school, area, village</td><td>fire, snow</td><td>landslide</td></tr>
<tr><td></td><td>Phrase type</td><td></td><td>NP</td><td>NP</td><td>PP (with)</td><td>PP (in)</td></tr>
<tr><td></td><td>Usage examples</td><td colspan="5">1. The landslide covered an entire school.
2. The following morning at roughly 4:00 a.m., a third landslide covered the whole area.
3. An avalanche covered Sherin Nazam village in Badakhshan province in northeast Afghanistan Sunday night causing over 140 people to be considered missing, and 39 confirmed dead.
4. The recent December floods covered an unexpectedly large area, again submerging districts upon which council houses have been built.
5. Train tunnel was covered in a landslide.
6. The rate of lava coverage—not how recently lava covered an area—is the basis of long-term lava flow hazard.
7. The entire northern side was covered with fire.
8. Four people died as an avalanche covered a village in the northeastern Badakhshan province.
9. Avalanche covered the resort village with huge amounts of snow.</td></tr>
<tr><td>2</td><td colspan="6">Lexical domain: POSITION
Frame: to_put_sth_on_top_of_or_over_sth
Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT].
Note: The PATIENT is usually a construction, human being or area. Very often the THEME is implicit. SITUATION/EXPERIENCE can also be specified.</td></tr>
<tr><td></td><td>Semantic role</td><td>ENGULF</td><td>Natural force</td><td>Patient</td><td>Theme</td><td>Sit./Exp.</td></tr>
<tr><td></td><td>Macrorole</td><td></td><td>Actor</td><td>Undergoer</td><td></td><td>_____</td></tr>
<tr><td></td><td>Label</td><td></td><td>natural disaster</td><td>construction, human being, area</td><td>material</td><td>_____</td></tr>
<tr><td></td><td>Linguistic realize-tions</td><td></td><td>avalanche, tsunami, landslide, fire</td><td>home, building, snowboarder, Hawaiian beach</td><td>flames</td><td>avalanche</td></tr>
<tr><td></td><td>Morpho-logy</td><td></td><td>NP</td><td>NP</td><td>PP (in)</td><td>PP (in)</td></tr>
</table>

	Usage examples	1. Two avalanches swept down on the village yesterday from the 6,500ft heights surrounding it, and engulfed the homes of the 404 people who live there. 2. Snowboarder was engulfed in avalanche. 3. Kamaishi engulfed by tsunami after earthquake. 4. Buildings engulfed in flames at Sendai Airport, Japan. 5. Hawaiian beach engulfed by 1957 tsunami. 6. Rescuers search for victims after a landslide engulfed a building of a primary school at Zhenhe village of Yiliang county. 7. Emergency officials say four people are unaccounted for after a landslide engulfed four homes in southern British Columbia. 8. A fire engulfed part of a Chevron oil refinery in Richmond. 9. The house was quickly engulfed in flames.

3 Lexical domain: POSITION
Frame: to_put_sth_on_top_of_or_over_sth
Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT].
Note: The PATIENT is usually a human being, construction, or area. Very often the THEME is implicit.

	BURY	Natural force	Patient	Theme
Semantic role				
Macrorole		Actor	Undergoer	_____
Labels		natural disaster, material entity	human being, construction, area	Material
Linguistic realizations		landslide, avalanche, lahars, volcano, lava, eruption, snow, ash	people, worker, villages, hectares of forest, the town of Goma, houses	mud, ash and mud
Phrase type		NP, PP (in, under)	NP	PP (in, under)

Usage examples	1. Four people buried by a landslide caused by heavy rain, in the north of the country. 2. 23,000 people are buried in the mud that killed them. 3. This facilitates fast search and rescue of people buried under the snow. 4. Towns were buried in mud, and shallow lakes appeared in the deserts. 5. An avalanche of similar size at Goldau in Switzerland, in 1806, buried four villages and killed 457 people. 6. Paricutín in Mexico, in 1946, buried 2400 hectares of forest and agricultural land. 7. They rescued many people buried by snow avalanches. 8. Bulgaria landslide buries 32-year-old worker. 9. It generated a landslide that within minutes killed 18,000 people and buried several towns. 10. An avalanche in the Siachen glacier in the Himalaya mountains buried at least 124 Pakistani soldiers 11. The Tarawera eruption of 10 June 1886, in New Zealand, buried Maori villages under more than 2 m of ash and mud. 12. Landslides triggered by the earthquake buried houses and people.

Table 139. Analysis of the frame *to put sth on top of or over sth* (English)

1	Lexical domain: POSITION Frame: to_put_sth_on_top_of_or_over_sth Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT]. Note: The PATIENT is usually a construction, area or human being. Very often the THEME is implicit. SITUATION/EXPERIENCE can also be specified.					
	Semantic role	(RE)CUBRIR	Natural force	Patient	Theme	Sit./Exp.
	Macrorole		Actor	Undergoer		_____
	Labels		natural disaster or part of natural disaster	construction, area, human being	material	_____
	Linguistic realiza-tions		inundaciones, lava, flujos piroclásticos, deslizamiento de tierra, avalancha	escuela, casas, vehículos, escaladores, barrios, hectáreas, localidad	escom-bros, todo tipo de objetos	avalancha de piedras
	Morpho-logy		NP	NP	PP (de, con)	PP (por)
	Usage examples	1. Las inundaciones cubrieron fundamentalmente los barrios pobres de la ciudad. 2. La erupción Laki de Islandia, en 1783, generó una colada de lava que cubrió 500 km2. 3. Los flujos piroclásticos y flujos de lodo cubrieron grandes zonas alrededor del volcán. 4. El deslizamiento de tierra cubrió 80 hectáreas 5. Un desprendimiento en la provincia de Yunnan (China) cubrió una escuela de escombros y afectó a seis viviendas cercanas. 6. La avalancha cubrió doce casas y destruyó totalmente seis de ellas, donde se encontraban las 33 personas que perecieron. 7. Los escaladores fueron cubiertos por una avalancha de piedras. 8. Hay que aclarar sin embargo que muchos vehículos particulares quedaron prácticamente cubiertos por la inundación. 9. Las inundaciones cubrieron la localidad con todo tipo de objetos.				
2	Lexical domain: POSITION Frame: put_sth_on_top_of_or_over_sth Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT]. Note: The PATIENT is usually a construction, human being, or area.					
	Semantic role	SEPULTAR	Natural force	Patient	Theme	
	Macrorole		Actor	Undergoer	_____	
	Label		natural disaster or part of natural disaster	construction, human being, area	material	
	Linguistic realizations		erupción volcánica, flujo de lodo, deslizamiento, ventisca, coladas de barro, alud de tierra, avalancha, tsunami, terremoto, inundaciones	ciudad, casas, Armero, niños, tren, viviendas, personas, paquistaníes	nieve	
	Morpho-logy		NP	NP	PP (de)	

	Usage examples	1. La erupción volcánica del Nevado del Ruiz en 1985 sepultó la ciudad de Armero, provocando la muerte de unas 21 000 personas. 2. El flujo de lodo sepultó la mitad sur de la ciudad arrancando por completo los edificios. 3. Este deslizamiento sepultó, derrumbó o dañó casas en la base de la ladera. 4. La ventisca de nieve de Saskatchewan de 1947 duró diez días y sepultó un tren completo. 5. Los flujos de lodo sepultaron Armero, en Colombia. 6. Unas coladas de barro invadieron el valle del río Lagunilla y sepultaron la ciudad de Armero, con 24.000 muertos y decenas de heridos. 7. Un alud de tierra sepulta a 16 niños en China. 8. Avalancha sepulta a 135 paquistaníes. 9. Deslizamiento sepultó cuatro casas en la localidad de Chapinero. 10. Un terremoto sepulta a 80 personas en el norte de Afganistán. 11. Un tsunami sepulta la ciudad de Constitución tras el terremoto que sacudió Chile. 12. Un tornado sepulta a 20 personas en Japón. 13. Las fuertes lluvias e inundaciones sepultaron 17.300 viviendas rurales y obligaron a evacuar a 149.000 residentes de toda la provincia. 14. Vientos de más de 160 kilómetros/hora sepultaron de nieve ciudades enteras.

3	Lexical domain: POSITION Frame: to_put_sth_on_top_of_or_over_sth Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT]. Note: The PATIENT is usually a human being, construction, or area. Very often the THEME is implicit.				
	Semantic role	ENTERRAR	Natural force	Patient	Theme
	Macrorole		Actor	Undergoer	
	Labels		natural disaster, part of natural disaster	human being, area, construction	material
	Linguis-tic realizations		avalancha, terremoto, deslizamiento, erupción	escuela, edificios, personas, viviendas, ciudad	escombros
	Phrase type		NP	NP	PP (con)
	Usage examples	1. La avalancha enterró la escuela y otros dos edificios con 160.000 metros cúbicos de escombros. 2. Aunque el terremoto enterró a miles de personas, el New York Times informo que únicamente 132 de ellas fueron rescatadas de entre los escombros. 3. El deslizamiento enterró 29 viviendas y dañó otras 70 en cuatro aldeas de la zona montañosa de Bududa. 4. En el centro de Angra dos Reis, otro deslizamiento enterró varias casas. 5. Hace casi 2.000 años, la erupción de un volcán enterró una ciudad entera que se llamaba Pompeya.			

4	Lexical domain: POSITION Frame: to_put_sth_on_top_of_or_over_sth Frame definition: [NATURAL DISASTER] puts sth [THEME] on top or over sth [PATIENT]. Note: The PATIENT is usually a construction, human being or area.				
	Semantic role	ENVOLVER	Natural force	Patient	Theme
	Macrorole		Actor	Undergoer	_____
	Label		natural disaster	construction, human	material

				being, area	
	Linguistic realizations		tormenta de granizo y nieve, fuego, avalancha, ola, incendio	alpinista, hectáreas, viviendas, estudiantes, guía de montaña, casa, urbanización	llamas
	Phrase type		NP	NP	PP (en)
	Usage examples	1. Una tormenta de granizo y nieve envolvió a los tres alpinistas. 2. El fuego envolvió 100 hectáreas en Cali. 3. El fuego envolvió con rapidez dos de las viviendas y dañó considerablemente la tercera. 4. La avalancha envolvió a 15 estudiantes y al guía de montaña. 5. La ola del tsunami los envolvió y los lanzó "en círculos, como si estuviéramos dentro de un lavarropas." 6. Un enorme incendio envolvió varias casas. 7. El incendio envolvió en llamas la urbanización del Altury. 8. Se oyó una fuerte explosión en el interior y la vivienda fue envuelta en llamas.			

Table 140. Analysis of the frame *to put sth on top of or over sth* (Spanish)

As shown in the analysis of Tables 139 and 140, this frame is usually instantiated by two arguments: (i) a NATURAL DISASTER or PART OF NATURAL DISASTER, which is the NATURAL FORCE that is the ACTOR; and (ii) a CONSTRUCTION, HUMAN BEING or AREA which is the PATIENT, i.e. the area or container being filled or covered by the natural force. Sometimes the MATERIAL with which the PATIENT is covered and other times, it is left implicit. The template is the following:

Frame: to put sth on top of or over sth				
Semantic role	Natural force	*cover* *engulf* *bury* *(re)cubrir* *sepultar* *enterrar* *envolver*	Patient	Theme
Macrorole	Actor		Undergoer	_____
Conceptual class	natural disaster		construction, human being, area	material
Phrase type	NP		NP	PP (in, under, with) (English) PP (de, con, en) (Spanish)

Table 141. Template of the frame *to put sth on top of or over sth*

To put water over/in a space

Finally, the frame *to put water over/in a space* includes the verbs 'flood', 'inundate', 'swamp', in English, and 'inundar', 'anegar', in Spanish. The complete analysis of all these verbs in shown in Tables 142 and Table 143:

1	Lexical domain: POSITION Frame: to_put_water_over/in_a_space Frame definition: [NATURAL FORCE] puts a lot of water [THEME] over/in a [PATIENT]. Note: The PATIENT is usually an area, landform, water course or construction. Both the THEME and NATURAL FORCE can be implicit.				
	Semantic role	FLOOD	Natural force	Patient	Theme
	Macrorole		Actor	Undergoer	_____
	Labels		water agent, atmospheric agent, natural disaster [water], natural disaster [earth], material entity, water course	area, construction, landform, water course	material [water]
	Linguistic realizations		eruption, storm, rain, tropical cyclones, tsunami, pyroclastic flows, rivers, lakes, storm surges, precipitation events	land, fields, town, rivers, homes, area, buildings, residences	water
	Phrase type		NP	NP	PP (with)
	Usage examples	1. 12.5 million hectares of land were flooded. 2. Eruptions from Mauna Loa and Kilauea in Hawaii have consistently flooded agricultural land and destroyed many small villages. 3. The storm flooded land for hundreds of miles along Nicaragua's river border. 4. Fields growing subsistence crops were flooded with brackish water. 5. One such low caused heavy rains that flooded Gippsland on 23–24 June. 6. Within an hour the town was flooded and helicopters evacuated the 2500 residents. 7. Some volcanoes have had enormous magma chambers that emptied and flooded landscapes over vast distances in a short period. 8. Pyroclastic flows overran 1000 m high ridges around the crater and flooded an area of 15 000 km2. 9. 1000 levees collapsed, 55 people lost their lives, 48 000 homes in 75 towns were flooded, and 74 000 people were left homeless. 10. Towards the end of June a new series of precipitation events flooded Hull, Sheffield and Doncaster. 11. There were 246 fatalities, 137 000 buildings flooded, and 700 000 people made homeless as a result of this flood. 12. 29 lives were lost and 23,572 residences were flooded across Japan. 13. The St Lawrence Valley in eastern Canada was effectively flooded by the ocean during the waning phases of continental glaciation. 14. Three cattle are drowned and his fields are flooded, their crops destroyed. 15. The State Library was also flooded, damaging rare exhibits. 16. Overflowing rivers and lakes flooded several towns and cities and caused landslides across northern Italy. 17. Oxfordshire, Berkshire and some areas in Lincolnshire, already flooded in June, were again under water. 18. The tsunami flooded the Pacific coast of Japan. 19. Tropical cyclones or monsoonal troughs have in the past severely flooded all of these rivers, but never at the same time. 20. Usually, many smaller rivers in the same region are flooded. 21. Old and abandoned river channels far from the active river			

		channels were also heavily flooded. 22. This river flooded after the second rainstorm.

2 Lexical domain: POSITION
Frame: to_put_water_over/in_a_space
Frame definition: [NATURAL FORCE] puts a lot of water [THEME] over/in a [PATIENT].
Note: The PATIENT is usually an area, landform, water course or construction. The THEME is often implicit.

Semantic role	INUNDATE	Natural force	Patient	Theme
Macrorole		Actor	Undergoer	_____
Labels		water agent, atmospheric agent, natural disaster [water]	area, construction	material [water]
Linguistic realizations		tide, sea wave, low tide, water overflows, rainfalls, tsunami, floods	coastal areas, land, region, school	water
Phrase type		NP	NP	PP (with)
Usage examples	1. Some tsunamis do not appear on shore as massive breaking waves but instead resemble a quickly surging tide that inundates coastal areas. 2. It generated a tsunami which inundated Hilo on the island of Hawaii. 3. A flood occurs when water overflows inundates land that's normally dry. 4. It describes an earlier eruption that generated a sea wave that inundated the land and killed many people. 5. The tsunami from the west and southwest inundates the relatively low-lying areas of Breuh Island at Lambaro. 6. Sydney was inundated with over 400 mm of rainfall. 7. This region has been inundated by destructive tsunami. 8. Even a low tide can still inundate coastal areas. 9. The floods inundated coastal communities and resulted in a catastrophic death toll of as many as 19,000 people. 10. Friday's flash flood inundated a school in Shalan, a remote town in China's far northeastern province of Heilongjiang. 11. Town inundated with water and traffic jam after heavy rain.			

3 Lexical domain: POSITION
Frame: to_put_water_over/in_a_space
Frame definition: [NATURAL FORCE] puts a lot of water [THEME] over/in a [PATIENT].
Note: The PATIENT is usually an area, landform, water course or construction. The THEME is often implicit.

Semantic role	SWAMP	Natural force	Patient	Theme
Macrorole		Actor	Undergoer	_____
Labels		water agent, atmospheric agent, natural disaster [water]	construction, area	material [water]
Linguistic realizations		wave, floodwaters, floods, tsunami, avalanche, waves	vehicles, riverboats, town, areas, beaches, people	water
Phrase type		NP	NP	PP (with)
Usage examples	1. Damage exceeded $A100 million, including over 3000 motor vehicles swamped by floodwaters. 2. Nine hours after the blast, three hundred riverboats were swamped and sunk at Kolkata. 3. The town of Anjer Lor was swamped by an 11 m high wave.			

		4. Eight climbers were missing and presumed dead Sunday after an avalanche swamped a commonly used hiking trail near Mont Blanc. 5. The floods swamped wide areas of the eastern United States, killing at least 32 people. 6. The tsunami swamped every boat in the harbor. 7. Tsunami waves swamped Hawaii beaches and severely damaged harbors in California. 8. The storm swamped the city with water.

Table 142. Analysis of the frame *to put water over/in a space* (English)

1	Lexical domain: POSITION Frame: to_put water over/in a space Frame definition: [NATURAL FORCE] puts a lot of water [THEME] over/in a [PATIENT]. Note: The PATIENT is usually an area, landform, water course or construction. Both the THEME and NATURAL FORCE can be be implicit				
	Semantic role	INUNDAR	Natural force	Patient	Theme
	Macrorole		Actor	Undergoer	_____
	Labels		water agent, atmospheric agent, natural disaster [water], natural disaster [earth], material entity, water course	area, construction, landform, water course	material [water]
	Linguistic realizations		huracán, riada, lluvia, agua, río, flujo	casas, ciudad, ha, viviendas, edificios, cultivos, isla	agua
	Phrase type		NP	NP	PP (de)
	Usage examples	1. Esta inundación fue mayor a la de 1933, cuando otro huracán inundó la ciudad. 2. El huracán inundó algunas casas en zonas rurales, causó apagones y sumergió algunas carreteras costeras. 3. El huracán inundó la ciudad y dejó unas 1.800 víctimas mortales a su paso. 4 Esta riada provocó 200 víctimas mortales, inundó 30.000 ha y destruyó 3.000 viviendas y 11 puentes. 5. Una fuerte lluvia inundó rápidamente el centro de la ciudad de agua. 6. La ola sísmica inundó las costas del sur de Portugal y del suroeste de España. 7. En la costa de Texas en 1980, una marejada de tormenta de cuatro metros inundó parte de la isla. 8. Los flujos inundaron muchas viviendas y edificios. 9. Las aguas también inundaron cultivos, destruyendo cosechas de arroz a lo largo de la franja costera de Indonesia. 10. Tres ríos inundaron las zonas bajas del centro de Manabí.			
2	Lexical domain: POSITION Frame: to_put water over/in a space Frame definition: [NATURAL FORCE] puts a lot of water [THEME] over/in a [PATIENT]. Note: The PATIENT is usually an area, landform, water course or construction. Both the THEME and NATURAL FORCE can be be implicit				
	Semantic role	ANEGAR	Natural force	Patient	Theme
	Macrorole		Actor	Undergoer	_____
	Labels		water agent,	area, construction	material

			atmospheric agent, natural disaster [water]		
	Linguistic realizations		diluvio, avalancha, tsunami, huracán, marejada, lluvia, inundaciones, marea	km2, poblados, edificios, calles, áreas, locales comerciales, garajes, sótanos	agua
	Phrase type		NP	NP	PP (de)
	Usage examples	1. El diluvio anegó 317.000 km2 y ocasionó la pérdida de 70.000 vidas. 2. La avalancha anegó de agua seis poblados en el municipio de Riohacha. 3. La tormenta tras convertirse en huracán, anegó buena parte de Nueva Orleans (Luisiana) y destruyó el área del Golfo de México. 4. Un tsunami anega varios edificios en la ciudad japonesa de Kamaishi, tras un terremoto de 8,9 de intensidad en la escala Richter. 5. El presidente Barack Obama declaró "zona catastrófica" la áreas anegadas por la marejada del ciclón, que ha dejado al menos 15 muertos. 6. La intensa lluvia anegó varias calles de distintos barrios. 7. La lluvia, por su lado, ha anegado algunos locales comerciales, garajes, sótanos. 8. Las inundaciones anegaron siete polígonos industriales en las afueras de Bangkok. 9. Las inundaciones anegaron decenas de poblados y afectaron las comunicaciones en la región. 10. Las fuertes mareas anegaron poblaciones costeras del sudoeste.			

Table 143. Analysis of the frame *to put water over/in a space* (Spanish)

As can be inferred from the analysis, verbs in the frame *to put water over/in a space* generally take at least one argument which can refer to the CONSTRUCTION, HUMAN BEING or AREA being filled with the role of PATIENT and macrorole of UNDERGOER. Very often the NATURAL FORCE that triggers the ACTION is also stated. It is normally a NATURAL FORCE instantiated by a NATURAL DISASTER involving water. Occasionally the THEME can also be included, which specifies the MATERIAL with which the PATIENT was filled (e.g. 'water'), but it is often implicit since it is part of the verb meaning.

The template of this frame is very similar to the frame *to put sth on top of or over sth*. Both involve a NATURAL FORCE, PATIENT and THEME. There are, however, differences between the two. The main difference is that in this frame *(to_put water over/in a space)*, the THEME is almost exclusively constrained to material water entities, and the PATIENT to an area. In addition, it does not normally have a HUMAN BEING as a PATIENT unlike the frame *to put sth on top or over sth*. Another difference is the presence of the category of a WATER COURSE which can work both as NATURAL FORCE or PATIENT. Finally, both frames also differ in the nature of the NATURAL DISASTER

which is the NATURAL FORCE. As such, whereas the frame *to put sth on top of or over sth* was mainly constrained to EARTH NATURAL DISASTERS ('landslide') or fire events, the frame *to put water on top or over sth* is usually constrained to NATURAL DISASTERS involving water (e.g. 'hurricane').

Frame: to put water over/in a space				
Semantic role	Natural force	*flood* *inundate* *swamp* *inundar* *anegar*	Patient	Theme
Macrorole	Actor		Undergoer	
Conceptual class	natural disaster [water] water agent water course		area, construction, landform, water course	material [water]
Phrase type	NP		NP	PP (with) (English) PP (de) (Spanish)

Table 144. Template of the frame *to put water over/in a space*

5.5 Implementation of phraseological information in EcoLexicon

As shown in 1.4.3.2, EcoLexicon provides a great quantity of linguistic information for each term, but does not provide any verbal phraseological information. Accordingly, the general objective of this research was the inclusion of a phraseology module providing the verb collocations for each term within the term information window in Figure 17 (§1.4.3.2.2.2).

This section explains the encoding, entry, and storage of phraseology in EcoLexicon (§5.5.1), the accessing of collocations in the macrostructure of Ecolexicon (§5.5.2), as well as the microstructure of a phraseological entry (§5.5.3).

5.5.1 Storing and recording information

Phraseological information was encoded in the storage database of EcoLexicon, which can only be accessed by the members of the Lexicon research group. To that end, in the term module, a separate section was created to record phraseology.

As shown (Figure 75), the template for phraseological information resembles the verb analysis templates in 5.4.3. However, in order to facilitate the access, no metalanguage was used to encode and describe collocations. For this reason information regarding semantic categories, roles and macroroles was excluded. These notions were used to configure our frames, and they will eventually be used to systematize verb patterns as well as to create a more formalized ontology. In addition, in the future, there are also plans to provide various types of information in consonance with the profile of

the user. As such, a more expert user with a linguistic profile will have access to the metalanguage.

Figure 75 is an extract of the encoding from the frame *to cause to change for the worse* associated with the term 'hurricane'. As shown, the frame includes the lexical domain (CHANGE), the name of the frame (*to cause to change for the worse)*, the definition of the frame (NATURAL DISASTER causes a PATIENT to change for the worse), the verb ('affect'), the usage examples which instantiate the verb in context, as well as a *note section* for meaning and usage restrictions. New verb entries in the same frame (e.g. 'damage', at the end of Figure 75) will only include its usage examples and notes since the lexical domain, frame, and frame definition have already been provided. This is the process that was followed to encode all the phraseological information regarding the EXTREME EVENT.

Figure 73. Extract of phraseological module for storing and recording information

5.5.2 The macrostructure: ways of accessing collocations

The phraseological module for EcoLexicon as proposed in this thesis enables users to access collocational information both via the base (term), as well as the collocate (verb).

For instance, 'hurricane destroys' can either be accessed in the corresponding entry for 'hurricane' (Figs. 78 and 79), or by searching for the verb 'destroy' (Figure 77). This method of encoding phraseology is helpful both for the translation of a collocation from L1 to L2 (e.g. the Spanish translation for 'the hurricane destroys'?) and for the production in L2 of a collocation in which the collocate conveys a specific meaning, i.e. the meaning of *to cause to change for the worse* as applied to 'hurricane'.

When searching via the term (i.e. 'hurricane'), the complete network of the concepts associated with the search term is displayed, as shown in Figure 76. Phraseological information is provided in the *Terms section* on the left side of the screen (Figure 76). As such, by clicking on the English term (i.e. 'hurricane'), a new window is displayed in which, along with the rest of the linguistic information regarding the term, verb collocations for the term are displayed. The microstructure of the phraseological entry for the entry for 'hurricane' is explained in 5.5.3 and displayed in Figures 78 and 79.

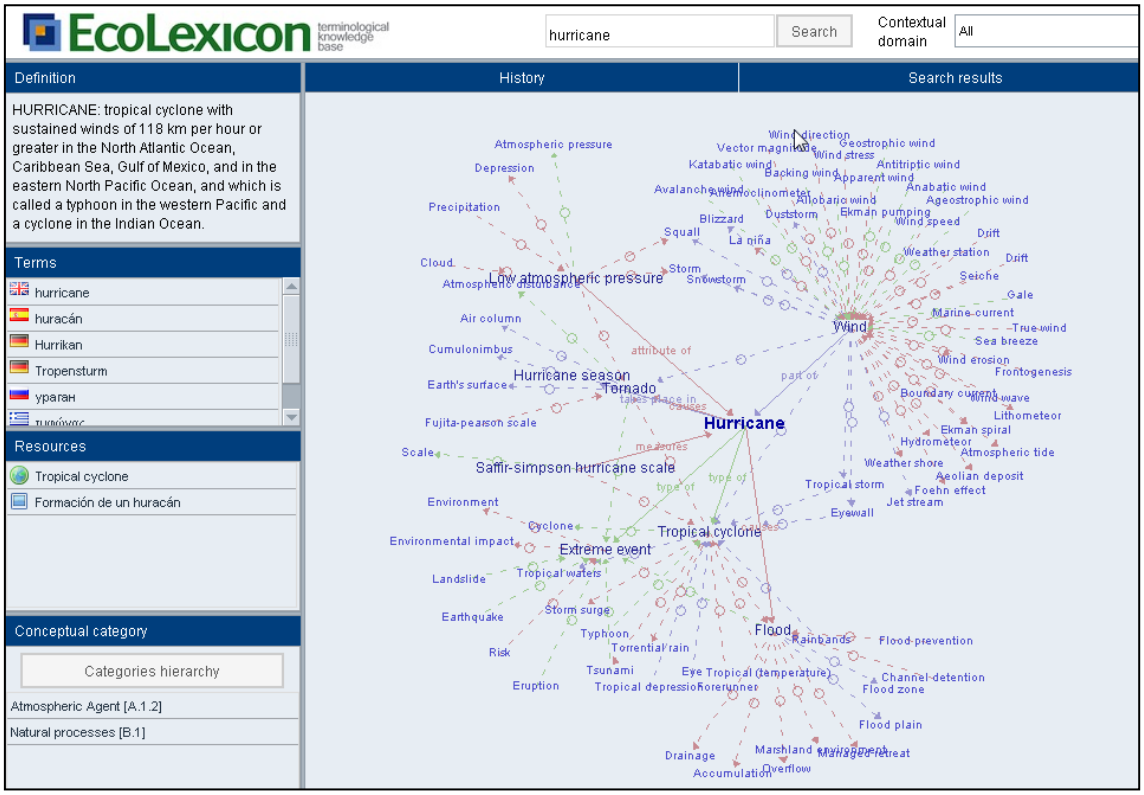


Figure 74. Search for collocations via the base

In addition, users can search via the collocate in order to retrieve the associated terms that collocate with the search verb. To that end, the user only has to enter the verb in the EcoLexicon search window and the system provides the *search results* for the

words introduced. By clicking on the verb, the user has access to the complete set of terms with which the verb can combine. For instance, as shown in Figure 77, the verb 'destroy' can combine with the terms 'tsunami', 'hurricane', 'eruption', 'fire', 'tornado', 'avalanche', 'flood', 'earthquake', 'flooding', 'landslide', 'quake', among others. By clicking on each term, a new window appears that redirects the user to the corresponding term entry.

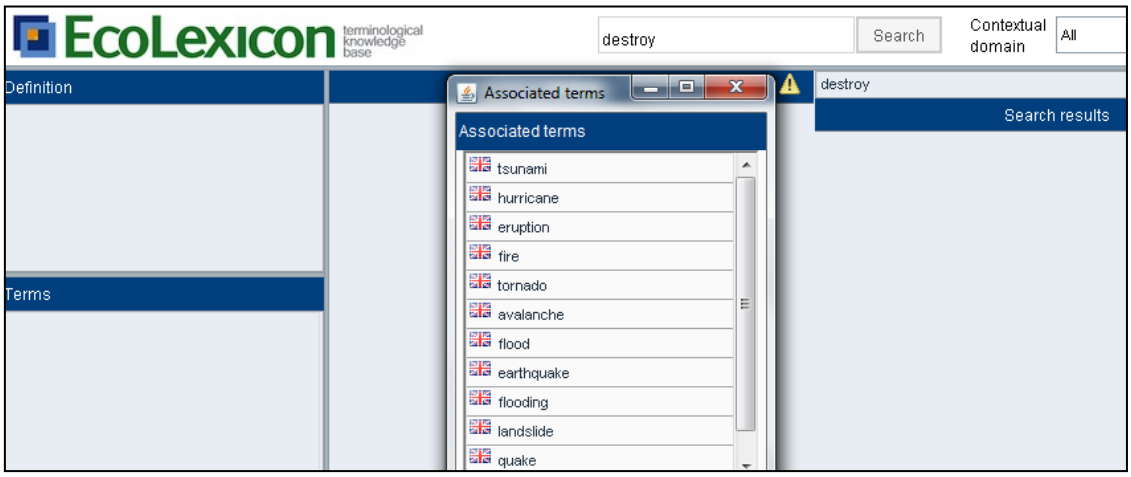


Figure 75. Search for collocations via the collocate

5.5.3 The microstructure of entries

Phraseological information is displayed in the term entry of each term in EcoLexicon along with morphosyntactic information regarding term type (main term, synonym, geographical variant, and acronym); gender (masculine, feminine, and neuter); contexts of use; and grammatical category (noun, verb, adjective or adverb). Figure 78 and 79 display the entry for the term 'hurricane' in EcoLexicon. As shown, once the term type is stated (main term) [término principal], a link to the usage contexts of 'hurricane' is provided (Fig. 78), the grammatical category is specified (noun), and the module for phraseology is offered:

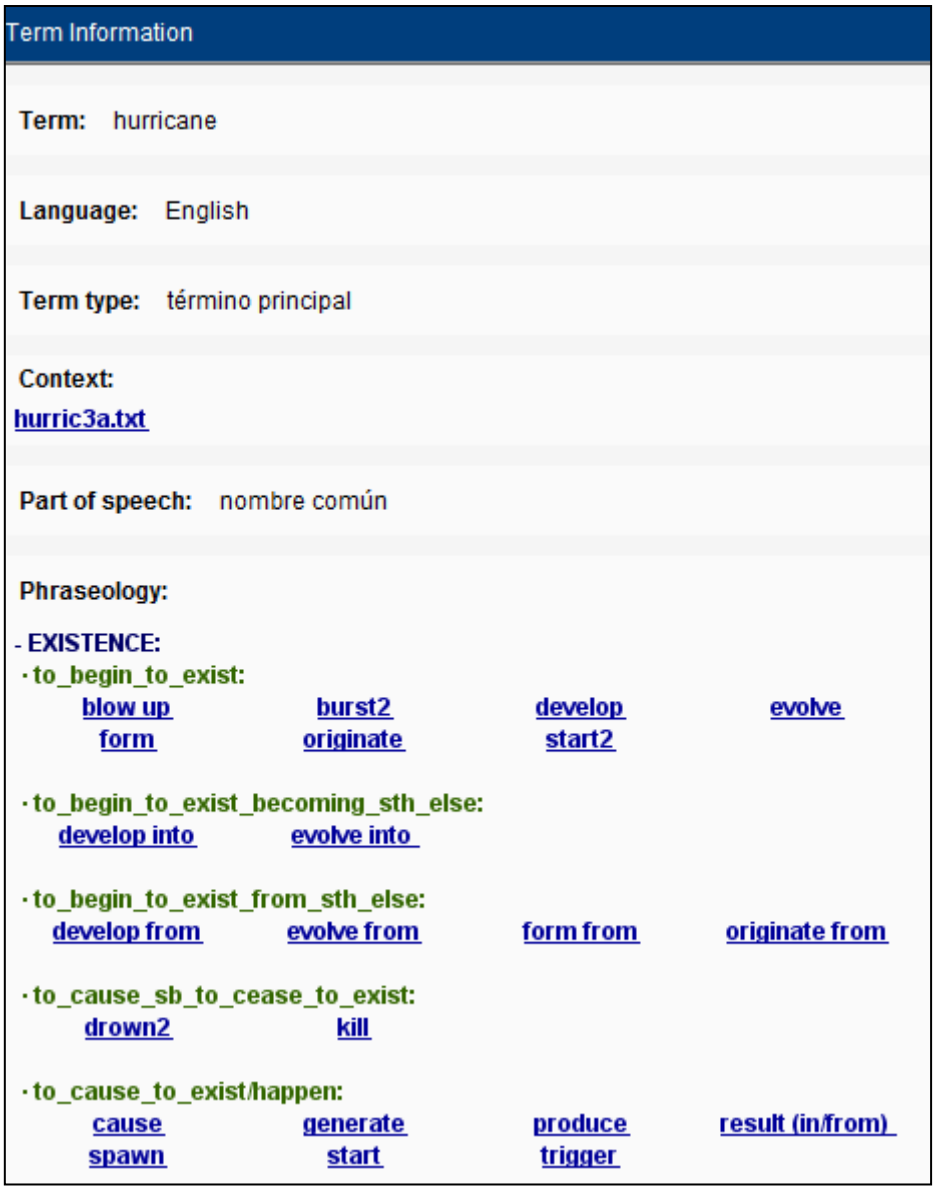


Figure 76. Extract 1 of the entry of the term 'hurricane' including the new module of phraseology

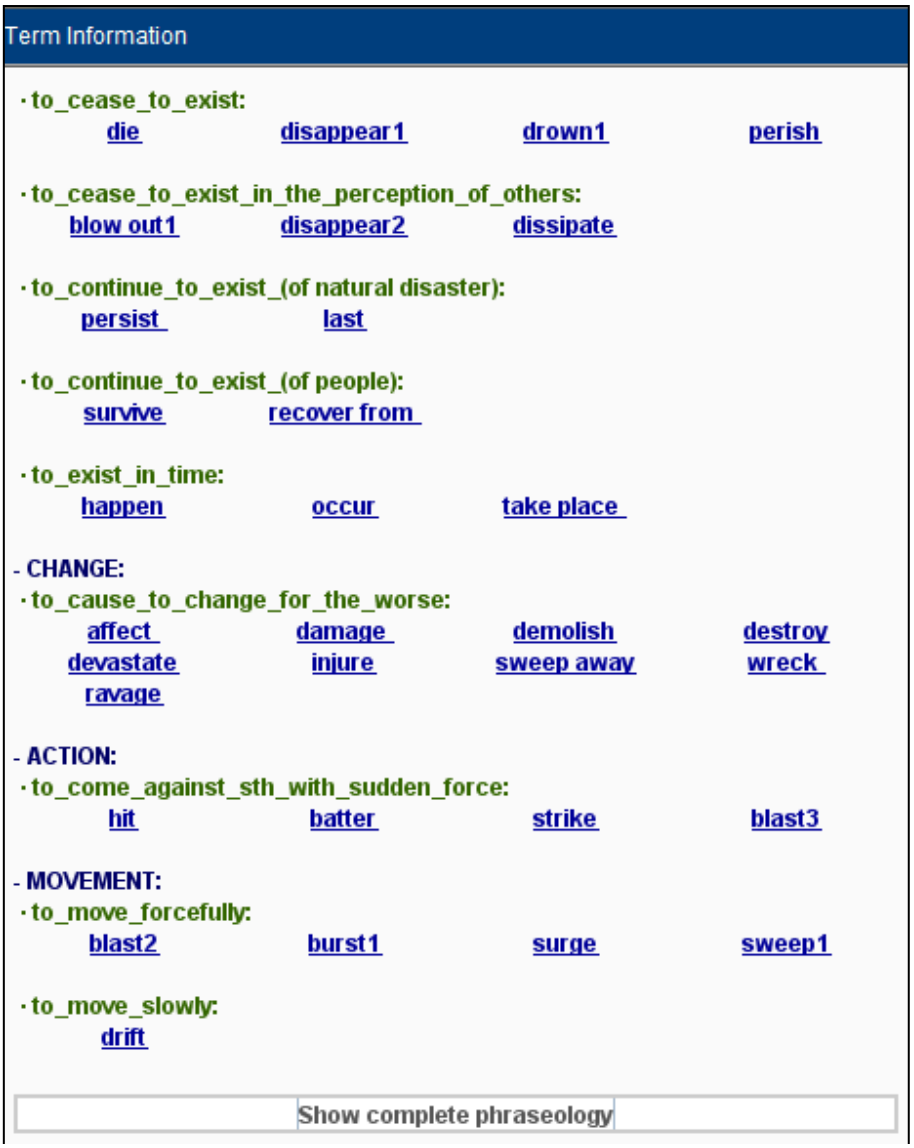


Figure 77. Extract 2 of the entry of the term 'hurricane' including the new module of phraseology

As shown in Figures 78 and 79, verb collocations are classified and described according to meaning. For this reason, they were primarily classified in terms of their lexical domain (capital letters in blue), and subsequently in terms of the frame activated within each lexical domain (roman letters in green). Once the lexical domain and frames are stated, the verbs are specified in blue and underlined.

For instance, in the frame *to cause to change for the worse*, the verbs 'affect', 'damage', 'destroy', 'devastate', 'ravage', 'demolish', 'wreck', 'sweep away', 'burn1', and 'injure' are given. By clicking on the verbs, the user has access to the usage sentences for the verb in question, as well as a note section with information about meaning restrictions. Accordingly, by clicking on the verb 'demolish', Figure 80 is displayed:

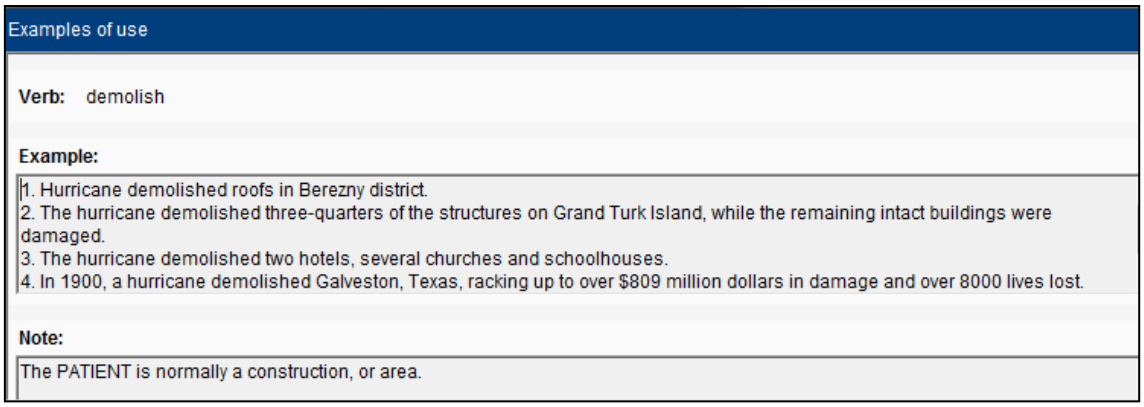


Figure 78. Verb usage examples in EcoLexicon for 'demolish'

As shown, concerning the verb 'demolish', the user is given four usage examples. There is also a note section that states that the PATIENT is usually a CONSTRUCTION ENTITY, or AREA.

In addition, the user can have access to the all phraseological information for a certain term by clicking on *show complete phraseology* at the end of the microstructure of entries (Figure 79). The phraseological information shown for 'hurricane' is the following (Figure 81):

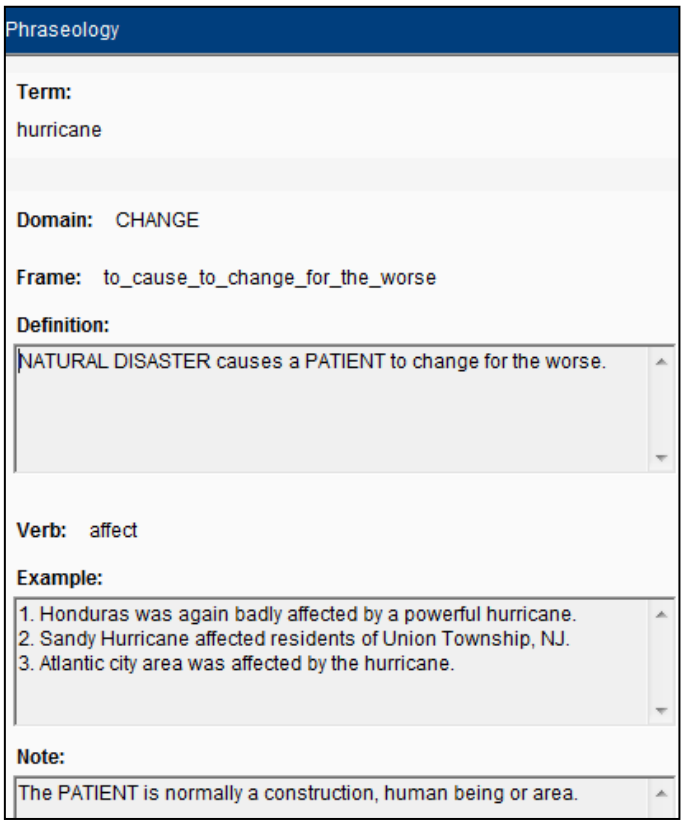


Figure 79. Extract of the complete phraseology concerning the term 'hurricane'

Finally, correspondences between English and Spanish collocations can also be found in EcoLexicon. All the verbs included in the term section in EcoLexicon for a specific concept have similar meaning. This means that the verbs related to 'hurricane' in Spanish can be accessed by clicking on the Spanish term 'huracán'. The same will then be provided for Spanish. As such, if the users are looking for a collocation in Spanish with 'huracán', which conveys the meaning *to cause to change sth for the worse*, they will only have to look for this frame in the term entry in Spanish and see the verbs belonging to the frame, as displayed in Figure 82:

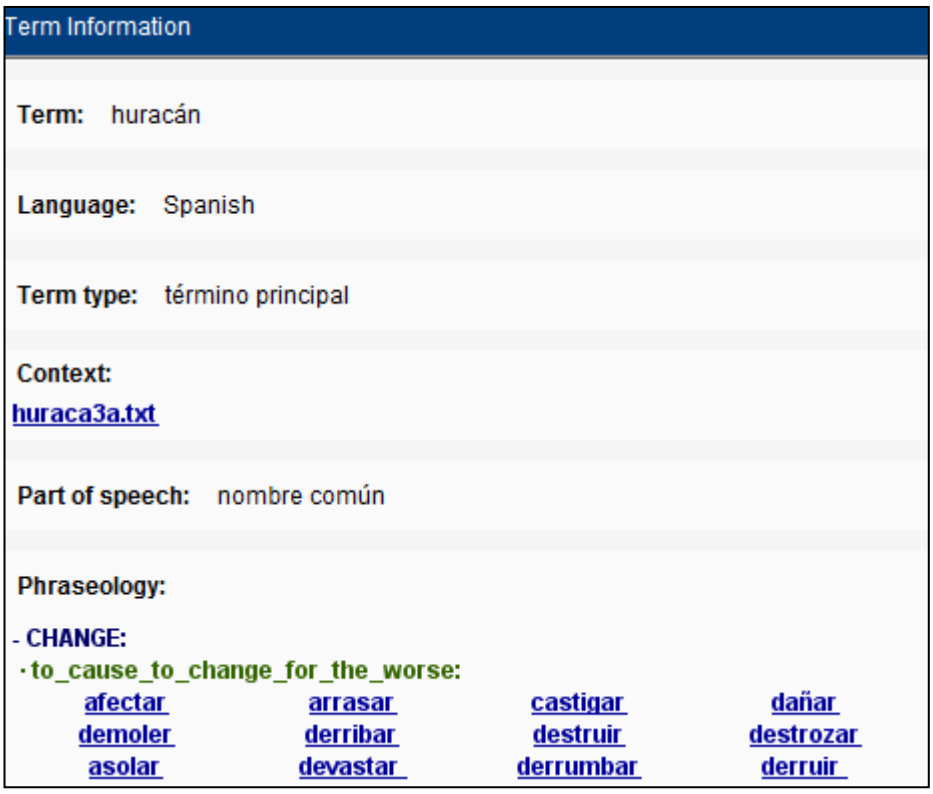


Figure 80. Extract of the term entry 'huracán'

In the future, the frames in one language will be linked to frames in the other. The correspondence between frames would thus be direct. From a computational perspective, this is more complex since it means restructuring the links in the whole knowledge base.

5.5.4. Summary

This chapter presents the analysis of the data performed in this research, as well as the results and discussion. First, by means of a bottom-up approach, candidate verbs were

extracted with the term extractor, TermoStat (§5.3). Then, verbs were classified in lexical domains as proposed by the Lexical Grammar Model, based on their definitions (top-down approach) (§5.4.1). This provided the most prominent lexical domains activated within the EXTREME EVENT: EXISTENCE, ACTION, CHANGE, MOVEMENT, and POSITION. Finally, verb meaning was refined by studying concordances that reflected the activation of these verbs in texts (bottom-up approach) (§5.4.2). More specifically, in this bottom-up analysis, predicates and their arguments were identified and analyzed as follows: (i) specification of the arguments and their linguistic realizations in order to semantically label the categories designated by the arguments (§5.4.2.1.1); (ii) assignment of semantic roles to the arguments (§5.4.2.1.2); (iii) description of the corresponding phrase type of the arguments in texts (§5.4.2.1.3).

Subsequently, all verbs within the same semantic category were grouped together, and assigned a name or *frame label*. Even though these frame labels resemble the lexical subdomains of the LGM (§3.1.2.1), we decided to call them *frames* because FBT is a frame-based approach that encompasses verbs and argument patterns. Each frame was defined in terms of its underlying lexical domain (e.g. EXISTENCE, POSITION, CHANGE, or MOVEMENT) and also in terms of the semantic categories assigned to its arguments. The verb analysis was first performed for English and then for Spanish. The Spanish correspondence with the English frames is based on Pimentel (2012). As such, once the semantic categories in each Spanish concordance line were identified, they were directly associated with the frames in English which had the same category labels.

The analysis for each verb was encoded in a template for the subsequent implementation of phraseological information in EcoLexicon. This allowed us to specify the frames in the EXTREME EVENT, and linked each frame to a template encoding the requirements and constraints of the frame and thus, of all the verbs that belong to the frame.

Finally, we described how phraseology is encoded, entered, and stored in EcoLexicon (§5.5.1). We also explained how collocations are reflected in the macrostructure of Ecolexicon (§5.5.2), as well as the microstructure of a phraseological entry (§5.5.3). As shown, the methodology proposed in this thesis for encoding and describing verbal collocations in terminographical resources is useful both for text comprehension and text production.

6. Conclusions

The objective of this thesis was to design a template for encoding and describing phraseological information in EcoLexicon, and to create a methodology for the codification of phraseological information that can be applied to other specialized fields as well.

As previously mentioned, this thesis focused on verbs since few terminological resources include information regarding predicates and their argument combinations. This is paradoxical given that verbs are regarded as the most important lexical and syntactic category in language. As such, the center of our analysis was verb collocations, namely combinations of the noun+verb type or verb+noun type. In our research, collocation is defined in its broadest sense. In other words, in contrast to other authors, who only focus on collocations such as 'to make the bed', we also regard as collocations, combinations as transparent as 'the fire burns'.

In such collocation, the predicate 'burn' requires noun phrases designating combustible entities. In addition, 'fire' requires a verb designating a combustion process ('burn'). Consequently, the semantic features encoded in a verb's meaning interact with, or are in some degree, imposed by those of the meaning of its argument/s. However, at the same time, the verb also constrains the arguments that can combine with it.

Our approach to collocations also includes a certain degree of compositionality in that each lexical unit in a collocation retains its meaning. In this regard, in 'hurricane damages' and 'hurricane hits', both elements in the collocation maintain their respective meanings. Nonetheless, they are not free combinations. Even though the verb selects its arguments, it is also true that to some extent, the verb is also imposed by the meaning of the noun. For this reason, both arguments and verbs were analyzed in this research.

The phraseological template proposed in this research takes its theoretical insights from cognitive approaches to Terminology, such as Sociocognitive Terminology and Frame-based Terminology. In addition, assumptions from linguistic models that deal with predicate argument structures were also included, namely the Lexical Grammar Model, Lexical Constructional Model, Role and Reference Grammar, and FrameNet.

The practical guidelines for the design of our terminological phraseological entry template were taken from the conclusions derived from the analysis of the most representative lexicographic and terminographic resources that contained phraseological

information, along with an analysis of the most important meaning-based resources for verbs.

For the methodology in this thesis, we combined the web for corpus (WfC) approach and the web as corpus (WaC) approach. Regarding the web for WfC approach, two subcorpora of specialized texts belonging to the domain of NATURAL HAZARDS or the EXTREME EVENT were compiled, one for English, and one for Spanish. For this purpose, we first designed a protocol for the evaluation of online resources from which texts are extracted. This protocol assured the reliability and quality of the texts in our corpus and can be applied to any type of corpus. In addition, usage contexts and phrases were directly retrieved from the web (i.e. following a *Web as Corpus* approach). This was done when the information in our corpus did not satisfy our needs, when examples were not found, or when a wider range of contexts was needed.

In consonance with Frame-based Terminology, we followed both a top-down and bottom-up approach. First, candidate verbs were extracted with the term extractor, *TermoStat*. Then, these verbs were classified in lexical domains as proposed by the LGM, based on their definitions. This permitted us to specify the most prominent lexical domains within the EXTREME EVENT, namely, EXISTENCE, ACTION, CHANGE, MOVEMENT, and POSITION. Finally, the meanings of the verbs were refined by studying their activation in texts as reflected in concordances. More specifically, both arguments and predicates were identified and analyzed. This allowed us to discard those verbs that were not directly associated with any kind of NATURAL HAZARD, i.e. verbs that did not have as one of their arguments an atmospheric disturbance or an associated phenomenon.

The underlying idea is that when verbs and their arguments are classified and structured in a set of conceptual-semantic categories typical of a given specialized domain, along with the semantic roles and macroroles as well as the phrase structure activated, it is then possible to establish templates that generalize this type of meaning for entire semantic subdomains. Accordingly, the range of verbs generally associated with a certain category can be predicted within the frame of a specialized event given that in specialized language, verb meaning is constrained by the specialized subject field.

As such, the semantic categories for EXTREME EVENT were found to be the following: NATURAL DISASTER, ATMOSPHERIC AGENT, WATER AGENT, ATMOSPHERIC CONDITION, MATERIAL ENTITY, AREA, CONSTRUCTION, ENERGY, HUMAN BEING,

LANDFORM, WATER COURSE, DEATH, DAMAGE, LOSS OF LIFE/PROPERTY, PLANT, and EXPLOSIVE. The inventory of semantic roles in this research was the following: AGENT, NATURAL FORCE, DESTINATION, EXPERIENCER, FREQUENCY, GEOGRAPHICAL LOCATION, MANNER, PATH, PATIENT, SITUATION/EXPERIENCE, ORIGIN, THEME, TIME and RESULT. Along with semantic categories and semantic roles, an additional categorization in terms of macroroles (i.e. ACTOR and UNDERGOER), as proposed by RRG, was also provided.

This type of predicate argument analysis to specify the frame templates linked to the EXTREME EVENT, allowed us to arrive at the following conclusions regarding verb behavior:

1. Verbs belonging to the same frame normally have the same number and type of arguments, which belong to the same type of semantic categories, and have the same semantic roles and macroroles. In consonance with the LGM, since all predicates in each frame are hierarchically organized according to their meaning, there is a focalization of verb meaning as the hierarchy becomes more specific. In other words, verbs inherit both semantic and syntactic patterns from their hyperonym, but not all the arguments of the hyperonym will be activated by the hyponyms. For example, in the frame *to begin to exist* (which includes 'start2', 'break out', 'originate', 'develop', 'evolve', 'blow up', 'form', 'burst2', 'erupt1', and 'explode2'), 'start2' can be used in a wide variety of contexts. In contrast, 'erupt1' and 'explode2' are constrained to THEME arguments that refer to volcanic natural disasters. As such, 'the **avalanche** started', 'the **hurricane** started suddenly' and 'the **volcano** started suddenly' are all acceptable. However, even though one cay say that 'the **volcano** exploded/erupted suddenly', this verb cannot be used with a wind disaster THEME, 'the **hurricane** exploded/erupted suddenly'*. Even though verbs within a frame are not total synonyms, and thus not interchangeable in all contexts, their meaning is similar.
2. The same verb can belong to different frames in different lexical domains, depending on the arguments activated, i.e. when the arguments belong to either different semantic categories, or have different roles or macroroles. For example, 'burst' was shown to have three senses within the EXTREME EVENT frame: (i) 'burst1' (MOVEMENT), meaning *to move forcefully (esp. of wind events)* with a first argument designating a NATURAL DISASTER restricted to wind events, and a second argument designating PATH (e.g. 'the **hurricane** burst **through the city's levees**'); (ii) 'burst2' (EXISTENCE), meaning *to begin to exist*

(esp. of volcanic events), with a first argument designating a NATURAL DISASTER restricted to volcanic events, and optional arguments designating GEOGRAPHICAL LOCATION, TIME or MANNER (e.g. 'the **volcano** burst **out fiercely in Italy'**; (iii) 'burst3' (ACTION) meaning *to (cause) to come apart, (esp. of construction artifacts)* with a first argument designating a construction entity with the role of PATIENT and a second argument with the role of SITUATION/EXPERIENCE (e.g. 'the **dam** burst **in torrential rain'**). When the construction is causative, the first argument is a NATURAL DISASTER usually associated with wind events with the role of NATURAL FORCE, whereas the second argument is a PATIENT which designates a CONSTRUCTION ARTIFACT (e.g. 'the **typhoon** burst **a dike** in Kaohsiung').

3. The same verb can belong to different frames within the same lexical domain depending on the arguments of the verbs in the frame. Sometimes this change of arguments is linguistically identifiable by the use of a new element, such as a change of preposition. For example, 'evolve' can be activated in three frames within the lexical domain of EXISTENCE: (i) *to begin to exist*; (ii) *to begin to exist from sth else*; (iii) *to begin to exist becoming sth else*. Although the first argument for the three frames is the same (NATURAL DISASTER, which is a THEME and UNDERGOER), the change is reflected in the semantic nature of the second argument which can be a GEOGRAPHICAL LOCATION, TIME, or MANNER for the first frame (e.g. 'the tropical cyclone evolved **in tropical and subtropical waters**'); an ORIGIN in the second frame (e.g. 'the tropical cyclone evolved **from a depression'**); or a RESULT as in the third frame (e.g. 'the tropical cyclone evolved **into an extratropical cyclone'**). Even though the frame *to begin to exist* can be instantiated by 'evolve', it is true that this verb is usually followed by *from* or *into* as shown in the concordances examples. Consequently, a change in the syntax of a verb phrase also reflects a change in semantic information, and thus activate different frames.
4. Frames belonging to different lexical domains can instantiate the same underlying template, but there will always be differences that determine which verbs belong to one frame and which belong to another. These differences can be derived from verb meaning since the definition of verbs in one frame will clearly differ from those in another. Verb meaning imposes the obligatoriness of certain arguments, which are typical of the frame. For example, the template of

frames *to come against sth with sudden force* (ACTION) coincides with that of *to cause to change for the worse* (CHANGE). Both templates are instantiated by a NATURAL DISASTER with the role of NATURAL FORCE and macrorole of ACTOR, and a second argument that can be a CONSTRUCTION, AREA, or HUMAN BEING with the role of PATIENT and macrorole of UNDERGOER. The difference between the two frames resides in the verbs contained in each. As such, in *to change sth for the worse* the only strictly obligatory argument is PATIENT (e.g. '**The coast** was devastated') since NATURAL FORCE is implicit. In contrast, in *to come against sth with sudden force*, the obligatory argument is NATURAL DISASTER '**the hurricane** struck with force' and the PATIENT can be implicit.

5. There are frames which have no verbal lexicalizations in the other language given the differences in the way a certain lexical domain or frame is instantiated in the two languages. This was the case of the frame *to move forcefully*. These frames contain the English verbs 'surge', 'sweep1', 'blast2', 'burst1', but there are no Spanish verbs with this meaning. The underlying reason is the difference in which MOVEMENT is lexicalized in Spanish. In Spanish, this type of meaning is conveyed by combining verbs from the frame *to move in a certain direction* (e.g. 'ascender', 'elevar', 'emerger', 'transportar', 'penetrar', 'adentrar', 'entrar', 'cruzar', 'atravesar'), or simply *to move* (e.g 'mover', 'desplazar', 'circular') with adverbial modification indicating intensity.

The template proposed in this thesis for encoding verb phraseological information within the domain of the ENVIRONMENT is believed to satisfy the whole set of criteria specified in chapter 2 (§2.4) for the design of a phraseological entry in a specialized resource:

1. EcoLexicon is an online resource and the phraseological entries are provided in electronic format.
2. Phraseological information is provided both for English and Spanish and users can easily find the correspondences between the phraseological units in both languages.
3. The phraseological template proposed is theoretically based on premises from Frame-based Terminology (§1.4.3), Lexical Grammar Model (§3.1.2.1), Role and Reference Grammar (§3.1.2.2), as well as FrameNet (§3.1.3.2).

4. The template is conceived for both encoding and decoding purposes, which is justified by the various ways of accessing phraseological information.
5. Difficult metalanguage is not used. As previously mentioned, linguistic notions, such as *role*, *macrorole*, or *conceptual category* were used in the analysis but are not encoded in the template in EcoLexicon.
6. The phraseological information module allows for various ways of searching and accessing collocations, via the collocate or via the base.
7. The module does not contain excessive collocational information per entry and the information within an entry is displayed in different colors.
8. The module provides a semantic classification of collocations, based on meaning in terms of lexical domains and frames.
9. The module provides a detailed description of the semantic and syntactic patterns associated with the meaning of each collocation in the definition of the frame and in the use note in each entry.
10. The module includes various usage examples showing the word in context.

This phraseological module integrates the positive aspects of existing knowledge resources to provide a frame-based representation that is in consonance with the premises of Frame-based Terminology and which is useful for text comprehension as well as production. The main objective of this research was thus attained: the design of a phraseological template for encoding and describing phraseological information in EcoLexicon.

Regarding future research, the inventory of categories for the EXTREME EVENT is only a preliminary classification. As such, our aim is to refine the inventory of categories and at the same time, to establish a complete set of categories for the domain of the ENVIRONMENT. This categorization will ultimately lead to a system of semantic tagging, which will be a step further towards the large-scale automatic retrieval of specialized knowledge. This will used in the implementation of the incipient ontology currently under construction in Ecolexicon.

Another issue to be dealt with in the near future is the establishment of direct correspondences between collocations in different languages. As shown, EcoLexicon provides an indirect correspondence between verbal collocations in English and Spanish. The user accesses the term and its collocations in one language and then must

search for the collocation that expresses the desired meaning in the other. It is our hope that soon it will be possible to directly link frames in one language to frames in another.

In consonance with the pevious idea, another aim is to eventually extend this methodology to the other languages in EcoLexicon. Frames could thus be generalized as structures underlying interlinguistic correspondence. It would thus be possible to establish a direct correspondence between collocations in different languages. This methodology would also be extended to the other subevents in EcoLexicon. This will be possible because the corpus will have been semantically tagged.

The phraseological template in this research is mainly for translators. For this reason, no metalanguage (e.g. the specification of semantic category, role, macrorole, etc.) was used. However, an expert with a more linguistic profile may find it useful to have access to this metalanguage. In this regard, another future objective is to include more user profiles and to provide various levels and types of phraseological information in consonance with the profile of the user.

7. References

Aisenstadt, E. 1979. "Collocability Restrictions in Dictionaries." In Hartmann, R.R.K. (ed.) *Dictionaries and Their Users*. 71–74. Exeter: University of Exeter.

Alcaraz-Varó, E. 2006a. *Diccionario De Términos De La Piedra Natural e Industrias Afines: inglés-español, Spanish-English*. Barcelona: Ariel.

———. 2006b. *Diccionario De Términos Del Calzado e Industrias Afines: inglés-español, Spanish-English*. Barcelona: Ariel.

Alcaraz-Varó, E., and J. Castro-Calvín. 2007. *Diccionario De Comercio Internacional: Importación y Exportación (inglés-español, Spanish-English)*. Barcelona: Ariel.

Alcaraz-Varó, E., and B. Hughes. 2007. *Diccionario De Términos Jurídicos: inglés-español, Spanish-English*. 10th edition. Barcelona: Ariel.

———. 2008. *Diccionario De Términos Económicos, Financieros y Comerciales: inglés-español = A Dictionary of Economic, Financial and Commercial Terms: Spanish-English*. 5th edition. Barcelona: Ariel.

Alexander, J.E., and M.A. Tate. 1999. *Web Wisdom. How to Evaluate and Create Information Quality on the Web*. Hillsdale, New Jersey: Lawrence Erlbaum.

Allaby, A., and M. Allaby. 1990. *The Concise Oxford Dictionary of Earth Sciences*. Oxford: Oxford University Press.

Allaby, M. 1988. *Macmillan Dictionary of the Environment*. London: MacMillan.

Alonso, M. 2010. "No Importa Si La Llamas o No Colocación, Descríbela." In Mellado, C., P. Buján, C. Herrero, N. Iglesias, and A. Mansilla (eds.) *La Fraseología Del S.XXI: Nuevas Propuestas Para El Español y El Alemán*. 55–80. Berlin: Frank & Timme. <http://www.dicesp.com/app/webroot/files/file/Alonso 2010(1).pdf>. [05/07/2011].

Alonso, M., A. Nishikawa, and O. Vincze. 2010. "DICE in the Web: An Online Spanish Collocation Dictionary." In Granger, S. and M. Paquot (eds.) *eLexicograpy in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009*. 369–374. Louvain-la-Neuve: Cahiers du Cental 7, Presses universitaires de Louvain. <http://www.dicesp.com/app/webroot/files/file/eLex 2009 DiCE in the Web.pdf>. [25/02/2011].

Alonso, A., and J. DeCesaris. 2007. "El Léxico Del Medio Ambiente En Los Diccionarios Generales". In *Actas X Simposio Intenacional De Comunicación Social (22-26 De enero de 2007)*. 68-72. Santiago de Cuba: Centro de Lingüística

Aplicada. Ministerio de Centro de Lingüística Aplicada. Ministerio de Ciencia, Tecnología y Medio Ambiente. <http://www.iula.upf.edu/agenda/age019_06.pdf>. [05/12/2011].

Amosova, N.N. 1963. *Osnovui Anglijskoy Frazeologii*. Leningrad: University Press.

Appleby, R. 2000. "Review of M. Benson, E. Benson, and R. Ilson. The BBI Dictionary of English Word Combinations; J. Hill and M. Lewis. Dictionary of Selected Collocations; K. Stern. Longman Idioms Dictionary." *ELT Journal* 54(1): 89–91.

Apresjan, J.D., I. Mel'čuk, and A.K. Žolkovsky. 1969. "Semantics and Lexicography: Towards a New Type of Unilingual Dictionary." In Kiefer, F. (ed.) *Studies in Syntax and Semantics*. 1–33. Dordrech/Boston: D. Reidel.

Atkins, S., J. Clear, and N. Ostler. 1992. "Corpus Design Criteria." *Literary and Linguistic Computing* 7(1): 1–16.

Auer, N.J. 1999. *Evaluating Internet Information*. University Libraries at Viginia Tech. <http://www.lib.vt.edu/help/instruct/evaluate/evaluating.html>. [22/02/2009].

Austermühl, F. 2001. *Electronic Tools for Translators*. Manchester: St. Jerome.

Baker, C.F., and J. Ruppenhofer. 2002. "FrameNet's Frames Vs. Levin's Verb Classes." In *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. 27–38. Berkeley, California: UC Linguistics Department.

Baker, M. 1995. "Corpora in Translation Studies. An Overview and Suggestions for Future Research." *Target* 7(2): 243–257.

Baldewein, U., K. Erk, S. Pado, and D. Prescher. 2004. "Semantic Role Labeling with Similarity Based Generalization Using EM-based Clustering." In *Proceedings of Senseval-3*, 64–68.

Bally, Ch. 1951. *Traité De Stylistique Française*. 3rd edition. Paris: Librairie C. Klincksieck.

Baroni, M., and S. Bernardini. 2004. "BootCaT : Bootstrapping Corpora and Terms from the Web." In Lino, Mª.T., Mª.F. Xavier, F. Ferreira, R. Costa, and R. Silva (eds.) *Proceedings of LREC 2004 (Fourth International Conference on Language Resources and Evaluation)*. 1313–1316. Lisbon. <http://sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf>. [22/04/2011].

———. 2006. *Wacky! Working Papers on the Web as Corpus*. Bologna: GEDIT.

Baroni, M., A. Kilgarriff, J. Pomikálek, and P. Rychlý. 2006. "WebBootCat: Instant Domain-specific Corpora to Support Human Translators." In *Proceedings of*

EAMT 2006 - 11th Annual Conference of the European Association for Machine Translation. 247–252. Norway: EAMT.

Baroni, M., S. Bernardini, A. Ferraresi, and E. Zanchetta. 2009. "The WaCky Wide Web: a Collection of Very Large Linguistically Processed Web-crawled Corpora." *Language Resources and Evaluation Journal* 43(2): 209–226.

Barrios, M.A. 2007. "Diccionarios Combinatorios Del Español: Diferencias y Semejanzas Entre Redes y Práctico." *redELE. Revista Electrónica De Didáctica Del Español Como Lengua Extranjera* 11: 1–14. <http://www.mecd.gob.es/dctm/redele/Material-RedEle/Revista/2007_11/2007_redELE_11_01Barrios.pdf?documentId=0901e72b80df2cb7>. [05/01/2010].

———. 2010. *El Dominio De La Funciones Léxicas En El Marco De La Teoría Sentido-texto*. Ph.D. Thesis presented at the University Complutense of Madrid, Spain. <http://elies.rediris.es/elies30/>. [07/01/2011].

Barrière, C. 2004. "Building a Concept Hierarchy from Corpus Analysis." *Terminology* 10(2): 241–263.

Barrière, C., and A. Agbago. 2006. "TerminoWeb: A Software Environment for Term Study in Rich Contexts." In *Proceedings of the International Conference on Terminology, Standardisation and Technology Transfer (TSTT 2006)*. Beijing, China.

Barsalou, L.W. 2003. "Situated Simulation in the Human Conceptual System." *Language and Cognitive Processes* 18(5-6): 513–562.

Bauer, L. 1988. "When Is a Sequence of Two Nouns a Compound in English?" *English Language and Linguistics* 2(1): 65–86.

Benito, G., and A. Díez. 2004. *Riesgos Naturales y Antrópicos En Geomorfologia: Libro De Actas De Los Simposios Desarrollados Durante La VIII Reunión Nacional De Geomorfología, Celebrada En Toledo, 22-25 de septiembre de 2004*. Madrid: Sociedad Española de Geomorfología.

Benson, M. 1989. "The Structure of the Collocational Dictionary." *International Journal of Lexicography* 2(1): 1–13.

Benson, M., E. Benson, and R. Ilson. 1986. *Lexicographic Description of English. Studies in Language Companion Series. Volume 14*. Amsterdam/Philadelphia: John Benjamins.

———. 2009. *The BBI Combinatory Dictionary of English*. 3rd edition. Amsterdam/Philadelphia: John Benjamins.

Bergenholtz, H., and S. Nielsen. 2002. "Terms in the Language of Culture-Dependent LSP Dictionaries." *Lexicographica. International Journal of Lexicography* 18: 5–18.

Bergenholtz, H., and S. Tarp. 2002. "Die Moderne Lexikographische Funktionslehre. Diskussionsbeitrag Zu Neuen Und Alten Paradigmen, Die Wörterbücher Als Gebrauchsgegenstände Verstehen." *Lexicographica* 18: 253–263.

———. 2003. "Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions." *Hermes. Journal of Linguistics* 31: 171–196.

———. 2004. "The Concept of Dictionary Usage." *Nordic Journal of English Studies* 3: 23–36.

———. 2010. "LSP Lexicography or Terminography? The Lexicographer's Point of View." In Fuertes-Olivera, P.A. (ed.) *Specialized Dictionaries for Learners*. 27–37. Berlin/New York: De Gruyter.

Bernardini, S. 2008. "Do We Still Need Corpora (now That We Have the Web)?" In *Conference Presented at the Postgraduate Conference in Corpus Linguistics*. Birmingham: Aston University.

Bernardini, S., M. Baroni, and S. Evert. 2006. "A WaCky Introduction." In Baroni, M., and S. Bernardini (eds.) *WaCky! Working Papers on the Web as Corpus*. 1–32. Bologna: GEDIT.

Bernardos, M.S., R. Jiménez, and M.B. Pérez. 2011. "Una Aplicación Informátic Para La Gestión De Las Plantillas Léxicas Del Modeloo Léxico Construccional." *Revista De Lingüística y Lenguas Aplicadas* 6: 53–69.

Biber, D. 1993. "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8(4): 243–257.

Binon, J., S. Verlinde, J. Van Dyck, and A. Bertels. 2001. *Dictionnaire D'apprentissage Du Français Des Affaires. Dictionnaire De Compréhension Et De Production De La Langue Des Affaires*. Paris: Didier.

Blackburn, S. 1996. *The Oxford Dictionary of Philosophy*. Oxford: Oxford University Press.

Blasco, E. 2000. *Los Límites Entre Perífrasis Verbales y Unidades Fraseológicas Verbales*. Ph.D. Thesis presented at the University of Barcelona, Spain.

Bogaards, P. 1996. “Dictionaries for Learners of English.” *International Journal of Lexicography* 9(4): 277–320.

Boguraev, B., and J. Pustejovsky. 1996. *Corpus Processing for Lexical Acquisition*. Cambridge: The MIT Press.

Bosque, I. 2004. *Redes. Diccionario Combinatorio Del Español Contemporáneo*. Madrid: S.M.

———. 2006. *Diccionario Combinatorio Práctico Del Español Contemporáneo*. Madrid: S.M.

Boulanger, J.C. 1991. “Une Lecture Socioculturelle De La Terminologie.” *Cahiers De Linguistique Sociale* 18: 13–30.

———. 1995. “Présentation: Images Et Parcours De La Socioterminologie.” *Meta* XL(2): 195–205.

Bowker, L. 1996. “Towards a Corpus-based Approach to Terminography.” *Terminology* 3(1): 27–52.

———. 1997. “Multidimensional Classification of Concepts and Terms.” In Wright, S.E., and G. Budin (eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Vol.1. 133–143. Amsterdam/Philadelphia: John Benjamins.

———. 2010. “The Contribution of Corpus Linguistics to the Development of Specialised Dictionaries for Learners.” In Fuertes-Olivera, P.A. (ed.) *Specialized Dictionaries for Learners*. 155–168. Berlin/New York: De Gruyter.

Bowker, L, and I. Meyer. 1993. “Beyond Textbook. Concept Systems: Handling Multidimensionality in a New Generation of Term Banks.” In Schmitz, K.D. (ed.) *TKE ’93 Terminology and Knowledge Engineering*. 123–137. Frankfurt: Indeks Verlag.

Bowker, L., and J. Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Bryant, E. 2005. *Natural Hazards*. 2nd edition. Cambridge: Cambridge University Press.

Buendía, M. 2012. “Verb Dynamics.” *Terminology* 18(2): 149–166.

Butler, C.S. 2009. “The Lexical Constructional Model : Genesis , Strengths and Challenges.” In Butler, C.S., and J. Martín (eds.) *Deconstructing Constructions*. 117–152. Amsterdam/ Philadelphia: John Benjamins. <http://www.lexicom.es/drupal/files/Butler_2009.pdf>. [08/02/2011].

Cabré, Mª.T. 1993. *La Terminología. Teoría, Metodología, Aplicaciones*. Barcelona: Antártida/Empúries.

———. 1999a. *Terminology. Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins.

———. 1999b. *Terminología: Representación y Comunicación. Una Teoría De Base Comunicativa y Otros Artículos*. Barcelona: Universitat Pompeu Fabra.

———. 2001a. "Consecuencias Metodológicas De La Propuesta Teórica I." In Cabré, Mª.T., and J. Feliu (eds.) *La Terminología Científico-técnica: Reconocimiento, Análisis y Extracción De Información Formal y Semántica*. 27–36. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

———. 2001b. "Sumario De Principios Que Configuran La Nueva Propuesta Teórica." In Cabré, Mª.T., and J. Feliu (eds.) *La Terminología Científicotécnica: Reconocimiento, Análisis y Extracción De Información Formal y Semántica*. 17–26. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

———. 2000. "Elements for a Theory of Terminology: Towards an Alternative Paradigm." *Terminology* 6(1): 35–57.

———. 2003. "Theories of Terminology: Their Description, Prescription and Explanation." *Terminology* 9(2): 163–199.

Cabré, Mª.T., and M. Lorente. 2005. "Panorama De Los Paradigmas En Lingüística." In Estany. A. (ed.) *Filosofía De Las Ciencias Naturales, Sociales y Matemáticas*. 433–467. Madrid: Consejo Superior de Investigaciones Científicas.

Caignon, P. 2000. *Essential Lexicon in Accounting*. Quebec: Fides.

———. 2001. *Vocabulaire Et Cooccurrents De La Comptabilité*. Montreal: Linguatech.

Calderón, M. 1994. *Sobre La Elaboración De Diccionarios Monolingües De Producción. Las Definiciones, Los Ejemplos y Las Colocaciones Léxicas*. Ph.D. Thesis presented at the University of Granada, Spain.

Campos, M.A., and E. Alcaraz-Varó. 2002. *Diccionario De Términos De La Propiedad Inmobiliaria: inglés-español, Spanish-English*. Barcelona: Ariel.

Carter, R. 1987. *Vocabulary: Applied Linguistics Perspectives*. London: Unwin Hyman.

———. 1988. "Compositionality and Polysemy." In Levin, B., and C. Tenny (eds.) *On Linking: Papers by Richard Carter*. 167–204. Cambridge: Center for Cognitive Science, MIT.

Castagnoli, S. 2006. "Using the Web as a Source of LSP Corpora in the Terminology Classroom." In Baroni, M., and S. Bernardini (eds.) *WaCky! Working Papers on the Web as Corpus*. 159–172. Bologna: GEDIT.

Castro-Calvín, J., and E. Alcaraz-Varó. 2003. *Diccionario De Términos De Seguros: inglés-español, Spanish-English*. Barcelona: Ariel.

Chacón, J., and C. Irigaray. 1996. *Riesgos Naturales, Ordenación Del Territorio y Medio Ambiente. Sexto Congreso Nacional y Conferencia Internacional De Geología Ambiental y Ordenación Del Territorio =Natural Hazards, Land-use Planning and Environment, 6th Spanish Congress and Internationa*. Madrid: Sociedad Española de Geologia Ambiental y Ordenación del Territorio.

Chafe, W. 1994. *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Janua Linguarum 4.

———. 1964. *Current Issues in Linguistic Theory*. 5th edition. The Hague: Mouton.

———. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

———. 1988. *Language and Problems of Knowledge*. Cambridge: MIT Press.

Codina, L. 2000. "Parámetros e Indicadores De Calidad Para La Evaluación De Recursos Digitales." In *Actas De Las VII Jornadas Españolas De Documentación. La Gestión Del Conocimiento: Retos y Soluciones De Los Profesionales De La Información*. 135–144. Bilbao: Universidad del País Vasco servicio editorial.

Coffey, S. 2011. "A New Pedagogical Dictionary of English Collocations." *International Journal of Lexicography* 24(3): 328–341.

Cohen, B. 1986. *Lexique De Cooccurrents. Bourse-conjoncture Économique*. Montreal: Linguatech.

Collins, A.M., and M.R. Quillian. 1969. "Retrieval Time from Semantic Memory." *Journal of Verbal Learning and Verbal Behavior* 8: 240–247.

Cooke, A. 1999. *A Guide to Finding Quality Information on the Internet: Selection and Evaluation Strategies*. London: Library Association Publishing.

Corpas, G. 1990. "Tratamiento De Las Colocaciones Del Tipo A+S/S+A En Diccionarios Bilingües y Monolingües (español-inglés)." In *Proceedings of the IV EURALEX International Congress*. 331–340. Málaga: University of Málaga. <http://www.euralex.org/elx_proceedings/Euralex1990>. [22/07/2010].

———. 1996. *Manual De Fraseología Española*. Madrid: Editorial Gredos.

———. 1998. “Criterios Generales De Clasificación Del Universo Fraseológico De Las Lenguas, Con Ejemplos Tomados Del Español y Del Inglés.” In Alvar Ezquerra, M., and G. Corpas (eds.) *Diccionarios, Frases, Palabras*. 157–187. Málaga: Servicio de Publicaciones de la Universidad.

———. 2001a. “En Torno Al Concepto De Colocación.” *Euskera* XLVI(1): 89–108.

———. 2001b. “Compilación De Un Corpus Ad Hoc Para La Enseñanza De La Traducción Inversa Especializada.” *TRANS. Revista De Traductología* (5): 155–184.

———. 2001c. “Apuntes Para El Estudio De La Colocación.” *LEA* XXIII: 41–56.

———. 2003. *Diez Años De Investigación En Fraseología: Análisis Sintáctico-semánticos, Contrastivos y Traductológicos*. Madrid: Iberoamericana.

Coseriu, E. 1981. *Lecciones De Lingüística General*. Madrid: Gredos.

Cowie, A.P. 1981. “The Treatment of Collocations and Idioms in Learners’ Dictionaries.” *Applied Linguistics* 2(3): 223–235.

———. 1990. “Pedagogical Descriptions of Language.” *Lexis. Annual Review of Applied Linguisitcs* 10: 196–209.

——— (ed.). 1998a. *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press.

———. 1998b. “Introduction.” In Cowie, A.P. *Phraseology. Theory, Analysis, and Applications*. 1–20. Oxford: Clarendon Press.

———. 1998c. “Phraseological Dictionaries: Some East-West Comparisons.” In Cowie, A.P. (ed.) *Phraseology. Theory, Analysis, and Applications*. 209–228. Oxford: Clarendon Press.

Croft, W. 2001. *Radical Constructional Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

———. 2003. “Lexical Rules Vs. Constructions. A False Dichotomy.” In Cuyckens, H., R. Dirven, and K.U. Panther (eds.) *Motivation in Language*. 49–68. Amsterdam/ Philadelphia: John Benjamins.

Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Danaher, D.S. 2007. “Úvod Do Kognitivní Gramatiky [An Introduction to Cognitive Grammar]’.” In Římalová Saicová, L. (ed.) *Čítanka Textů z Kognitivní Lingvistiky II [Reader in Cognitive Linguistics II]*. 89–103. Prague: Charles University, Faculty of Arts.

De Schryver, G.M. 2002. "Web For/as Corpus: a Perspective for the African Languages." *Nordic Journal of African Studies* 11, 266-282. <http://tshwanedje.com/publications/webtocorpus.pdf>. [25/07/2009].

Dik, S.C. 1978a. *Functional Grammar*. Dordrecht: Foris.

———. 1978b. *Stepwise Lexical Decomposition*. Lisse: de Ridder.

———. 1989. *The Theory of Functional Grammar. Part I: The Structure of the Clause. (Functional Grammar Series 9)*. Dordrecht: Foris.

———. 1997a. *The Theory of Functional Grammar: The Structure of the Clause*. Berlin: Mouton de Gruyter.

———. 1997b. *The Theory of Functional Grammar: Complex and Derived Constructions*. Berlin: Mouton de Gruyter.

Dolbey, A, M. Ellsworth, and J. Scheffczyk. 2006. "BioFrameNet: a Domain-specific FrameNet Extension with Links to Biomedical Ontologies." In *The Second International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2006), Biomedical Ontology in Action*. Baltimore, Maryland, USA. <http://ceur-ws.org/Vol-222/krmed2006-p10.pdf>. [04/07/2009].

Domínguez, A., K. Kerremans, and R. Temmerman. 2012. "Strategies in Automatic Traversal of Wikipedia Articles for Mining Multilingual Resources." In *Proceedings of the Workshop on Challenges to Knowledge Representation in Multilingual Contexts (TKE 2012 Conference). Presented at the Challenges to Knowledge Representation in Multilingual Contexts*. Madrid.

Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Foris.

Drouin, P. 2003. "Term Extraction Using Non-technical Corpora as a Point of Leverage." *Terminology* 9(1): 99–117.

EAGLES (EXPERT ADVISORY GROUP ON LANGUAGE ENGINEERING). 1996a. *Text Corpora Working Group Reading Guide*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

———. 1996b. *Preliminary Recommendations on Corpus Typology*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.

Elkiss, A., and P. Resnik. 2004. *The Linguist's Search Engine User's Guide.* <http://lse.umiacs.umd.edu/lseuser.pdf>. [06/08/2009].

Engwall, G. 1992. "Comments." In Svartvik, J. (ed.) *Directions in Corpus Linguistics*. 164–169. New York/Berlin: Mouton de Gruyter.

European Environment Agency. 2010. *Mapping the Impacts of Natural Hazards and Technological Accidents in Europe*. Luxembourg: Publications Office of the European Union.

Evans, V., and M. Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Faber, P. 1999. "Conceptual Analysis and Knowledge Acquisition in Scientific Translation." *Terminologie Et Traduction* 2: 97–123.

———. 2002. "Terminographic Definition and Concept Representation." In Maia, B., J. Haller, and M. Ulyrich (eds.) *Training the Language Services Provider for the New Millennium*. 343–354. Oporto: University of Oporto.

———. 2009. "The Cognitive Shift in Terminology and Specialized Translation." *MonTI. Monografías De Traducción e Interpretación* 1(1): 107–134.

———. 2011. "The Dynamics of Specialized Knowledge Representation: Simulational Reconstruction or the Perception-action Interface." *Terminology* 17(1): 9–29.

——— (ed.). 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin, Boston: Mouton de Gruyter.

Faber, P., and C. Jiménez. 2002. *Investigar En Terminología*. Granada: Comares.

Faber, P., P. León, J.A. Prieto, and A. Reimerink. 2007. "Linking Images and Words: The Description of Specialized Concepts." *International Journal of Lexicography* 20(1): 39–65.

Faber, P., C.I. López, and M. Tercedor. 2001. "La Utilización De Técnicas De Corpus En La Representación Del Conocimiento Médico." *Terminology* 7(2): 167–197.

Faber, P., and C.I. López. 2012. "Terminology and Specialized Language." In Faber, P. (ed.) *A Cognitive Linguistics View of Terminology and Specialized Language*. 9–31. Berlin, Boston: Mouton de Gruyter.

Faber, P, and R Mairal. 1999. *Constructing a Lexicon of English Verbs*. Berlin/New York: Mouton de Gruyter.

Faber, P, C. Márquez, and M. Vega. 2005. "Framing Terminology: a Process-oriented Approach." *META* 50(4). CD-ROM. <http://www.erudit.org/livre/meta/2005/000255co.pdf> [06/04/2010].

Faber, P., and J.M. Ureña. 2012. "Specialized Language Translation." In Faber, P. (ed.) *A Cognitive Linguistics View of Terminology and Specialized Language*. 73–92. Berlin, Boston: Mouton de Gruyter.

Felber, H. 1984. *Terminology Manual*. Viena: Infoterm.

Fellbaum, C. 1990. "English Verbs as a Semantic Net." *International Journal of Lexicography1* 3(4): 279–301.

———. 1998. "A Semantic Network of English Verbs." In Fellbaum, C. (ed.) *WordNet: An Electronic Lexical Database1*. 69–104. MA: The MIT Press.

———. 2006. "WordNet and Wordnets." In Brown, K. et al. (eds.) *Encyclopedia of Language and Linguistics*. 665–670. Oxford: Elsevier.

Fellbaum, C., and G.A. Miller. 1990. "Folk Psychology or Semantic Entailment? A Reply to Rips and Conrad." *The Psychological Review* 97: 565–570.

Fernández, J. 2011. *Representación Del Conocimiento Terminológico Difundido En La Web : Estudio Del Dominio Del Cambio Climático*. Ph.D. Thesis presented at the University of Granada, Spain.

Fernández, M., A. Gómez-Pérez, and N. Jurista. 1997. "METHONTOLOGY: From Ontologica Art Towards Ontological Engineering Workshop on Ontological Engineering." In *Spring Symposium Series*. AAAI 97. Stanford, USA.

Fernández, A., and G. Vázquez. 2012. "Análisis Cuantitativo Del Corpus SenSem." In Elorza, I., O. Carbonell, R. Albarrán, B. García, and M. Pérez (eds.) *Empiricism and Analytical Tools for 21st Century Applied Linguistics*. 157–170. Salamanca: Ediciones de la Universidad de Salamanca. <http://grial.uab.es/archivos/analisis cuantitativo del corpus sensem - aesla 2011 - def.pdf>. [22/10/2012].

Ferrando,V. 2012. *Aspectos Teóricos y Metodológicos Para La Compilación De Un Diccionario Combinatorio Destinado a Estudiantes De E/LE*. Ph.D. Thesis presented at the University Rovira i Virgili, Spain. <http://www.tesisenred.net/handle/10803/84025?show=full>. [02/02/2013].

Fillmore, C.J. 1968. "The Case for Case." In Bach, E., and E. Harms (eds.) *Universals in Linguistic Theory*. 1–90. New York: Holt, Rinehart and Wilson.

———. 1971. "Some Problems for Case Grammar." In O'Brien, R.J. (ed.) *Report on the Twenty-Second Annual Round Table Meeting on Languages and Linguistics*. 35–56. Washington: Georgetown University Press.

———. 1976. "Frame Semantics and the Nature of Language." In Harnad, S.R., H.D. Steklis, and J. Lancaster (eds.) *Origins and Evolution of Language and Speech*. 20–32. Annals of the NY Academy of Sciences, Vol. 280.

———. 1977. "Scenes and Frame Semantics." In Zampolli, A. (ed.) *Linguistic Structures Processing*. 55–83. Amsterdam: North Holland Publishing Company.

———. 1982. “Frame Semantics.” In The Linguistic Society of Korea (ed.) *Linguistics in the Morning Calm*. 111–137. Seoul: Hanshin.

———. 1985. “Frames and the Semantics of Understanding.” *Quaderni Di Semantica* 6: 222–254.

Fillmore, C.J., and B.T.S. Atkins. 1992. “Toward a Frame-based Lexicon: The Semantics of RISK and Its Neighbors.” In Lehrer, A., and E. Kittay (eds.) *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*. 75–102. Hillsdale: Erlbaum.

Fillmore, C.J., C.R. Johnson, and M.R.L. Petruck. 2003. “Background to Framenet.” *International Journal of Lexicography* 16(3): 235–250.

Fillmore, C.J., and P. Kay. 1993. *Construction Grammar*. Unpublished Manuscript. Berkeley: University of California.

Fillmore, C.J., P. Kay, and C. O’Connor. 1988. “Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone.” *Language* 64: 501–538.

Fillmore, C.J., M.R.L. Petruck, J. Ruppenhofer, and A. Wright. 2003. “FrameNet in Action: The Case of Attaching.” *International Journal of Lexicography* 16(3): 297–332.

Firth, J.R. 1957. *Papers of Linguistics 1939-1951*. London: Oxford University Press.

———. 1968. “Linguistic Analysis as a Study of Meaning.” In Palmer, F.R. (ed.) *Selected Papers of J.R. Firth 1952-59*. 12–26. London and Harlow: Longmans.

Fleischer, W. 1997. *Phraseolgie Der Deutschen Gegenwartssprache*. 2nd edition. Leipzig: Bibliographisches Institut, VEB Verlag Enzyklopädie.

Fletcher, W.H. 2001. “Concordancing the Web with KWiCFinder.” In *3rd North American Symposium on Corpus Linguistics and Language Teaching*. Boston. <http://kwicfinder.com/FletcherCLLT2001.pdf>. [02/07/2010].

———. 2004. “Making the Web More Useful as a Source for Linguistic Corpora.” In Connor, U. and T. Upton (eds.) *Corpus Linguistics in North America*. 191–205. Amsterdam: Rodopi.

———. 2007. “Implementing a BNC-Compare-able Web Corpus.” In Fairon, C., H. Naets, A. Kilgarriff, and G.M. De Schryver (eds.) *Building and Exploring Web Corpora*. 43–56. Lovain-la-Neueve: Cahiers du Central.

———. 2012.“Corpus Analysis of the World Wide Web.” In *The Encyclopedia of Applied Lingusitics*. Chapelle: Wiley-Blackwell.

<http://webascorpus.org/Corpus_Analysis_of_the_World_Wide_Web.pdf>. [02/12/2012].

Fontanelle, T. 1998. "Discovering Significant Lexical Functions in Dictionaries Entries." In Cowie, A.P. (ed.) *Phraseology. Theory, Analysis, and Applications*. 189–207. Oxford: Clarendon Press.

Francis, W.N. 1982. "Problems of Assembiling and Computerizing Large Corpora." In Johansson, S. (ed.) *Computer Corpora in English Language Research*. 7–24. Bergen: Norwegian Computing Centre for the Hummanities.

Fraser, B. 1970. "Idioms Within a Transformational Grammar." *Foundations of Language* 6: 22–42.

Frawley, W. 1988. "New Forms of Specialized Dictionaries." *International Journal of Lexicography1* 1(3): 189–213.

Freixa, J. 2002. V*ariació Terminològica: Anàlisi De La Variació Denominativa En Textos De Diferent Grau D'especialització De L'àrea De Medi Ambient*. Ph.D. Thesis presented at the University Pompeu Fabra, Barcelona, Spain. <http://www.tdx.cesca.es/ TDCat->. [24/01/2009].

———. 2006. "Causes of Denominative Variation in Terminology: A Typology Proposal." *Terminology* 12(1): 51–77.

Fuertes-Olivera, P.A. 2011. "Review of Macmillan Collocations Dictionary for Learners of English, edited by Michael Rundell." *Language Value* 3(1): 153–161.

Fuertes-Olivera, P.A., and M. Niño-Amo. 2011. "Internet Dictionaries for Communicative and Cognitive Functions: El Diccionario Inglés-Español De Contabilidad." In Fuertes-Olivera, P.A., and H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 168–186. London and New York: Continuum.

Gambier, Y. 1991. "Travail Et Vocabulaire Spécialisés: Prolegomènes à Une Socioterminologie." *Meta* 36(1): 44–60.

———. 1993. "Vers Une Histoire Sociale De La Terminologie." *Le Langage Et L'homme* 28(4): 233–246.

García, J.M., L. Costas, and S. Martínez. 2005. "Diátesis Verbales y Esquemas Construccionales. Verbos, Clases Semánticas y Esquemas Sintáctico-semánticos En El Proyecto ADESSE." In Wotjak, G., and J. Cuartero Otal (eds.) *Entre Semántica Léxica, Teoría Del Léxico y Sintaxis*. 373–384. Frankfurt am Main: Peter Lang.

Gaudin, F. 1993. *Pour Une Socioterminologie: Des Problèmes Pratiques Aux Pratiques Institutionnelles*. Rouen: Publications de l'Université de Rouen.

———. 2003. *Socioterminologie: Une Approche Sociolinguistique De La Terminologie*. Brussels: Duculot.

Geentjens, S., R. Temmerman, K. Kerremans, and P. De Baer. 2006. "Sociocognitive Terminology and Termontography." In *Proceedings of the Journées d'Etudes Sur Le Traitement Automatique De La Langue Arabe*, 138–151. Rabat, Morocco. <http://taalkunde.ehb.be/sites/www2.ehb.be/files/u96/JETALA2006_SKRP.pdf> [21/03/2009].

Geeraerts, D. 1995. "Cognitive Linguistics." In Verschueren, J., J.O. Östman, and J. Blommaert (eds.) *Handbook of Pragmatics*. 111–116. Amsterdam/ Philadelphia: John Benjamins.

Genesereth, M.R., and N.J. Nilsson. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo, CA.: Morgan Kaufmann Publishers.

Gildea, D, and D. Jurafsky. 2002. "Automatic Labeling of Semantic Roles." *Computational Linguistics* 28(3): 245–288.

Ginzburg, R.S., S.S. Khidekel, G.Y. Knyazeva, and A.A. Sankin. 1979. *A Course in Modern English Lexicology*. 2nd edition. Moscow: Vuisshaya Shkola.

Giuglea, A., and A. Moschitti. 2006. "Semantic Role Labeling via FrameNet, VerbNet and PropBank." In *Proceedings of the 21st International Conference on Computational Linguistics and The44th Annual Meeting of the ACL*. 929–936.

Gläser, R. 1986. *Phraseologie Der Englischen Sprache*. Leipzig: VEB Verlag Enzyklopädie.

———. 1988. "The Grading of Idiomaticity as a Presupposition for a Taxonomy of Idioms." In Hüllen, W. and R. Schulze (eds.) *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*. 264–279. Tübingen: Max Niemeyer.

———. 1994/95. "Relations Between Phraseology and Terminology with Special Reference to English." *ALFA* 7/8: 41-60.

Goddard, C. 2010. "The Natural Semantic Metalanguage Approach." In Heine, B., and H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis*. 459–484. Oxford: Oxford University Press.

Goddard, C., and A. Wierzbicka. 2002. *Meaning and Universal Grammar: Theory and Empirical Findings*. Amsterdam/ Philadelphia: John Benjamins.

———. 2005. “Universal Human Concepts as a Basis for Contrastive Linguistics.” In *Fourth International Contrastive Linguistics Conference*. Santiago de Compostela: University of Santiago de Compostela.

———. 2007. “NSM Analyses of the Semantics of Physical Qualities.” *Studies in Language* 31(4): 761–800.

Goldberg, A. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

———. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

———. 2010. “Verbs, Constructions and Semantic Frames.” In Rappaport Hovav, M., E. Doron, and I. Sichel (eds.) *Syntax, Lexical Semantics and Event Structure*. 39–58. Oxford: Oxford University Press.

Gordon-Murname, L. 1999. “Evaluating Net Evaluators.” *Searcher* 7(2): 57–66.

Grinev, S., and I. Klepalchenko. 1999. “Terminological Approach to Knowledge Representation.” In *TKE ’99: Proceedings of the 5th International Congress on Terminology and Knowledge Engineering*. 147–151. Vienna: TermNet.

Gross, G. 1994. “Classes d’objets et description des verbes.” *Langages* 115: 15–30.

Gross, M. 1968. *Grammaire Transformationnelle Du Français. Vol. 1 Syntaxe Du Verbe*. Paris: Larousse.

———. 1975. *Méthodes En Syntaxe*. Paris: Hermann.

———. 1977. *Grammarie Transformationnelle Du Français. Vol. 2 Syntaxe Du Nom*. Paris: Larousse.

———. 1981. “Les Bases Empiriques De La Notion De Prédicat Sémantique.” *Langages* 63: 7–52.

———. 1986. *Grammaire Transformationnelle Du Français. Vol. 3 Syntaxe De L’adverb*. Paris: ASSTRIL.

Gruber, T. R. 1993. “A Translation Approach to Portable Ontologies.” *Journal on Knowledge Acquisition* 5(2): 199–220.

Guilbert, L. 1973. “La Spécificité Du Terme Scientifique Et Technique.” *Langue Française* 17: 5–17.

Guillet, A., and C. Leclère. 1992. *La Structure Des Phrases Simples En Français : Les Constructions Transitives Locatives*. Geneva: Droz.

Gómez, A., M. Fernández, and O. Corcho. 2004. *Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web*. London: Springer.

Halliday, M.A.K. 1966. "Lexis as a Linguistic Level." In Bazell, C.E., J.C Catford, M.A.K. Halliday, and R.H. Robins (eds.) *In Memory of J.R. Firth*. 148–162. London: Longmans, Green & Co. Ltd.

Hanks, P. 2008. "Lexical Patterns. From Hornby to Hunston and Beyond." In Bernal, E., and J. DeCesaris (eds.) *Proceedings of the 13th EURALEX International Congress*. 89–129. Barcelona: Institute for Applied Linguistics, Pompeu Fabra University.

Harris, Z. 1968. *Mathematical Structures of Language*. New York: John Wiley and Son.

———. 1970. *Papers in Structural and Transformational Linguistics*. Dordrecht: D. Reidel.

Hausmann, F.J. 1979. "Un Dictionnaire Des Collocations Est-il Posible?" *Travaux De Linguistique Et De Littérature* 17(1): 187–195.

———. 1984. "Wortschatzlernen Ist Kollokationslernen: Zum Lehren Und Lernen Französischer Wortverbindungen." *Praxis Des Neusprachlichen Unterrichts* 31: 395–406.

———. 1985. "Kollokationen Im Deutschen Wörterbuch: Ein Beitrag Zur Theorie Des Lexicographischen Beispiels." In Bergenholtz, H., and J. Mugdan (eds.) *Lexicographie Und Grammatik*. 118–129. Tübingen: Niemeyer.

———. 1989. "Le Dictionnaire Des Collocations." In Hausmann, F.J., O. Reichmann, H.E. Wiegand, and L. Zgusta (eds.) *Wörterbücher/ Dictionaries/ Dictionnaires — Ein Internationals Handbuch Zur Lexikographie/ An International Enyclopedia of Lexicography/ Enyclopédie Internationale De Lexicographie*. 1010–1019. Berlin and New York: Walter de Gruyter.

Heid, U. 2001. "Collocations in Sublanguage Texts: Extraction from Corpora." In Wright, S.E. (ed.) *Handbook of Terminology Management: Applications-Oriented Terminology Management. Vol. II*. 788–808. Amsterdam/Philadelphia: John Benjamins.

Hill, J., and M. Lewis. 1997. *The LTP Dictionary of Selected Collocations*. Hove: Language Teaching publications.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Howarth, P. 1996. *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making (Lexicographica, Series Maior 75)*. Tübingen: Max Niemeyer.

Hundt, C. 1994. *Untersuchungen Zur Portugiesischen Phraseologie*. Willhemsfeld: Gottfried Egert.

Hundt, M., N. Nesselhauf, and C. Biewer. 2007. "Corpus Linguistics and the Web." In Hundt, M., N. Nesselhauf, and C. Biewer (eds.) *Corpus Linguistics and the Web*. 1–5. Amsterdam: Rodopi.

Jakobson, R., and M. Halle. 1956. *Fundamentals of Language*. The Hague: Mouton de Gruyter.

Jiménez, R. 2008. "Los Metalenguajes y Las Representaciones Léxico-semánticas: Una Propuesta Desde El Modelo Léxico-Construccional." In Moreno, A. (ed.) *Actas Del VIII Congreso De Lingüística General*. Madrid: Universidad Autónoma de Madrid. <http://www.lllf.uam.es/clg8/actas/pdf/paperCLG58.pdf>. [15/04/2009].

Jiménez, M., and V. Ortiz-Repiso. 2007. *Evaluación y Calidad De Sedes Web*. Gijón: Ediciones Trea, S.L.

Kageura, K. 1997. "Multifaceted/Multidimensional Concept Systems." In Wright, S.E., and G. Budin (eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Vol.1. 119–132. Amsterdam/Philadelphia: John Benjamins.

Kehoe, A., and A. Renouf. 2002. "WebCorp: Applying the Web to Linguistics and Linguistics to the Web." In Lassner, D., D. De Roure, and A. Iyengar (eds.) *Proceedings of the WWW 2002 Conference*. New York: ACM. <http://www2002.org/CDROM/poster/67/>. [25/01/2009].

Keller, E.A., and R.H. Blodget. 2007. *Riesgos Naturales: Procesos De La Tierra Como Riesgos, Desastres y Catástrofes*. Madrid: Pearson Educación.

Kerremans, K., R. Temmerman, and J. Tummers. 2003. "Representing Multilingual and Culture-specific Knowledge in a VAT Regulatory Ontology: Support from the Termontography Approach." *Lecture Notes in Computer Science*. 662–674. <http://taalkunde.ehb.be/sites/www2.ehb.be/files/u96/WORMCoRe2003-KTT.pdf>. [15/05/2010].

Kilgarriff, A. 2003. "Linguistic Search Engine." In Simov, K. and P. Osenova (eds.) *Proceedings of the Workshop on Shallow Processing of Large Corpora*. 53–58. Lancaster.

Kilgarriff, A, and G. Grefenstette. 2003. "Introduction to the Special Issue on the Web as Corpus." *Computational Linguistics* 29(3): 333–347.

Kilgarriff, A, P. Rychly, P. Smrz, and D. Tugwell. 2004. "The Sketch Engine." In *Proceedings of the 11th EURALEX International Congress*. 105–116. Lorient: EURALEX.

Kilgarriff, A., V. Kovář, S. Krek, I. Srdanovic, and C. Tiberius. 2010. "A Quantitative Evaluation of Word Sketches." In *Proceedings of the XIV Euralex International Congress*. 372–379. Leeuwarden: Fryske Academy. <http://www.euralex.org/elx_proceedings/Euralex2010/019_Euralex_2010_1_KILGARRIFF KOVAR KREK SRDANOVIC TIBERIUS_A Quantitative Evaluation of Word Sketches.pdf> [03/07/2012].

Kingsbury, P., and M. Palmer. 2002. "From Treebank to PropBank." In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*. 1989–1993. Las Palmas, Spain.

Kipper, K. 2005. *VerbNet: a Broad-coverage, Comprehensive Verb Lexicon*. Ph.D. Thesis presented at the University of Pennsylvania. <http://verbs.colorado.edu/~kipper/Papers/dissertation.pdf>. [24/05/2010].

Kipper, K., A. Korhonen, N. Ryant, and M. Palmer. 2006. "Extending VerbNet with Novel Verb Classes." In *Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. 1027-1032. Genoa, Italy.

Kipper, K., M. Palmer, and O. Rambow. 2002. "Extending PropBank with VerbNet Semantic Predicates." In *AMTA 2002 Workshop: Workshop on Applied Interlinguas*. Tiburon, CA. USA.

Kipper, K., B. Snyder, and M. Palmer. 2004. "Extending a Verb-lexicon Using a Semantically Annotated Corpus." In *Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. 1557-1560. Lisbon, Portugal.

Kjellmer, G. 1994. *Dictionary of Collocations*. Oxford: Clarendon Press.

Klappenbach, R. 1968. "Probleme Der Phraseologie." *Wissenschaftliche Zeitschrift Der Karl-Marx-Universität* 17(5): 22–227.

Klotz, M. 2003. "Review of Oxford Collocations Dictionary for Students of English." *International Journal of Lexicography* 16(1): 57–61.

Korhonen, A., and T. Briscoe. 2004. "Extended Lexical-Semantic Classification of English Verbs." In *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*. 38–45. Boston, MA.

<http://www.cl.cam.ac.uk/~alk23/newclass.pdf>. [09/09/2010].

Korhonen, A., and N. Ryant. 2005. "53 Novel Lexical-Semantic Verb Classes." *Unpublished Manuscript*. <http://www.cl.cam.ac.uk/~alk23/classes2/>. [25/03/2010].

Lainé, C. 1993. *Vocabulaire Combinatoire De La CFAO Mécanique = Combinatory Vocabulary of CAD/CAM in Mechanical Engineering*. Ottawa: Secrétariat d'État du Canada = Dept. of the Secretary of State of Canada.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

———. 1988. "Cognitive Semantics." In Eco, U., M. Santambragio, and P. Violi (eds.) *Meaning and Mental Representations*. Bloomington and Indianapolis: Indiana University Press.

———. 1993. "The Contemporary Theory of Metaphor." In Ortony, A. (ed.) *Metaphor and Thought*. 202–251. Cambridge: Cambridge University Press.

Lakoff, G., and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago.

———. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Langacker, R.W. 1982. "Space Grammar, Analysability, and the English Passive." *Language* 58(1): 22–80.

———. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisits. Vol.1*. Stanford: Stanford University Press.

———. 1991. *Foundations of Cognitive Grammar: Descriptive Application. Vol. 2*. Stanford: Stanford University Press.

———. 2006. "Dimensions of Defocusing." In Tsunoda, T., and T. Kageyama (eds.) *Voice and Grammatical Relations*. 115–137. Amsterdam/Philadelphia: John Benjamins.

———. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.

Laviosa, S. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam and Atlanta: Rodopi.

Leech, G. 1997. "Introducing Corpus Annotation." In Garside, R., G. Leech, and T. McEnery (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. 1–18. London: Longman.

———. 2004. “Adding Linguistic Annotation.” In Wynne, M. (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. 17–29. Oxford: Oxbow Books. <http://ahds.ac.uk/linguistic-corpora>. [15/02/2010].

Leroyer, P. 2006. “Dealing with Phraseology in Business Dictionaries: Focus on Functions — Not Phrases.” *Linguistik Online* 27(2): 183–194.

Leturia, I., A. Gurrutxaga, I. Alegria, and E. Aitzol. 2007. “CorpEus, a ‘Web as Corpus’ Tool Designed for the Agglutinative Nature of Basque.” In Fairon, C., A. Kilgarriff, H. Naets, and G.M. De Schryver (eds.) *Building and Exploring Web Corpora*. 69–82. Lovain-la-Neueve: Cahiers du Central.

Levie, W.H., and R. Lentz. 1982. “Effects of Text Illustrations: A Review of Resarch.” *Educational Communication and Technology Journal* 30: 195–232.

Levin, B. 1993. *English Verb Class and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Levin, B., and M. Rappaport Hovav. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge: MIT Press.

León, P. 2009. *Representación Multidimensional Del Conocimiento Especializado: El Uso De Marcos Desde La Macroestructura Hasta La Microestructura*. Ph.D. thesis presented at the University of Granada, Spain.

León, P., and P.J. Magaña. 2010. “EcoLexicon: Contextualizing an Environmental Ontology.” In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin. <http://lexicon.ugr.es/pdf/leonmagana2010.pdf>. [12/02/2011].

León, P., and A. Reimerink. 2010. “Knowledge Extraction and Multidimensionality in the Environmental Domain.” In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University. <http://lexicon.ugr.es/pdf/leonreimerink2010.pdf>. [12/02/2011].

León, P., and A. San Martín. 2011. “Distinguishing Polysemy from Contextual Variation in Terminological Definitions.” In Carrió, Mª.L., J. Contreras, F. Olmo, H. Skorczynska, I. Tamarit, and D. Westall (eds.) *Actas Del X Congreso De La Asociación Europea De Lenguas Para Fines Específicos: La Investigación y La Enseñanza Aplicadas a Las Lenguas De Especialidad y a La Tecnología*. 173–186. Valencia: Universitat Politècnica de València. <http://lexicon.ugr.es/pub/leo-dis>. [05/01/2012].

Lipka, L. 1974. “Probleme Der Analyse Englischer Idioms Aus Struktureller Und Generativer Sicht.” *Linguistik Und Didaktik* 20: 274–285.

Lorente, M. 2000. “Tipología Verbal y Textos Especializados.” In González, M., and M. Souto (eds.) *Cuestiones Conceptuales y Metodológicas De La Lingüística*. 143–153. Santiago de Compostela: Universidad de Santiago de Compostela. <http://www.upf.edu/pdi/iula/merce.lorente/docums/ml01.pdf>. [22/02/2010].

———. 2002. “Verbos y Discurso Especializado.” *Estudios De Lingüística Española* 16. <http://elies.rediris.es/elies16/Lorente.html>. [22/04/2011].

———. M. 2007. “Les Unitats Lèxiques Verbals Dels Textos Especialitzats. Redefinició D’una Proposta De Classificació.” In Lorente, M., R. Estopà, J. Freixa, J. Martí, and C. Tebé (eds.) *Estudis De Lingüístics i De Lingüística Aplicada En Honor De Mª Teresa Cabré Castellví. Volum 2: De Deixebles*. 365–380. Barcelona: Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. <http://ricoterm.iula.upf.edu/docums/16_lorente.pdf>. [22/02/2010].

Lyons, J. 1977. *Semantics. Vols.1-2*. London: Cambridge University Press.

López, C.I. 2000. *Tipología Textual y Cohesión En La Traducción Biomédica Inglés-español: Un Estudio De Corpus*. Ph.D. Thesis presented at the University of Granada, Spain.

———. 2007. “Understanding Scientific Communication Through the Extraction of the Conceptual and Rhetorical Information Codified by Verbs.” *Terminology* 13(1): 61–84.

López, C.I., and M. Buendía. 2011. “En Busca De Corpus Online a La Carta En El Aula De Traducción Científica y Técnica.” *Trans-kom* 4(1): 1–22. <http://www.trans-kom.eu/bd04nr01/trans-kom_04_01_01_Lopez_Buendia_Corpus.20110614.pdf>. [23/07/2011].

López, C.I., M. Buendía, and A. García. 2012. “User Needs to the Test: Evaluating a Terminological Knowledge Base on the Environment by Trainee Translators.” *Jostrans. The Journal of Specialized Translation* (18): 57–76. <http://www.jostrans.org/issue18/art_lopez.pdf>. [20/12/2012].

Lüdeling, A., S. Evert, and M. Baroni. 2007. “Using Web Data for Linguistic Purposes.” In Hund, M., N. Nesselhauf, and C. Biewer (eds.) *Corpus Linguistics and the Web*. 7–24. Amsterdam/New York: Rodopi.

L’Homme, M.C. 1998. “Le Statut Du Verbe En Langue De Spécialité Et Sa Description Lexicographique.” *Cahiers De Lexicographie* 73(2): 61–84.

<http://www.ling.umontreal.ca/lhomme/docs/cahiers-lexico-98.PDF>. [22/02/2011].

———. 2009. "A Methodology for Describing Collocations in a Specialized Dictionary." In Nielsen, S. and S. Tarp (eds.) *Lexicography in the 21st Century*. 237–256. Amsterdam/Philadelphia: John Benjamins.

———. 2010. "Designing Terminological Dictionaries for Learners Based on Lexical Semantics: The Representation of Actants." In Fuertes-Olivera, P.A. (ed.) *Specialised Dictionaries for Learners*. 141–153. Berlin/New York: Mouton de Gruyter.

L'Homme, M.C., and P. Leroyer. 2009. "Combining the Semantics of Collocations with Situation-driven Search Paths in Specialized Dictionaries." *Terminology* 15(2): 258–283.

Mairal, R., and F.J. Cortés. 2006. "An Overview of Role and Reference Grammar." In Mairal, R., M.A. Escobar, M.S. Peña, E. Samaniego, and F.J. Cortés (eds.) *Current Trends in Linguistic Theory*. 97–176. Madrid: Universidad Nacional de Educación a Distancia.

Mairal, R., and P. Faber. 2005. "Decomposing Semantic Decomposition: Towards a Semantic Metalanguage in RRG." In *Proceedings of the 2005 International Conference on Role and Reference Grammar*. 279–308. Taiwan: Academia Sinica. <http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/MairalRRG05.pdf>. [25/01/2010].

Mairal, R., and P. Faber. 2007. "Lexical Templates Within a Functional Cognitive Theory of Meaning." *Annual Review of Cognitive Linguistics* 5: 137–172.

Mairal, R., M.S. Peña, F.J. Cortés, and F. Ruiz de Mendoza. 2012. *Teoría Lingüística: Métodos, Herramientas y Paradigmas*. 2nd edition. Madrid: Editorial Universitaria Ramón Areces.

Martín-Mingorance, L. 1984. "Lexical Fields and Stepwise Lexical Decomposition in a Contrastive English-Spanish Verb Valency Dictionary." In Hartmann, R. (ed.) *LEXeter 83: Proceedings of the International Conference on Lexicography*. 226–236. Tubinga: Niemeyer.

———. 1989. "Functional Grammar and Lexematics." In Tomaszcyk, J. and B. Lewandowska (eds.) *Meaning and Lexicography*. 227–253. Amsterdam/ Philadelphia: John Benjamins.

———. 1995. “Lexical Logic and Structrual Semantics: Methodological Underpinnings in the Structuring of a Lexical Database for a Natural Language Processing.” In Hoinkes, U. (ed.) *Parorama Der Lexikalischen Semantik*. 461–474. Tubinga: Gunter Narr.

———. 1998. *El Modelo Lexemático-funcional: El Legado Lingüístico De Leocadio Martín Mingorance. (The Functional Lexematic Model: The Linguistic Legacy of Leocadio Martín Mingorance)*. Ed. Amalia Marín Rubiales. Granada: University of Granada.

Mateo-Martínez, J., and E. Alcaraz-Varó. 2003. *Diccionario De Términos De La Bolsa: inglés-español, Spanish-English*. Barcelona: Ariel.

Matsubayashi, Y., N. Okazaki, and J.I. Tsujii. 2009. “A Comparative Study on Generalization of Semantic Roles in FrameNet.” In *ACL’09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. 19–27. Stroudsburg, PA: Association for Computational Linguistics.

McEnery, T., and A. Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

———. 2004. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McGee, I. 2012. “Collocation Dictionaries as Inductive Learning Resources in Data-driven Learning: an Analysis and Evaluation.” *International Journal of Lexicography* 25(3): 319–361.

McGuire, B., I. Mason, and C. Kilburn. 2002. *Natural Hazards and Environmental Change*. New York: Oxford University Press.

McIntosh, C., B. Francis, and R. Poole. 2009. *Oxford Collocations Dictionary for Students of English*. 2nd edition. Oxford: Oxford University Press.

Mel’čuk, I. 1988. “Semantic Description or Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria.” *International Journal of Lexicography* 1(3): 165–188.

———. 1995. “Phrasemes in Language and Phraseology in Linguistics.” In Everaert, M., E.J. Van der Linden, A. Schenk, and R. Schreuder (eds.) *Idioms: Structural and Psychological Perspectives*. 167–232. Hillsdale, NJ: Lawrence Erlbaum Associates.

———. 1996. "A Tool for the Description of Lexical Relations in the Lexicon." In Wanne, L. (ed.) *Lexical Functions in Lexicography and Natural Language Processing*. 37–71. Amsterdam/ Philadelphia: John Benjamins.

———. 1998. "Collocations and Lexical Functions." In Cowie, A.P. (ed.) *Phraseology. Theory, Analysis, and Applications*. 23–53. Oxford: Clarendon Press.

———. 2004. "Actants in Semantics and Syntax. I." *Linguistics* 42(1): 1–66.

Mel'čuk I. et al. 1984-1999. *Dictionnaire explicatif et combinatoire du français contemporain. Recherche lexico-sémantiques I, II, III, IV*. Montréal: Les Presses de l'Université de Montréal.

Mel'čuk, I, A. Clas, and A. Polguère. 1995. *Introduction à La Lexicologie Explicative Et Combinatoire*. Lovain-la-Neueve, Belgium: Duculot/Aupelf-UREF.

Mel'čuk, I., and A. Polguère. 2007. *Lexique Actif Du Français*. Brussels: Duculot.

Mel'čuk, I., and A. Zholkovskij. 1970. "Towards a Functioning Meaning-Text Model of Language." *Linguistics* 57: 10–47.

Meyer, C.F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press. <http://catdir.loc.gov/catdir/samples/cam033/2001052491.pdf>. [02/04/2011].

Meyer, I. 2001. "Extracting Knowledge-rich Contexts for Terminography. A Conceptual and Methodological Framework." In Bourigalt, D., C. Jacquemin, and M.C. L'Homme (eds.) *Recent Advances in Computational Terminology*. 279–302. Amsterdam/Philadelphia: John Benjamins.

Meyer, I., L. Bowker, and K. Eck. 1992. "COGNITERM: An Experiment in Building a Knowledge-based Term Bank." In *Proceedings of the Fifth EURALEX International Congress*. 159–172. Tampere, Finland: EURALEX.

Meyer, I, K. Eck, and D. Skuce. 1997. "Systematic Concept Analysis Within a Knowledge-based Approach to Terminology." In Wright, S.E., and G. Buedin (eds.) *Handbook of Terminology Management. Volume I. Basic Aspects of Terminology Management*. 98–118. Amsterdam/ Philadelphia: John Benjamins.

Meyer, I., and K. Mackintosh. 1996. "Refining the Terminographer's Concept-analysis Methods: How Can Phraseology Help?" *Terminology* 3(1): 1–26.

Meynard, I. 2000. *Internet. Répertoire Bilingue De Combinaisons Lexicales Spécialisées Français-Anglais*. Montreal: Linguatech.

Michiels, A. 1975. "Idiomaticity in English." *Revue Des Langues Vivantes* 43(2): 184–189.

Miller, G.A. 1986. "Dictionaries in the Mind." *Language and Cognitive Processes* 1(3): 171–185.

Model, B. 2008. "Review of the Book BOSQUE, I. (2006): Diccionario Combinatorio Práctico Del Español Contemporáneo, Madrid, SM." *Revista Lexicografia* XXIII: 196. <http://ruc.udc.es/dspace/bitstream/2183/5520/2/RL_13_REC-2.pdf.txt>. [24/01/2012].

Moliner, M. 2008. *Diccionario de uso del español*. 2nd edition. Madrid: Gredos.

Montero, S. 2002. *Estructuración Conceptual y Formalización Terminográfica De Frasemas En El Subdominio De La Oncología*. Ph.D. Thesis presented at the University of Valladolid, Spain. <http://www.cervantesvirtual.com/FichaObra.html?Ref=14547>. [22/02/2010].

———. 2008. "A Constructional Approach to Terminological Phrasemes." In Bernal, E., and J. DeCesaris (eds.) *Proceedings of the XIII International EURALEX Congress*. 1015–1022. Barcelona: IULA.

Montero, S., and M. Buendía. 2010. "Las Construcciones Fraseológicas en La Terminología basada en Marcos." In *Plenary Session Presented at the Seminario De Lingüística Española: Los Retos Informáticos Del Español En Lexicología, Terminología y Fraseología*, 26–27 Februrary. Antwerpen: Fundación Duques de Soria and the University of Antwerpen.

———. 2012. "La Sistematización En El Tratamiento De Las Construcciones Fraseológicas: El Caso Del Medio Ambiente." In Elorza, I., O. Carbonell, R. Albarrán, B. García, and M. Pérez (eds.) *Empiricism and Analytical Tools for 21st Century Applied Linguistics*. 711–724. Salamanca: Ediciones de la Universidad de Salamanca.

Montero, S., P. Faber, and M. Buendía. 2011. *Terminología Para Traductores e Intérpretes*. Granada: Ediciones Tragacanto.

Moreno, A.J. 1998. *Diseño e Implementación De Un Lexicón Computacional Para Lexicología y Traducción Automática*. Ph.D. Thesis presented at the University of Córdoba, Spain.

Nakamura, J. 1993. "Statistical Methods and Large Corpora. A New Tool for Describing Text Types." In Baker, M., G. Francis, and E. Tognini-Bonelli (eds.) *Text and Techology: In Honour of John Sinclair*. 293–312. London: John Benjamins.

United Nations. 2009. *UNISDR Terminology on Disaster Risk Reduction*. Geneva: United National International Strategy for Disaster Reduction.

Nesselhauf, N. 2004. "What Are Collocations?" In Allerton, D.J., N. Nesselhauf, and P. Skandera (eds.) *Phraseological Units: Basic Concepts and Their Application*. 1–21. Basel: Schwabe Verlag.

Nielsen, S. 2010. "Specialized Translation Dictionaries for Learners." In Fuertes-Olivera, P.A. (ed.) *Specialized Dictionaries for Learners*. 69–82. Berlin/New York: De Gruyter.

Niles, I., and A. Pease. 2001. "Toward a Standard Upper Ontology." In Welty, C., and B. Smith (eds.) *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001).*

Nkwenti-Azeh, B. 1998. "Information Mediation: The Interface Between Terminology and Translation." In Kent, A. (ed.) *Encyclopedia of Library and Information Science.* Vol. 62. 157–170. New York/Basel/Hong Kong: Marcel Dekker, Inc.

Ntoulas, A., J. Cho, and C. Olston. 2004. "What's New on the Web? The Evolution of the Web from a Search Engine Perspective." In Feldman Stuart, I., M. Uretsky, M. Najork, and C.E. Wills (eds.) *Proceedings of the 13th International Conference on World Wide Web*. 1–12. New York: ACM Press.

Nuccorini. 2003. "Towards an Ideal Dictionary of Collocations." In Van Sterkenburg, P. (ed.) *A Practical Guide to Lexicography*. 366–387. Amsterdam/ Philadelphia: John Benjamins.

Olcina, J. 2006a. *¿Riesgos Naturales? Vol.1*. Barcelona: Davinci.

———. 2006b. *¿Riesgos Naturales? Vol.2*. Barcelona: Davinci.

ONCOTERM. 2002. "Sistema Bilingüe De Información y Recursos Oncológicos." <http://www.ugr.es/~oncoterm/oncodesc.htm>. [22/01/2011].

Orliac, B. 2004. *Automatisation Du Repérage Et De L'encodage Des Collocations En Langue De Spécialité*. Ph.D. Thesis presented at the University of Montreal, Canada. <http://olst.ling.umontreal.ca/pdf/PhDOrliac.pdf>. [22/01/2010].

Palmer, H.E. 1933. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.

Palmer, M., D. Gildea, and P. Kingsbury. 2005. "The Proposition Bank: An Annotated Corpus of Semantic Roles." *Computational Linguistics* 31(1): 71–105.

Park, O.C., and R. Hopkins. 1993. "Instructional Conditions for Using Dynamic Visual Displays: A Review." *Instructional Science* 22: 1–24.

Pavel, S., and M. Boileau. 2003. *Vocabulaire Combinatoire De L'imagerie Fractale - Combinatory Vocabulary of Fractal Imagery*. Canada: Bureau de la Traduction. <http://www.btb.gc.ca/publications/documents/fractal.pdf>. [29/04/2012].

Pawley, A. 2001. "Phraseology. Linguisitcs and the Dictionary." *International Journal of Lexicography* 14(2): 122–134.

Pazienza, M.T., M. Pennacchiotti, and F.M. Zanzotto. 2006. "Mixing WordNet, VerbNet and PropBank for Studying Verb Relations." In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. 1372–1377. Genova, Italy. <http://www.lrec-conf.org/proceedings/lrec2006/pdf/379_pdf>. [25/01/2010].

Pearson, J. 1998. *Terms in Context*. Amsterdam/ Philadelphia: John Benjamins.

Penadés, I. 2001. "¿Colocaciones o Locuciones Verbales?" *Lingüística Española Actual* XXIII: 57–88.

———. 2012. "La Fraseología y Su Objeto De Estudio." *Linred: Lingüística En La Red*. <http://www.linred.es/monograficos_pdf/LR_monografico10-articulo2.pdf>. [10/11/2012].

Perera, I. 2012. "Semantic Parsing." <http://www.seas.upenn.edu/~muri/y3review/SemanticParsing.pdf>. [22/02/2012].

Peters, C., E. Picchi, and L. Biagini. 1996. "Parallel and Comparable Bilingual Corpora in Language Teaching and Learning." In Botley, S., J. Glass, T. McEnery, and A. Wilson (eds.) *Proceedings of Teaching and Learning*. 68–82. Lancaster: UCREL.

Pimentel, J. 2012. *Criteria for the Validation of Specialized Verb Equivalents: Application in Bilingual Terminography*. Ph.D. Thesis presented at the University of Montreal, Canada. <http://olst.ling.umontreal.ca/pdf/Pimentel_J_thesis_2012.pdf> [15/09/2012].

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. New York: William Morrow.

Polguère, A. 2000. "Towards a Theoretically-motivated General Public Dictionary of Semantic Derivations and Collocations for French." In *Proceedings of the 9th EURALEX International Congress*. 517–528. Stuttgart: EURALEX <http://olst.ling.umontreal.ca/pdf/APolEURALEX.pdf>. [25/01/2010].

Prieto, J.A. 2009. *Traducción e Imagen: La Información Visual En Textos Especializados*. Granada: Tragacanto.

Prieto, J.A., and C.I. López. 2009. "Managing Graphic Information in Terminological Knowledge Bases." *Terminology* 15(2): 179–213.

Pérez, C. 2002. E*xplotación De Los Córpora Textuales Informatizados Para La Creación De Bases De Datos Terminológicas Basadas En El Conocimiento*. Ph.D. Thesis presented at the University of Málaga. <http://elies.rediris.es/elies18>. [07/12/2009].

Randall, J.H. 1983. "A Lexical Approach to Causatives." *Journal of Linguistics Research* 2(3): 77–105.

Rappaport Hovav, M., and B. Levin. 1998. "Building Verb Meaning." In Butt, M. and W. Geuder (eds.) *The Projection of Arguments. Lexical and Compositional Factors*. 96–124. California: CSLI Publications Stanford.

Real Academia Española. 2001. *Diccionario De La Real Academia Española*. 22nd edition. Madrid: Real Academia Española. <http://www.rae.es> [02/01/2009].

Reimerink, A., M. García, and S. Montero. 2010. "Contextual Information in Terminological Knowledge Bases: A Multimodal Approach." *Journal of Pragmatics* 42(7): 1928–1950.

Rey, A. 1975. *La Terminologie: Noms Et Notions*. Paris: Presses Universitaires de France.

Rieber, L.P. 1994. *Computers, Graphics, and Learning*. Madison, WI: Brown & Benchmark.

Rogers, M. 2004. "Multidimensionality in Concept Systems: a Bilingual Textual Perspective." *Terminology* 10(2): 215–240.

Rosch, E. 1973a. "Natural Categories." *Cognitive Psychology* 4(3): 328–350.

———. 1973b. "On the Internal Structure of Perceptual and Semantic Categories." In Moore, T. (ed.) *Cognitive Development and the Acquisition of Language*. 111–144. New York: Academia Press.

Rosch, E., and B.B. Lloyd. 1978. *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Rossenbeck, K. 1989. "Lexikologische Und Lexikographische Probleme Fachsprachlicher Phraseologie Aus Kontrastiver Sicht." In Snell-Hornby, M., and E. Pöhl (eds.) *Translation and Lexicography*. 197–210. Amsterdam/ Philadelphia: John Benjamins.

Roussy, K.R. 2006. "Hurricanes." *University of Illinois at urbana-Champaign Website*. <http://www.atmos.illinois.edu/earths_atmosphere/hurricanes.html>. [05/01/2011].

Ruiz de Mendoza, F.J. 1999. *Introducción a La Teoría Cognitiva De La Metonimia*. Granada: Método Ediciones.

Ruiz de Mendoza, F.J., and R. Mairal. 2008. "Levels of Description and Constraining Factors in Meaning Construction: An Introduction to the Lexical Constructional Model." *Folia Lingüística* 42(2): 355–400.

Ruiz-Gurillo, L. 1997. *Aspectos De Fraseología Teórica Española (Anejo n^{o}XXIV De La Revista Cuadernos De Filología)*. Valencia: Servicio de Publicaciones de la Universidad.

Rundell, M. 2010. *Macmillan Collocations Dictionary for Learners of English*. Oxford: Macmillan Publishers Ltd.

Ruppenhofer, J., M. Ellsworth, M.R.L. Petruck, C.R. Johnson, and J. Scheffczyk. 2006. *FrameNet II: Extendend Theory and Practice*. Berkeley, California: International Computer Science Institute. <http://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>. [27/01/2011].

Sager, J.C. 1990. *A Practical Course in Terminology Processing*. Amsterdam/ Philadelphia: John Benjamins.

———. 1992. "Future Developments and Research in Phraseology and Terminology Related to Translation." *Terminologie Et Traduction* 2(3): 583–588.

———. 1997. "Term Formation." In Wright, S.E., and G. Budin (eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Vol.1. 25–41. Amsterdam/Philadelphia: John Benjamins.

Sanz, Mª.L. 2011. *Análisis Contrastivo De La Terminología De La Teledetección. La Traducción De Compuestos Sintagmáticos Nominales Del Inglés Al Español*. Ph.D. Thesis presented at the University of Salamanca, Spain.

Saussure, F. [1916] 1990. *Cours De Linguistique Générale*. Paris: Payot.

Schmidt, T. 2009. "The Kicktionary - a Multilingual Lexical Resources of Football Language." In Boas, H.C. (ed.) *Multilingual Framenets in Computational Lexicography*. 101–134. New York: Mouton de Gruyter.

Scott, M. 2008. "WordSmith Tools Version 5". Liverpool: Lexical Analysis Software.

Sebastián, N., M.A. Martí, M.F. Carreiras, and F. Cuetos. 2000. *LEXESP. Léxico Informatizado Del Español*. Barcelona: Ediciones de la Universitat de Barcelona.

Seco, M. 1981. "María Moliner, Una Obra No Un Nombre." *El País, 35*. <http://www.mariamoliner.com/centenario_2.pdf>. [22/04/2010].

Seghiri, M. 2006. *Compilación De Un Corpus Trilingüe De Seguros Turísticos (español-inglés-italiano): Aspectos De Evaluación, Catalogación, Diseño y Representatividad*. Ph.D. Thesis presented at the University of Málaga, Spain.

Shi, L., and R. Mihalcea. 2005. "Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing." In *Proceedings of CICLingu-2005*, 100–111.

Siepmann, D. 2005. "Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects." *International Journal of Lexicography* 18(4): 409–443.

Sinclair, J. 1987. "Introduction." In *Collins Cobuild English Language Dictionary*, xv–xxi. London and Glasgow: Collins.

———. 1995. "Corpus Typology: A Framework for Classification." In Melchers, G., and B. Warren (eds.) *Studies in Anglistics*. 17–34. Stockhom: Almquist and Wiksell International.

———. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

———. 2000. "Lexical Grammar." *Darbai Ir Dienos*. <http://donelaitis.vdu.lt/publikacijos/sinclair.pdf>. [08/02/2010].

———. 2004. "Intuition and Annotation -the Discussion Continues." In Aijmer, K., and B. Altenberg (eds.) *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23) Göteborg 22-26 May 2002*. 39–61. Amsterdam/New York: Rodopi.

———. 2005. "Corpus and Text: Basic Principles." In Wynne, M. (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. 1–16. Oxford: Oxbow Books. <http://ota.ahds.ac.uk/documents/creating/dlc/appendix.htm>. [08/08/2009].

Staab, S., and R. Studer. 2009. *International Handbooks on Information Systems*. 2nd edition. London: Springer Verlag.

Stubbs, M. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.

Studer, R., R. Benjamins, and D. Fensel. 1998. "Knowledge Engineering: Principles and Methods." *Data and Knowldege Engineering* 25(1-2): 161–198.

Subirats, C. 1984. *Las Completivas Sin Preposición En La Frase De Dos Argumentos En Castellano. Resumen De Tesis*. Bellaterra: Universidad Autónoma de Barcelona.

———. 1987. *Sentential Complementation in Spanish: a Lexico-grammatical Study of Three Classes of Verbs*. Amsterdam/ Philadelphia: John Benjamins.

———. 2001. *Introducción a La Sintaxis Léxica Del Español*. Madrid: Vervuert.

———. 2006. "El Concepto De Predicado En La Tradición Gramatical y Lexicográfica." *Estudios De Lingüística Del Español (ELiEs)2* 23. <http://elies.rediris.es/elies23/subirats.htm>. [19/10/2009].

Sánchez, B., and M. Buendía. 2012. "Inclusion of Verbal Syntagmatic Patterns in Specialized Dictionaries: The Case of EcoLexicon." In Vatvedt Fjeld, R., and J. Matilde Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress*. 554–562. Oslo: EURALEX.

Talmy, L. 1976. "Semantic Causative Types." In Shibatini, M. (ed.) *Syntax and Semantics 6: The Grammar of Causative Constructions*. 43–116. New York: Academic Press.

———. 2000a. *Toward a Cognitive Semantics, Vol. 1: Concept Structuring Systems*. Cambridge, Mass.: MIT Press.

———. 2000b. *Cognitive Semantics, Vol. 2: Typology and Process in Concept Structuring*. Cambridge, Mass.: MIT Press.

Tarp, S. 2006. "Lexicografía De Aprendizaje." In Xatara, C., and P. Humblé (eds.) *Traduçao e Lexicografia Pedagógica*. 295–317. Florianópolis: Universidade Federal de Santa Catarina.

———. 2007. "¿Qué Requisitos Debe Cumplir Un Diccionario De Traducción Del Siglo XXI?" In Fuertes-Olivera, P.A. (ed.) *Problemas Lingüísticos En La Traducción Especializada*. 227–256. Valladolid: Universidad de Valladolid.

———. 2008a. *Lexicography in the Borderland Between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer Verlag.

———. 2008b. "The Third Leg of Two-legged Lexicography." *Hermes. Journal of Language and Communication Studies* 40: 117–131.

———. 2010. "Functions of Specialized Learners'Dictionaries." In Fuertes-Olivera, P.A. (ed.) *Specialized Dictionaries for Learners*. 39–54. Berlin/New York: De Gruyter.

Temmerman, R. 1997. "Questioning the Univocity Ideal. The Difference Between Socio-cognitive Terminology and Traditional Terminology." *Hermes, Journal of Linguistics* 18: 51–90.

———. 1998a. *Terminology Beyond Standardisation. Language and Categorisation in the Life Sciences*. Ph.D. Thesis presented at the University of Leuven, Belgium.

———. 1998b. "Terminology Theory and Terminography in a Natural Language Processing Environment." *Revue Française De Linguistique Appliqué* 3(2): 29–46.

———. 2000a. *Towards New Ways of Terminology Description*. Amsterdam/ Philadelphia: John Benjamins.

———. 2000b. "Why Traditional Terminology Theory Impedes a Realistic Description of Categories and Terms in the Life Sciences." *Terminology* (special issue): 73–78.

———. 2001. "Sociocognitive Terminology Theory." In Cabré, Mª.T., and J. Feliu (eds.) *Terminología y Cognición*. 75–92. Barcelona: Universitat Pompeu Fabra.

Temmerman, R., and K. Kerremans. 2003. "Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description." In Hajicová, E., A. Kotešovcová, and J. Mírovský (eds.) *Proceedings of CIL17, Matfyzpress, MFF UK*. Prague, Czech Republic. <http://taalkunde.ehb.be/sites/www2.ehb.be/files/u96/temmerman_art_prague03.pdf>. [05/01/2010].

Tercedor, M. 1999. *La Fraseología En El Lenguaje Biomédico: Análisis Desde Las Necesidades Del Traductor*. Ph.D. Thesis presented at the University of Granada, Spain. <http://elies.rediris.es/elies6/>. [25/07/2009].

Tercedor, M., P. Faber, and C.I. López. 2012. "Working with Words: Research Approaches in Translation-oriented Lexicographic Practice." *TTR: Traduction, Terminologie, Rédaction* XXV(1): 181–214.

Teubert, W. 1996. "Comparable or Parallel Corpora." *International Journal of Lexicography* 9(3): 238–264.

Thomas, P. 1993. "Choosing Headwords from Language-for-Special-Purposes (LSP) Collocations for Entry into a Terminology Data Bank (Term Bank)." In Sonneveld, H.B., and K.L. Loening (eds.) *Terminology: Applications in Interdisciplinary Communication*. 43–68. Amsterdam/Philadelphia: John Benjamins.

Tisdal, D. 2012. *Hurricane FAQS. Travel Insurance Review: The Web's Best Travel Insurance Resource*. <http://www.travelinsurancereview.net/hurricane-travel-resource-center/hurricane-faqs/> [08/08/2012].

Tolone, E., S. Voyatzi, and C. Leclère. 2010. "Constructions Définitoires Des Tables Du Lexique-Grammaire." In *Actes Du 29ème Colloque International Sur Le*

Lexique Et La Grammaire. 321–331. Belgrade, Serbia: Faculty of Mathematics of University of Belgrade.

Vaamonde, G., F. González, and J.Mª. García. 2010. "ADESSE. A Database with Syntactic and Semantic Annotation of a Corpus of Spanish." In Calzolari, N., K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner,and D. Tapias (eds.) *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. 1903–1910. Valletta: European Language Resources Association (ELRA).

Van Valin, R.D.Jr. 2005. *The Syntax-semantics-pragmatics Interface: An Introduction to Role and Reference Grammar*. Cambridge: Cambridge University Press.

Van Valin, R.D.Jr., and R. LaPolla. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaca, New York: Cornell University Press.

Verlinde, S., J. Folon, J. Binon, and J. Van Dyck. *Dictionnaire Contextuel Du Français Économique. (4 Tomes)*. Antwerpen: Garant.

Vincze, O., E. Mosqueira, and M. Alonso. 2011. "An Online Collocation Dictionary of Spanish." In Boguslavsky, I. and L. Wanner (eds.) *Proceedings of the 5th International Conference on Meaning-Text Theory*. 275–286. Barcelona. <http://www.dicesp.com/app/webroot/files/file/MTT 2011 Vincze et al_.pdf>. [22/02/2012].

Vinogradov, V.V. 1947. *Ob Osnovnij Tipaj Fraseologicheskij Iedinits v Ruskom Yazike*. Moscow.

Volk, Martin. 2002. "Using the Web as Corpus for Linguistic Research." In Pajusalu, R. and T. Hennoste (eds.) *Tähendusepüüjda. Hatcher of the Meaning. A Fetshcrift for Professor Haldur Õim*. Tartu: University of Tartu.

Vossen, P. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht, Holland: Kluwer.

Vázquez, G., and A. Fernández Montraveta. 2011. "Paralelización Del Corpus SenSem: Español-catalán." *Anuari De Filologia. Estudis De Lingüística* 1: 167–193. <http://revistes.ub.edu/index.php/AFEL/article/view/2252/2399>. [05/07/2012].

Walker, C. 2009. "Dictionaries, Collocational Dictionaries and Dictionaries of Business English." *International Journal of Lexicography* 22(3): 281–299.

Wanner, L. 1996. *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/ Philadelphia: John Benjamins.

Weinreich, U. 1969. "Problems in the Analysis of Idioms." In Puhvel, J. (ed.) *Substance and Strucutre of Language*. 23–81. Berkeley and Los Angeles: University of California Press.

Wierzbicka, A. 1996. *Semantics: Primes and Universals*. Oxford: Oxford University Press.

———. 1999. *Emotions Across Languages and Cultures. Diversity and Universals*. Cambridge: Cambridge University Press.

Wilkinson, G.L., L.T. Bennett, and K.M. Oliver. 1997. "Evaluating the Quality of Internet Information Sources." In *Paper Presented at ED-MEDIA/ED-TELECOM*. Calgary, Alberta, Canada. <http://www.eric.ed.gov/PDFS/ED412927.pdf>. [09/08/2009].

Wisner, B., P. Blaikie, T. Cannon, and I. Davis. 2004. *At Risk. Natural Hazards, People's Vulnerability and Disasters*. Second Edition. London and New York: Routledge.

Wittgenstein, L. 1968. *Philosophical Investigations*. Oxford: Blackwell.

Wray, A. 2000. "Formulaic Sequences in Second Language Teaching: Principle and Practice." *Applied Linguistics* 21(4): 463–489.

Wright, S.E. 1997. "Representation of Concept Systems." In Wright, S.E., and G. Budin (eds.) *Handbook of Terminology Management*. Vol.1. 89–97. Amsterdam/Philadelphia: John Benjamins.

Wüster, E. 1968. *The Machine Tool. An Interlingual Dictionary of Basic Concepts*. London: Technical Press.

———. [1979] 1998. *Introducción a La Teoría General De La Terminología y a La Lexicografía Terminográfica*. Barcelona: IULA.

You, L., and K. Liu. 2005. "Building Chinese FrameNet Database." In *Proceedings of 2005 IEEE International Conference on Processing and Knowledge Engineering (IEEE NLP-KE'05)*. 301–306.

Zanettin, F., S. Bernardini, and D. Stewart. 2003. *Corpora in Translator Education*. Manchester: St. Jerome.

Zuluaga, A. 1980. *Introducción Al Estudio De Las Expresiones Fijas*. Frankfurt: Verlag.

———. 1997. “Sobre Las Funciones De Unidades Fraseológicas En Textos Literarios.” *Paremia* 6: 631–640.

Appendix 1. Definitions of English and Spanish verbs classified in lexical domains

EXISTENCE
to begin to exist
to begin to exist from sth else
to begin to exist becoming sth else
to cause to exist/happen
to prevent sth from existing/happening
to be likely to happen (especially sth bad)
to exist in time
to continue to exist in time (of natural disaster)
to continue to exist (of people)
to cease to exist
to cease to exist in the perception of others
to cause sb to cease to exist
to cause to cease to exist (of fire disaster)

Table 145. Frames of the lexical domain of EXISTENCE

To begin to exist (English)

start2: *to begin to exist.*

break out: *to start suddenly (esp. of undesirable events).*

originate: *to begin to exist.*

develop: *to begin to exist becoming larger/fuller/better.*

evolve: *to develop gradually.*

blow up: *to begin to develop (esp. of a storm or wind event).*

form: *to begin to exist and have a shape.*

burst2: *to begin to exist (esp. of volcanic events).*

erupt1: *to begin to exist (esp. of volcanic events).*

explode2: *to begin to exist (esp. of volcanic events).*

To begin to exist (Spanish)

empezar: *dar comienzo a algo.*

despertar: *empezar (especialmente actividad volcánica).*

comenzar: *empezar a existir.*

originarse: *empezar a existir, normalmente a partir de algo.*

desarrollarse: *empezar a existir, aumentando o acrecentando.*

formarse: *empezar a existir y desarrollarse.*

To begin to exist from sth else (English)

originate from: *to begin to exist from sth else.*

develop from: *to begin to exist from sth else and becoming larger/fuller/better.*

evolve from: *to develop from sth gradually.*

form from: *to begin to exist from sth else and have a shape.*

To begin to exist from sth else (Spanish)

originarse a partir de: *empezar a existir a partir de otro fenómeno.*

desarrollarse (a partir) de: *empezar a existir a partir de otro fenómeno más grande.*

formarse de: *empezar a existir a partir de otro fenómeno y desarrollarse.*

To begin to exist becoming sth else (English)

develop into: *to begin to exist becoming sth else.*

evolve into: *to develop into sth gradually.*

To begin to exist becoming sth else (Spanish)

transformarse en: *hacer que algo empiece a existir siendo algo distinto a lo que empezó.*

convertirse en: *transformarse en algo.*

evolucionar a: *convertirse en algo gradualmente.*

To cause to exist/happen (English)

cause: *to make sth happen.*

start1: *to cause sth to exist/happen.*

form2: *to cause to exist/happen.*

generate: *to cause sth to begin to exist.*

produce: *to cause sth to begin to exist.*

spawn: *to cause sth new or many new things, to grow or start suddenly.*

result (from/in): *to cause to exist/happen (as a consequence of sth/becoming sth else).*

trigger: *to cause an event or situation to happen or exist.*

spread2: *to cause a fire event to exist/happen.*

To cause to exist/happen (Spanish)

causar: *hacer que algo ocurra.*

generar: *causar.*

producir1: *causar.*

provocar: *producir algo como respuesta.*

ocasionar: *(referido especialmente a un suceso), causarlo o ser su origen.*

originar: *(referido a un hecho), causarlo o dar lugar a él.*

To prevent sth from existing/happening (English)

prevent: *to stop sth from happening.*

avert: *to prevent sth from happening by taking action.*

To prevent sth from existing/happening (Spanish)

impedir: *(referido especialmente a una acción), dificultar o imposibilitar su realización.*

evitar: *impedir que suceda (especialmente algo malo).*

prevenir: *evitar un mal.*

To be likely to happen (especially sth bad) (English)

threaten1: *to be likely to happen (esp. of sth bad).*

risk: *to threaten.*

endanger: *to cause to be threatened.*

To be likely to happen (especially sth bad) (Spanish)

amenazar1: *(referido a algo malo), haber indicios de que va a ocurrir.*

To exist in time (English)

happen: *to be/become real in time.*

take place: *to happen in a particular place or at a given time.*

occur: *to happen (esp. of unplanned events).*

recur: *to happen many times or to happen again.*

To exist in time (Spanish)

producirse: *originarse, ocasionar, causar.*

ocurrir: *producirse un hecho.*

tener lugar: *ocurrir.*

suceder: *producirse un hecho (más formal).*

sobrevenir: *suceder algo de forma inesperada.*

pasar: *producirse un hecho.*

To continue to exist (of natural disaster) (English)

last: *to continue to exist in time.*

persist: *to continue to exist esp. after trying to make it disappear.*

extent2: *to continue to exist, especially the effects, consequences or sth derived from natural disasters.*

To continue to exist (of natural disaster) (Spanish)

durar: *prolongarse o extenderse en el tiempo.*

perdurar: *durar mucho.*

persistir: *durar por largo tiempo.*

extenderse2: *durar cierto tiempo.*

To continue to exist (of people) (English)

survive: *to continue to live or exist, after being in a difficult or dangerous situation.*

recover from: *to return to a normal state of health, mind or strength.*

To continue to exist (of people) (Spanish)

sobrevivir a: *vivir después de determinado suceso.*

recuperarse de: *volver a un estado de normalidad después de haber pasado por una situación difícil.*

To cease to exist (English)

die: *to cease to exist.*

perish: *to die as a result of very hard conditions.*

drown1: *to die because of being under the water and unable to breathe.*

disappear1: *to die by ceasing to be visible.*

To cease to exist (Spanish)

morir: *dejar de vivir.*

perecer: *morir, especialmente si es de forma violenta.*

fallecer: *morir (más formal).*

ahogarse: *morirse por no poder respirar, especialmente por estar debajo de agua.*

desaparecer1: *dejar de existir.*

To cease to exist in the perception of others (English)

disappear2*: to cease to exist in the perception of others.*

dissipate: *to disappear gradually.*

blow out1: *to disappear by finally losing its force (esp. of wind events).*

To cease to exist in the perception of others (Spanish)

desparecer2*: dejar de existir ante la percepción de otros.*

disiparse: *hacer desaparecer.*

To cause sb to cease to exist (English)

kill: *to cause to cease to exist.*

drown2: *to kill sb by keeping them under water so that they cannot breathe*.

To cause sb to cease to exist (Spanish)

matar: *hacer que alguien deje de existir.*

To cause to cease to exist (of fire disaster) (English)

extinguish: *to cause a fire to cease to exist.*

douse: *to extinguish a fire.*

To cause to cease to exist (of fire disaster) (Spanish)

apagar: *(referido al fuego), hacer que termine.*

extinguir: *(referido al fuego), apagarlo o hacer que cese.*

sofocar: *extinguir un fuego.*

CHANGE
to become different by going from one state to another
to become larger in size/amount/number/degree
to become less in size/amount/degree/importance/intensity
to cause sth to become less in size
to cause sth to increase in moisture
to cause sth to lessen in moisture
to cause sth to increase in temperature
to cause sth to decrease in temperature
to cause a change in sth so that it loses its power or control
to cause sth/sb to change for the worse
to begin to be different in the way that is stated
to separate into two or more pieces
to cause sth not to be changed by sth else

Table 146. Frames of the lexical domain of CHANGE

To become different by going from one state to another (English)

melt: *to become softer or more liquid (esp. of sth solid).*

liquefy: *to become liquid.*

condense: *to become liquid by becoming cooler.*

To become different by going from one state to another (Spanish)

condensar: *(referido a un cuerpo gaseoso), convertirlo en líquido o en sólido.*

derretir: *(referido a algo sólido), convertirlo en líquido por medio del calor.*

fundir: *(referido a un cuerpo sólido, especialmente a un metal), derretirlo y convertirlo en líquido.*

solidificar: *(referido especialmente a un líquido), hacerlo sólido.*

helar: *(referido especialmente a un líquido) solidificarlo o congelarlo por la acción del frío, especialmente el agua en hielo.*

evaporar: *(referido especialmente a un líquido), convertirlo en vapor.*

To become larger in size/amount/number/degree (English)

increase: *to become larger in size/amount/number/degree.*

accumulate1: *to gradually increase in number or amount.*

strengthen: *to increase sth in force/power.*

build up: *to increase in strength or intensity.*

intensify: *to increase sth in force/amount/degree.*

surge: *to increase suddenly and strongly*.

amplify: *to increase the size or effect of sth.*

accelerate: *to increase in speed.*

rise: *to increase in force*.

To become larger in size/amount/number/degree (Spanish)

aumentar: *hacer mayor en tamaño, en cantidad o en intensidad.*

amplificar: *(referido a la intensidad de un fenómeno físico, especialmente el sonido) aumentarla por procedimientos técnicos.*

acrecentar: *aumentar.*

intensificar: *aumentar la intensidad de algo.*

intensar: *intensificar*.

acentuar: *intensificar algo.*

activar1: *aumentar, avivar o acelerar la intensidad o la rapidez de algo.*

reforzar: *aumentar la fuerza de algo.*

acelerar: *aumentar la velocidad de algo.*

To become less in size/amount/degree/importance/intensity (English)

lessen: *to become less in size/amount/degree/importance/intensity.*

mitigate: *to lessen the effects of sth bad or harmful.*

decrease: *to lessen in size/amount/degree/intensity (esp. in numbers).*

fall: *to lessen in size, amount or strength, becoming lower*.

reduce: *to lessen in size, amount, degree, importance, etc*.

damp: *to reduce the flow of air of a fire in order to make it less strong.*

absorb1: *to reduce the effect of a physical force, shock or change.*

minimize: *to reduce sth to the least possible level or amount.*

log: *to reduce an area of forest in order to exploit the timber commercially*.

weaken: *to lessen the intensity of sth.*

To become less in size/amount/degree/importance/intensity (Spanish)

disminuir: *hacer o hacerse menor en tamaño, en cantidad o en intensidad.*

mitigar: *disminuir, moderar o hacer más suave o más soportable.*

aliviar: *mitigar algo.*

paliar: *(referido especialmente a algo negativo), aliviarlo.*

reducir: *disminuir en tamaño, en cantidad o en intensidad.*

comprimir: *reducir a menor volumen.*

minimizar: *reducir a lo mínimo.*

angostar: *reducir o estrechar.*

descender: *reducir en intensidad, cantidad o valor.*

bajar1: *descender (registro más familiar).*

decrecer: *disminuir en tamaño, en cantidad o en intensidad.*

debilitar: *disminuir en fuerza, energía o resistencia.*

To cause sth to become less in size (English)

erode: *to cause sth to gradually lessen in size (esp. because of wind/sea/weather).*

To cause sth to become less in size (Spanish)

erosionar: *producir desgaste por la acción de agentes externos, especialmente, por el agua y el viento.*

To cause sth to increase in moisture (English)

wet: *to cause sth to increase in moisture.*

To cause sth to increase in moisture (Spanish)

humidificar: *aumentar la humedad de algo.*

To cause sth to lessen in moisture (English)

dry*: to cause sth to lessen in moisture.*

To cause sth to lessen in moisture (Spanish)

secar: *(especialmente referido a las plantas), dejar sin agua, sin líquido pudiendo causar su muerte.*

agostar: *secar las plantas o abrasarlas por el excesivo calor.*

To cause sth to increase in temperature (English)

warm: *to cause sth to increase in temperature.*

To cause sth to increase in temperature (Spanish)

calentar: *comunicar calor, hacienda aumentar la temperatura.*

To cause sth to decrease in temperature (English)

cool: *to cause sth to decrease in temperature.*

To cause sth to decrease in temperature (Spanish)

enfriar: *disminuir o hacer que disminuya la temperatura.*

To cause a change in sth so that it loses its power or control (English)

destabilize: *to cause a change in sth so that it loses its power or control.*

To cause a change in sth so that it loses its power or control (Spanish)

desestabilizar: *(referido especialmente a una situación), perturbar o comprometer su estabilidad.*

To cause sth/sb to change for the worse (English)

affect: *to cause sth to change for the worse.*

damage: *to cause physical harm to sth/sb without destroying them.*

destroy: *to damage sth so much that it does not exist anymore.*

devastate: *to destroy sth (usually an area or place) completely.*

ravage: *to devastate.*

demolish: *to destroy sth (usually a building) completely.*

wreck*: to destroy sth completely (especially as the result of an accident).*

sweep away: *to destroy sth completely and without leaving a trace.*

burn1: *to cause sth to be damaged, injured or destroyed by fire or extreme heat.*

injure: *to damage some part of sb's body.*

deepen: *to cause sth to become worse.*

exacerbate: *to cause sth to become a lot worse (formal).*

To cause sth/sb to change for the worse (Spanish)

afectar: *producir cambios en algo, a menudo perjudicándolo o influyéndolo desfavorablemente.*

dañar: *producir daño.*

castigar: *dañar o estropear, especialmente un fenómeno natural.*

destruir: *(referido especialmente a algo material), dañarlo o deshacerlo totalmente.*

destrozar: *destruir, causando mucho daño.*

asolar: *destruir por completo.*

arrasar: *asolar.*

devastar: *(referido especialmente a un territorio), destruirlo arrasando sus edificios y asolando o echando a perder sus campos.*

demoler: *destruir una construcción echándola abajo.*

derrumbar: *demoler.*

derribar: *demoler.*

derruir: *demoler.*

quemar: *destruir con fuego.*

arder1: *quemar.*

calcinar: *quemar por completo.*

dificultar: *(referido a la consecución de algo), ponerle dificultades o inconvenientes.*

empeorar: *(referido a fenómenos atmosféricos), pasar a un estado peor.*

agravar: *empeorar*.

<u>To begin to be different in the way that it is stated (English)</u>

become: *to begin to be different in the way that it is stated.*

change: *to become different.*

vary: *to change usually continuously, from one occasion to another.*

range: *to vary between specified limits.*

oscillate: *to vary in magnitude or position in a regular manner about a central point.*

modify: *to change slightly, usually becoming better.*

<u>To begin to be different in the way that is stated (Spanish)</u>

cambiar: *hacer que algo sea distinto.*

alterar: *cambiar algo.*

modificar: *cambiar algo sin hacer que varíe mucho.*

transformar: *hacer que algo se convierta en algo distinto.*

variar: *hacer que algo sea diferente o distinto de como era antes.*

oscilar: *variar los límites de una cantidad o de un valor.*

fluctuar: *oscilar de forma más o menos regular***.**

To separate into two or more pieces (English)

break: *to separate suddenly or violently into two or more pieces, or to stop working because of being damaged.*

shear: *to break into pieces, usually because of sideways force (esp. of sth made of metal).*

fracture: *to break or crack (esp. of sth hard, such as a bone).*

breach: *to break in a wall or fence, esp. in order to attack sb or sth behind it.*

ground: *to break sth into small pieces or a powder by pressing between hard surfaces.*

crack*: to break without separating, but with thin lines appearing on the surface.*

rip: *to break violently and quickly.*

rupture: *to break.*

To separate into two or more pieces (Spanish)

romper: *quebrar o hacer pedazos.*

desgarrar: *(referido especialmente a algo de poca consistencia), romperlo o hacerlo pedazos mediante la fuerza y sin ayuda de ningún instrumento.*

fracturar: *(referido especialmente a algo duro y resistente como un hueso), romperlo con violencia o con brusquedad.*

rajar: *romper algo haciendo una abertura larga.*

agrietar: *hacer aberturas largas y estrechas en una superficie.*

To cause sth not to be changed by sth else (English)

resist: *to cause not to be changed by sth.*

withstand: *to resist sth bad successfully.*

To cause sth not to be changed by sth else (Spanish)

resistir: *mantener con fuerza.*

soportar: *(referido a una carga o un peso), resistir algo.*

POSSESSION
to cause sth to have sth as a part
to have sth at one place
to cause sth/sb to have sth
to cause sth not to have sth anymore
to have a large number of sth

Table 147. Frames of the lexical domain of POSSESSION

To cause sth to have sth as a part (English)

include: *to cause sth to have sth as a part.*

incorporate: *to include sth as part of a whole.*

encompass: *to include sth, especially different types of things.*

involve: *to include sth/sb in sth.*

integrate: *to include two or more things together in order to become more effective.*

aggregate: *to include sth as part of sth larger.*

absorb2: *to include sth inside gradually.*

To cause sth to have sth as a part (Spanish)

incluir: *hacer formar parte de algo.*

absorber: *hacer que algo incluya algo, especialmente de forma gradual.*

incorporar: *incluir algo como parte de otra cosa.*

integrar: *incluir algo en otra cosa para formarlo o componerlo.*

agregar: *incluir algo como parte de algo más grande.*

To have sth at one place (English)

trap: *to have sth such as heat or water in one place, especially because it is useful.*

***entrap**: to trap.*

To have sth at one place (Spanish)

retener: *tener algo en algún sitio.*

To cause sth/sb to have sth (English)

supply: *to cause sth/sb to have sth.*

irrigate: *to supply water to (land or crops) to help growth, typically by means of channels.*

To cause sth/sb to have sth (Spanish)

suministrar: *dar algo que resulta necesario.*

abastecer: *suministrar.*

To cause sth not to have sth anymore (English)

remove: *to cause sth not to have sth anymore.*

unload: *to remove the contents of sth.*

denude: *to remove the covering of sth, esp. land.*

To make sth not to have sth anymore (Spanish)

eliminar: *hacer que algo desaparezca.*

retirar: *eliminar o separar de algo.*

To have a large number of sths (English)

accumulate2: *to have a large number of things over a long period of time.*

To have a large number of sth (Spanish)

acumular: *juntar y amontonar, especialmente si se hace en gran cantidad.*

almacenar: *acumular en gran cantidad.*

embalsar: *acumular en un embalse o en un hueco de terreno, especialmente agua.*

SPEECH
to say that sth will happen
to say that sth bad will happen
to say sth to sb

Table 148. Frames of the lexical domain of SPEECH

To say that sth will happen (English)

predict: *to say that an event or action will happen in the future, especially as a result of knowledge or experience.*

forecast: *to predict.*

project: *to forecast sth on the basis of present trends.*

foreshadow: *to predict sth will happen as a result of a warning sign or hint.*

To say that sth will happen (Spanish)

predecir: *(referido a algo que va a suceder), decirlo con antelación.*

pronosticar: *(referido a algo que sucederá en un futuro), predecirlo a raíz de determinados indicios.*

To say that sth bad will happen (English)

warn: *to say that sth bad will happen.*

alert: *to warn sb of a possibly dangerous situation.*

threaten2: *to warn sb that sth very bad will happen soon.*

To say sth bad will happen (Spanish)

avisar: *(referido a un asunto), prevenir, advertir o informar de ello.*

alertar: *avisar de una amenaza o de un peligro.*

amenazar2: *(referido a algo malo o desagradable) avisar de que va a ocurrir en un futuro próximo.*

To say sth to sb (English)

report: *to say that sth has happened, especially an accident or crime (usually of an authority).*

To say sth to sb (Spanish)

informar: *(referido especialmente a una noticia o a un dato), transmitirlo.*

EMOTION
to feel/experience a particular situation

Table 149. Frame of the lexical domain of EMOTION

To feel/experience a particular situation (English)

feel: *to be aware of a particular emotion or sensation.*

experience: *to feel an emotion or sensation.*

suffer: *to experience or be subjected to sth bad or unpleasant.*

To feel/experience a particular situation (Spanish)

sentir: *recibir a través de los sentidos.*

experimentar: *(referido a una sensación, o situación), sentirla, sufrirla o padecerla en uno mismo.*

sufrir: *experimentar un daño o algo que resulta doloroso.*

padecer: *sufrir algo malo.*

ACTION
to do sth
to make sth
to stop doing sth
to come against sth with sudden force
to (cause) to come apart (of construction artifact)
to produce fire
to make an opening in sth with a sharp-edged tool
to make sth (especially referred to a building, road or machine)

Table 150. Frames of the lexical domain of ACTION

To do sth (English)

perform: *to do sth.*

carry out: *to perform a task.*

undertake: *to start to perform.*

To do sth (Spanish)

llevar a cabo: *hacer algo.*

realizar: *(referido a una acción), hacerla.*

efectuar: *realizar***.**

ejecutar: *realizar*.

To make sth (English)

create: *to make sth new.*

produce: *to make sth by doing sth.*

To make sth (Spanish)

fabricar1: *(referido especialmente a un producto) prepararlo, transformarlo o producirlo por medio del trabajo adecuado.*

elaborar: *fabricar algo por medio de un trabajo adecuado.*

crear: *(referido a un objeto), fabricarlo o elaborarlo.*

producir2: *(referido a un objeto), fabricarlo o elaborarlo.*

To stop doing sth (English)

interrupt: *to stop the continuous progress of an activity or process.*

disrupt: *to interrupt (an event, activity, or process) by causing a disturbance or problem.*

To stop doing sth (Spanish)

interrumpir: *(referido especialmente a una acción) impedirlo o suspender su continuación.*

To come against sth with sudden force (English)

hit: *to come against sth/sb with sudden force.*

batter: *to hit sth many times.*

strike: *to hit sth/sb violently.*

blast2: *to hit sth as with a strong gust of wind or explosion.*

crash : *to hit against sth and/or making a loud noise.*

collide: *to crash into each other (esp. of two moving objects).*

To come against sth with sudden force (Spanish)

golpear: *producir choque repentino y violento de un cuerpo contra otro.*

impactar: *golpear violentamente.*

batir: *golpear.*

azotar: *(referido especialmente al viento y demás fenómenos atmosféricos), golpear repetida y violentamente, produciendo daños o destrozos.*

sacudir: *azotar.*

chocar: *(referido a un cuerpo), encontrarse violentamente con otro.*

colisionar: *(referido especialmente a un vehículo), chocar violentamente con otro.*

To (cause) to come apart (of construction artifact) (English)

burst: *to come apart suddenly and violently, especially as a result of an impact or internal pressure.*

collapse: *to burst, break down.*

explode: *to come apart as a result of rapid combustion.*

To (cause) to come apart (of construction artifact) (Spanish)

reventar: *(referido a algo cerrado), abrirse bruscamente por no poder soportar la presión interior o como consecuencia de una fuerte presión exterior.*

explotar: *causar la división violenta de algo en trozos, que se acompaña de la salida de gran cantidad de energía en forma de calor, de luz y de gases.*

explosionar: *explotar.*

estallar: *explotar de golpe y con gran ruido.*

desplomarse: *caerse lo que está levantado.*

To produce fire (English)

burn2: *to produce fire.*

blaze: *to burn.*

smolder/smoulder: *to burn slowly with smoke, but no flames.*

flare: *to burn with a sudden intensity.*

To produce fire (Spanish)

arder2: *producir llamas por estar en combustión.*

To make an opening in sth with a sharp-edged tool (English)

cut: *to make an opening, incision, or wound in sth with a sharp-edged tool or object.*

crop: *to cut sth very short.*

To make an opening in sth with a sharp-edged tool (Spanish)

cortar: *dividir o separar en varias partes.*

picar: *cortar en pequeños trozos.*

To make sth (esp. a building, road, or machine) (English)

construct: *to make sth, esp. a building, road, or machine.*

build: *to construct (sth) by putting parts or material together.*

dam: *to build a dam across a river in order to store the water.*

rebuild: *to build (sth) again after it has been damaged or destroyed.*

To make sth (especially referred to a building, road, or machine) (Spanish)

fabricar2: *producir en serie, generalmente por medios mecánicos.*

construir: (*referido especialmente a una obra de albañilería), fabricarla o hacerla juntando los elementos necesarios para ello.*

alcantarillar: *construir o poner alcantarillas.*

MANIPULATION
to put sth into use
to put a special substance on sth
to regulate a mechanical or scientific process
to mark sth in order to prepare it for later use

Table 151. Frames of the lexical domain of MANIPULATION

English:

use: take, hold, or deploy (sth) as a means of accomplishing or achieving sth.

Spanish:

usar: *hacer server como instrumento para un fin.*

utilizar: *usar algo.*

To put sth into use (English)

implement: *to put sth into use.*

To put sth into use (Spanish)

activar2: *(referido a un mecanismo), ponerlo en funcionamiento.*

implementar: *(referido a algo que se quiere realizar), facilitar los medios necesarios para llevarlo a cabo.*

To put a special substance on sth (English)

treat: *to put a special substance on material such as wood, cloth, metal, etc. or put it through a special process, in order to protect it from damage or decay.*

To put a special substance on sth (Spanish)

tratar: *(referido a una sustancia), someterla a la acción de otra.*

To regulate a mechanical or scientific process (English)

regulate: *to maintain the rate or speed of (a machine or process) so that it operates properly.*

control: *to regulate a mechanical or scientific process.*

manage: *to control the use of sth.*

To regulate a mechanical or scientific process (Spanish)

regular: *ajustar el funcionamiento de un sistema.*

controlar: *(referido a un sistema), regularlo de forma manual o automática.*

monitorizar: *controlar mediante aparatos especiales el curso de uno o varios parámetros fisiológicos o de otra naturaleza para detectar posibles anomalías.*

To mark sth in order to prepare it for later use (English)

mark: *to indicate the position of sth.*

calibrate: *to mark (a gauge or instrument) with a standard scale of readings.*

To mark sth in order to prepare it for later use (Spanish)

marcar: *ajustar la posición de algo.*

calibrar: *ajustar, con la mayor exactitud posible, las indicaciones de un instrumento de medida con los valores de la magnitud que ha de medir.*

COGNITION/MENTAL PERCEPTION
to think carefully about sth

Table 152. Frames of the lexical domain of COGNITION/MENTAL PERCEPTION

To think carefully about sth (English)

consider: *to think carefully about sth.*

equate: *to consider one thing to be the same as or equal to another thing.*

determine: *to think carefully in order to establish an idea.*

calculate: *to determine by reasoning, experience or common sense.*

estimate: *to determine roughly the value, number, quantity or extent of sth.*

underestimate: *to estimate sth to be smaller or less important than it really is.*

overestimate: *to estimate sth to be higher or most favourable than it really is.*

evaluate: *to think carefully about the amount or number of value of sth.*

assess: *to evaluate the nature, ability or quantity of sth.*

calibrate: *to assess carefully sth abstract.*

measure: *to assess the importance, effect, or value of sth.*

<u>To think carefully about sth (Spanish)</u>

considerar: *pensar algo despacio y con atención.*

identificar: *considerar que varias cosas distintas son la misma.*

suponer: *considerar algo como cierto o posible.*

calcular: *pensar en algo con cuidado.*

evaluar: *calcular el valor de algo.*

estimar: *calcular algo de forma aproximada.*

averiguar: *(referido especialmente a un asunto), pensar en algo con cuidado con el fin de descubrir la verdad.*

conocer: *averiguar o descubrir por el ejercicio de las facultades intelectuales.*

examinar: *averiguar, investigar o estudiar con minuciosidad y cuidado.*

estudiar: *examinar detenidamente.*

analizar: *estudiar las distintas partes que componen algo.*

determinar: *averiguar, distinguir o conocer al establecer las diferencias o características:*

decretar: determinar o decidir porque se tiene autoridad para ello.

MOVEMENT
to move in a certain direction (the wind)
to move forcefully
to move slowly
to move circularly
to (cause) to move vertically
to move in a certain direction
to move horizontally in a certain direction, over a large space
to cause to move people from one place to another
to cause_motion
to move (especially for liquids)
to move sth without changing direction
to cause sth/sb to change direction
to cause sth to stop moving

Table 153. Frames of the lexical domain of MOVEMENT

English:

move: *to cause to change position.*

Spanish:

mover: *cambiar de posición o de lugar.*

circular: *moverse*.

desplazar: *mover.*

To move in a certain direction (the wind) (English)

blow: *to move or to be moved by the wind.*

To move in a certain direction (the wind) (Spanish)

soplar: *(referido al viento), correr de forma que se note.*

To move forcefully (English)

surge: *to move quickly and powerfully.*

sweep1*: to move, especially quickly and powerfully.*

blast2: *to move in a specified direction, usually very quickly and noisily by impact or explosion.*

burst1: *to move in a specified direction with power (usually a wind event).*

To move slowly (English)

drift: *to move slowly, especially as a result of outside forces, with no control over direction.*

slide: *to move smoothly.*

creep: *to move slowly, quietly and carefully, usually in order to avoid being noticed.*

shift: *to (cause sth/sb) to move or change from one position or direction to another, especially slightly.*

To move slowly (Spanish)

deslizar: *mover con suavidad sobre una superficie lisa o mojada.*

To move circularly (English)

spin: *to move in a circular manner, especially fast*.

rotate: *to move in a circle, especially around a fixed point.*

To move circularly (Spanish)

girar: *mover sobre un eje o un punto, o dar vueltas sobre ellos***.**

orbitar: *girar describiendo órbitas.*

rotar: *(referido a un cuerpo), mover alrededor de su eje.*

To (cause to) move vertically (English)

shake: *to(cause to) move backwards and forwards or up and down in quick, short movements*.

To (cause) to mover vertically (Spanish)

agitar: *moverse con pequeños y rápidos movimientos de un lado a otro o de arriba a abajo.*

temblar: *agitarse con sacudidas breves, rápidas y frecuentes.*

remover: *mover repetidas veces agitando.*

To move in a certain direction (English)

rise: *to move upwards.*

surface*: to rise to the surface of water*.

fall: *to move downwards.*

descend: *to move downwards (more formal).*

avalanche: *to descend the side of a mountain, especially referred to a large amount of ice, snow and rock.*

converge: *to move towards the same point where they join or meet.*

transport: *to move from one place to another.*

penetrate: *to move into or through sth*.

submerge: *to move below the surface of the sea or a river or lake.*

funnel: *to move or make sth move through a narrow space, or as if through a funnel.*

<u>To move in a certain direction (Spanish)</u>

ascender: *mover de un lugar a un punto o a un grado más alto.*

elevar: *ascender*.

emerger: *mover hacia la superficie del agua o de otro líquido.*

transportar: *llevar de un lugar a otro.*

penetrar: *introducirse en el interior de algo.*

adentrar: *penetrar hacia el interior*.

entrar: *mover de fuera adentro o al interior de algo.*

cruzar: *(referido a un lugar), recorrerlo desde una parte a otra.*

atravesar: *cruzar.*

trasladar: *mover de un lugar a otro.*

trasvasar: *mover (especialmente un líquido) de un recipiente a otro.*

caer: *mover de arriba abajo por la acción del propio peso.*

bajar2: mover a un lugar o posición inferiores.

descender: *bajar*.

volcar: *(referido a un objeto), moverlo hacia un lado o totalmente de forma que su contenido caiga o se vierta.*

avanzar: *mover hacia adelante.*

<u>To move horizontally in a certain direction over a large space (English)</u>

spread: *to move horizontally in a certain direction, covering a larger space.*

extend: *to spread.*

propagate: *to cause to spread out.*

To move horizontally in a certain direction over a large space (Spanish)

extenderse1: *hacer que algo aumente su superficie u ocupe más espacio.*

propagarse: *extenderse.*

To cause to move people from one place to another (English)

evacuate: *to move people from a dangerous place to somewhere safe.*

migrate: *to move from one place to another.*

relocate*: to (cause a person or company to) move to a new place.*

resettle: *to (be helped or forced to) move to another place to live.*

uproot: *to make sb leave a place where they have been living for a long time.*

displace: *to move people away from their home to another place.*

flee: *to move by running away, especially because of danger or fear.*

drown: *to move by running away, especially because of danger or fear.*

To cause to move people from one place to another (Spanish)

evacuar: *(referido especialmente a un lugar) desocuparlo o desalojarlo las personas que en él se encuentren.*

migrar: (referido a una persona) desplazarse para cambiar su lugar de residencia.

To cause_motion (English)

loft: *to cause to move by striking.*

release: *to allow a substance to flow out from somewhere with force.*

eject: *to release.*

erupt2: *to release by exploding and allowing flames and rocks to come out of it.*

blast1: *to release very quickly and loudly in a specified direction.*

emit: *to release (esp. a gas).*

expel: *to release (esp. air or liquid) out of sth.*

blow_out2: *to release suddenly and forcefully.*

spew: *to release (esp. liquid or gas) in large amounts.*

spit: *to release forcibly from a mouth or a mouth-like orifice.*

To cause_motion (Spanish)

echar: *hacer salir algo en una dirección.*

expulsar: *echar del interior de algo*.

desprender: *echar de sí.*

liberar: *desprender o dejar escapar.*

lanzar: *echar algo con fuerza en una dirección.*

despedir: *echar algo con fuerza hacia fuera.*

soltar: *despedir algo.*

emitir: *despedir algo hacia fuera*.

emanar: *emitir*.

arrojar: *despedir o hacer salir con fuerza de un lugar*.

To move (especially for liquids) (English)

flow: *(especially liquids) to move in one direction, especially continuously and easily.*

spill: *to flow, move, fall or spread over the edge or outside the limits of sth.*

spill over/ over spill: *to spill over.*

overflow: *to flow over the edges of a container, etc. because there is too much of it.*

To move (especially for liquids) (Spanish)

fluir: *(referido a un líquido o a un gas), mover por algún lugar o brotar de él.*

bombear: *(referido a un líquido), moverlo e impulsarlo en una dirección determinada.*

derramar: *(referido sobre todo a un líquido), hacer que salga o caiga de donde está y se esparza.*

To move sth without changing direction (English)

channel: *to move sth into a particular place or situation (especially liquids), without changing direction.*

To move sth without changing direction (Spanish)

drenar: *(referido a un lugar) dar salida al agua acumulada en él, generalmente mediante zanjas o cañerías.*

avenar: *drenar*.

desaguar: *(referido a un lugar), extraer o sacar el agua que hay en él.*

To cause sth/sb to change direction (English)

divert: *to cause sth/sb to change direction.*

deflect: *to cause to change direction.*

To cause sth/sb to change direction (Spanish)

desviar: *hacer que algo cambie la dirección que lleva.*

To cause sth to stop moving (English)

paralyze: *to cause sth to stop moving.*

stabilize: *to cause sth to stop changing or moving.*

block: *to cause sth to stop moving through sth.*

To cause sth to stop moving (Spanish)

paralizar: *(referido especialmente a un movimiento), detenerlos, entorpecerlos o impedirlos.*

estabilizar: *hacer que algo pare de cambiar o moverse.*

contener: *(referido a un movimiento), impedirlo o reprimirlo.*

bloquear: *(referido a un lugar), impedir o interrumpir el paso o el movimiento a través de él.*

GENERAL PERCEPTION
to cause sth to be seen
to see sth

Table 154. Frames of the lexical domain of GENERAL PERCEPTION

To cause sth to be seen (English)

show: *to cause sth to be seen.*

To cause sth to be seen (Spanish)

mostrar: *poner algo en presencia de alguien.*

To see sth (English)

observe: *to see sth and know that it is important.*

detect: *to see sth that is partly hidden or not clear, especially using a special method.*

register: *(of an instrument) to detect and show a reading automatically.*

record: *to register a measurement or a result.*

To see sth (Spanish)

observar: *percibir algo.*

ver: *percibir algo a través de la vista.*

detectar: *percibir algo o notar.*

registrar: *(especialmente de un aparato) detectar automáticamente ciertos datos propios de su función, como una cantidad o una magnitud.*

POSITION
to (cause) to be in a certain place or position
to put sth in a certain place
to put sth on top of or over sth
to put water over/in a space

Table 155. Frames of the lexical domain of POSITION

To (cause) to be in a certain place or position (English)

surround: *to be around sth.*

locate: *to be in a particular place.*

localize: *to locate sth precisely.*

inhabit: *to be in a place for a time.*

populate: *to inhabit an area often in large numbers, forming its population.*

To (cause) to be in a certain place or position (Spanish)

situar: *poner en un determinado lugar o tiempo.*

localizar: *situar algo de forma precisa.*

habitar: *(referido a un lugar), ocuparlo y hacer vida en él.*

poblar: *habitar.*

To put sth in a certain place (English)

deposit: *to put sth in a certain place.*

To put sth in a certain place (Spanish)

depositar: *poner algo en un sitio.*

verter: *(especialmente referido a un líquido), ponerlo en un sitio, vaciándolo de otro.*

To put sth on top of or over sth (English)

cover: *to put sth on top or over sth.*

engulf: *to surround and cover sth/sb completely*.

bury: *to cover sth/sb completely with a large quantity of sth.*

To put sth on top of or over sth (Spanish)

cubrir: *poner algo encima de algo.*

sepultar: *cubrir algo de forma que no se vea.*

enterrar: *cubrir algo por completo bajo tierra.*

envolver: *cubrir algo rodeándolo con algo.*

recubir: *cubrir por completo.*

To put water over/in a space (English)

flood: *to put water over/in a space, especially in a way that causes problems.*

inundate: *to flood.*

swamp: *to flood with a large amount of water.*

To put water over/in a space (Spanish)

inundar: *llenar un espacio de agua, especialmente de manera que cause problemas.*

anegar: *inundar.*