Multi-word term variation

Prepositional and adjectival complex nominals in Spanish

Melania Cabezas-García & Santiago Chambó
University of Granada

Complex nominals (CNs) are frequently found in specialized discourse in all languages, since they are a productive method of creating terms by combining existing lexical units. In Spanish, a conceptual combination may often be rendered with a prepositional CN (PCN) or an equivalent adjectival CN (ACN), e.g., *demanda de electricidad* vs. *demanda eléctrica* [electricity demand]. Adjectives in ACNs – usually derived from nouns – are known as 'relational adjectives' because they encode semantic relations with other concepts. With recent exceptions, research has focused on the underlying semantic relations in CNs. In natural language processing, several works have dealt with the automatic detection of relation adjectives in Romance and Germanic languages. However, there is no discourse studies of these CNs, to our knowledge, for the goal of establishing writer recommendations. This study analyzed the co-text of equivalent PCNs and ACNs to identify factors governing the use of a certain form. EcoLexicon ES, a corpus of Spanish environmental specialized texts, was used to extract 6 relational adjectives and, subsequently, a set of 12 pairs of equivalent CNs. Their behavior in co-text was analyzed by querying EcoLexicon ES and a general language corpus with 20 expressions in CQP-syntax. Our results showed that immediate linguistic co-text determined the preference for a particular structure. Based on these findings, we provide writing guidelines to assist in the production of CNs.

Keywords: multi-word term, complex nominal, denominative variation, relational adjective, co-text, specialized language

1. Introduction

Complex nominals (CNs) are units consisting of a head noun modified by other elements, such as other nouns, adjectives, prepositional phrases, etc. They fre-

https://doi.org/10.1075/resla.19012.cab
Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics 34:2 (2021), pp. 402–434.
ISSN 0213-2028 | E-ISSN 2254-6774

quently appear in specialized discourse (Sanz Vicente, 2012) because they are productive instantiations of conceptual combination (Sager et al., 1980), in which new concepts are formed by integrating two or more pre-existing concepts (Murphy, 1988, 1990; Wisniewski, 1996; Gagné, 2000).

However, languages often generate multiple variants resulting from different CN-forming mechanisms. For instance, English produces many compounds by juxtaposing nouns, which have a corresponding expression formed by an adjective and a noun, e.g., *electricity demand* vs. *electric demand* (Maniez, 2014).

Although noun-noun compounds, such as *camión cisterna* [tanker truck] can be found in Romance languages, such as Spanish, the general preference is noun-adjective combinations (e.g., *parque eólico* for *wind farm*). Noun-adjective combinations can also correspond to a noun modified by a prepositional phrase (Liceras et al., 2002), e.g., *gestión ambiental* vs. *gestión del ambiente* [environmental management]. Such pairs can be a problem when it is necessary to choose the best option.

Context is an important factor when analyzing language units (Lyons, 1995). Awareness of contextual preference for a particular morphosyntactic form can make a text more or less acceptable. In fact, poor writing choices in Spanish (or any language) may obscure the meaning of certain complex expressions, such as CNs, and hinder knowledge transfer. Until now, research has mainly focused on the identification of sematic relations in CNs (Downing, 1977; Levi, 1978; Nakov, 2013; López & Bernardos, 2018). Although CN variation has been addressed in studies such as Cartoni (2008, 2009), Harastani et al. (2013), Daille (2017) or Gledhill & Pecman (2018), some problems remain to be solved, especially regarding the linguistic factors that can lead to the preference for one form or another in Spanish.

Our study examined 12 pairs of adjectival and prepositional CNs (henceforth, ACN and PCN), e.g., *producción eléctrica* [electrical production] and *producción de electricidad* [production of electricity]. Six of the most frequent relational adjectives (i.e., adjectives that encode semantic relations with other concepts) were extracted from a Spanish corpus of environmental texts. Twelve CNs were also selected, including those formed by relational adjectives, as well as their prepositional counterparts. The immediate co-text (understood as the parts of discourse surrounding a word, sentence, or passage [Faber and León-Araúz, 2016]) of these variants was then analyzed to identify the factors governing the use of a certain form. The purpose was to discover whether these variants could be handled systematically. As a factor influencing linguistic usage (Faber and León-Araúz, 2016), our findings showed that co-text may determine the preference for a specific variant.

The rest of this article is organized as follows. Section 2 briefly presents the theoretical framework of this study, the role of CNs in specialized communication, and their variation. Section 3 describes the materials and methods of our study. In Sections 4 and 5, our results are presented and discussed. Finally, Section 6 lists the conclusions that can be derived from this research.

2. Theoretical background

2.1 Complex nominals in specialized texts

Multi-word expressions are lexical items composed of more than one element. They are found in all languages since they facilitate lexical expansion (Baldwin & Kim, 2010, p. 2). There are different types of multi-word expressions, such as idioms, collocations, phrasal verbs, and complex nominals (see Bally, 1909 [1951]; Cowie, 1981; Sinclair, 1991; Langacker, 2008; Lorente Casafont et al., 2017; *inter alia*).

Complex nominals (CNs) allow the formation of new concepts and are particularly frequent in specialized discourse. Elements forming CNs can have different positions depending on the term formation rules of different languages (Mairal Usón & Cortés Rodríguez, 2000; Liceras, 2001; Fernández Fuertes et al., 2008). Such multi-word terms (MWTs) are problematic because of the disambiguation of their internal structure, their semantics, translation, and heterogeneous treatment in lexicographic and terminographic resources (Cabezas-García & Faber, 2017).

Correctly identifying CNs in texts is crucial, because this is the first step to understanding and producing these MWTs. However, their identification is not always easy since they are often formed by general language words (e.g., *general* in *general circulation model*). Furthermore, their number of constituents can vary, ranging from two (*organic matter*) to even five (*river water kinetic energy conservation*). The identification of long CNs is thus often difficult, because some of the elements may be left outside the chain.

Additionally, determining the internal associations of CNs is key to eliciting the semantic relations between constituents, and thus, to discovering their meaning. Nevertheless, their dependency analysis often requires domain knowledge. For instance, *offshore wind power* is structured as *offshore [wind power]*, whereas *wind power output* is interpreted as *wind [power output]*. This means that two units (*wind power*) that were grouped together in one CN can be separated in another.

Not surprisingly, meaning access has been the main research focus (Levi, 1978; Warren, 1978; Rosario et al., 2002; Nakov, 2013; López & Bernardos, 2018). Factors that can obscure CN meaning include the specialization of their constituents and the deletion of a crucial element. For example, *stall-regulated wind turbine* alludes to the regulation of a turbine by stopping it. However, the conditions causing this event (i.e., high wind speeds) are not specified even though this is an essential part of the meaning.

Another widely discussed aspect of CNs is the non-specification of the semantic relation between their elements. For instance, in *oil pollution*, pollution *is_caused_by* oil. However, in *water pollution*, pollution *affects* water (Cabezas-García & León-Araúz, 2018). Since the semantic relations between CN constituents are not always transparent, proposals for their identification include sets of semantic relations such as *cause, affect*, etc. (Vanderwende, 1994; Rosario et al., 2002; Nastase & Szpackowicz, 2003; *inter alia*). Another way of clarifying CNs is with paraphrases that represent the sentential structure of the CNs, such as a *power curve* is a curve that represents/calculates/simulates power (Nakov & Hearst, 2006).

When translating or producing CNs in a different language, one must be aware that their structure can vary. For example, Germanic languages are more synthetic and produce packed CNs (Štekauer et al., 2012). The head is generally premodified by nouns or adjectives (e.g., *sediment transport rate*). In contrast, Romance languages are characterized by postmodification, in which adjectives or prepositional phrases (Escandell-Vidal, 1995) are placed to the right of the head. The use of adjectives or prepositional phrases as modifiers is a frequent source of variation since both structures can usually be alternated, e.g., *energía eólica/ energía del viento* [wind power].

In addition, the translation of a CN does not always correspond to the translations of its parts. It can often include more elements or be translated as a single term (Daille et al., 2004; Carrió-Pastor & Candel-Mora, 2013). Even if a literal translation is possible, the type of text, as well as its style, are factors that can determine the choice of the most naturally sounding structure, as occurs in technical or scientific texts (Bocanegra-Valle et al., 2008), such as those studied here.

Still another difficulty is that the entries for CNs in terminographic resources are usually not helpful. In most cases, if they are included at all, they are often listed alphabetically. This means that in Spanish, where postmodification is the rule, related CNs, such as *aerogenerador* and *aerogenerador de eje horizontal*, appear together. However, in the case of English, which uses premodification, their equivalents (e.g., *wind turbine* and *horizontal-axis wind turbine*) are located on separate pages. In other resources, CNs appear as subentries of the head noun.

This usually results in long lists of CNs without a head (e.g., the entry of *erosion* includes *wind ~*, *water ~*, *river ~*, etc.).

As for CN variation, the adjectives related to a noun are not usually shown. For example, the relation between *agua* and *hídrico* [water] is not indicated, even though this would be useful information. Moreover, when synonyms are included, no guidelines are offered about the preferred use of one or another.

2.2 Variation in complex nominals

For many years, terminological variation was ignored because the General Theory of Terminology claimed that it did not exist. However, when the cognitive and communicative dimensions of terminology were acknowledged, research emerged that dealt with the changing nature of terms and concepts in specialized texts (Cabré, 1999; Temmerman, 2000; Freixa, 2006; Pecman, 2012; León-Araúz, 2017).

Variation can be motivated by either user or usage differences. On the one hand, depending on the users, variation can be temporal, geographic or social (Smith et al., 2013; Palacios Martínez, 2014). On the other hand, variation based on usage (i.e., functional variation) can be motivated by field, tenor, or mode (Cabré, 1999).

Additionally, variation can be of two types: term variation or concept variation (León-Araúz, 2017). Term variation or denominative variation occurs when different terms are used to name the same concept (Geeraerts et al., 1994; Carrió-Pastor & Candel-Mora, 2012) (e.g., *wind power* and *wind energy*). Concept variation is when one term designates more than one concept (Geeraerts et al., 1994) (e.g., *inflammation* can be a physiological function, a condition, or the body area suffering from inflammation [Gangemi et al., 2000]). Concept variation can also allude to the formation of new terms as a result of concept expansion (Daille, 2017), or to the contextual modulation of concepts that highlights certain semantic traits while obscuring and suppressing others (Cruse, 1986; León-Araúz, 2017). This type of modulation is directly linked to multidimensionality or the classification of concepts based on different characteristics (Bowker, 1998; León-Araúz, 2017).

Term variation is caused by different processes, such as derivation with Latin or Greek prefixes and suffixes (Sager et al., 1980), simplification (Collet, 2003; Daille, 2017), or multidimensionality. For instance, depending on the dimension emphasized, the same wind turbine can be referred to as a *horizontal-axis wind turbine* or as a *fixed-speed wind turbine* (Cabezas-García & Faber, 2017).

Term variation can also be morphosyntactic when constituents are substituted by other parts of speech. In English, certain N+N compounds can be

replaced by Adj+N equivalents (*atom bomb* and *atomic bomb*). They can also have an alternate postmodification structure: *aspirin synthesis* and *synthesis of aspirin*[1] (Gledhill & Pecman, 2018). In Romance languages, such postmodification also enables alternation between adjectival and prepositional modification (Maniez, 2009; Daille, 2017) e.g., *transporte aéreo* [air transport] and *transporte por aire* [transport by air].

The adjectives that can be replaced by a prepositional phrase are usually 'relational adjectives' (Bally, 1965; Maniez, 2009; Daille, 2017). These have received different names in the literature: (1) 'relational adjectives' in Bally (1965), Maniez (2009), and Daille (2017); (2) 'pseudo adjectives' in Postal (1969); or (3) 'nominal non predicating adjectives' in Levi (1978). Nonetheless, these authors agree on their denominal nature. They are usually derived from a noun by means of a suffix or the use of a Latin or Greek root, which is typical of a highly specialized register (Levi, 1978; Daille, 2001; Maniez, 2009; Sanz Vicente, 2012).

Since such adjectives are derived from nouns, they represent concepts and establish semantic relations with the head of the CN. For example, in *degradación medioambiental* [environmental degradation], degradation *affects* the environment. This is not true for qualifying adjectives, which add a property to the head noun but are not linked to it by a semantic relation. Although qualifying adjectives can also be derived from nouns, they do not usually allude to the same concept designated by the noun. In *error garrafal* [terrible mistake], there is no semantic relation between both constituents, because *garrafal* alludes to the intensity of the mistake rather than to a *garrafa* [container].

Relational adjectives differ from qualifying adjectives because of the following characteristics:

a. They usually cannot be placed in a predicative position (Levi, 1978; Daille, 2001), e.g., *demanda hídrica* [water demand] > **la demanda es hídrica* [*the demand is hydric]. However, we agree with Maniez (2009), who argues that the identification of relational adjectives cannot be solely based on the non-predication criterion. Some relational adjectives can be placed after the verb, e.g., *coche eléctrico* [electric car] > *el coche es eléctrico* [the car is electric].
b. They cannot be modified by degree adverbs (Lees, 1960; Levi, 1978; Daille, 2001; Maniez, 2009), as in **muy hídrico* [*very hydric].
c. They can be coordinated with nouns or other relational adjectives, though coordination with qualifying adjectives is not possible (Levi, 1978), e.g.,

1. For Gledhill & Pecman (2018), this type of alternation can be regarded as two distinctly specialized multi-word terms, rather than variations of the same structure (Gledhill & Pecman, 2018, p. 30). However, we consider that both forms represent CNs (either premodified or postmodified).

**demanda energética e importante* [energetic and important demand].[2] For Daille (2001), they do not admit coordination with other relational adjectives either, although we found that this was frequent in our corpus, e.g., *suministro eléctrico y energético* [energy and water supply].

d. Certain relational adjectives are countable, like the nouns from which they are derived (Levi, 1978), e.g., *trifásico* [three-phase].
e. They can be assigned semantic roles (Levi, 1978). For example, in *impacto ambiental* [environmental impact], *ambiental* [environmental] is the PATIENT. In the same vein, they can be assigned semantic categories, as in *ambiental* [environmental], which is a NATURAL GEOGRAPHIC FEATURE.

Even though Levi (1978) argues that relational adjectives do not permit the formation of new nouns, we found nouns derived from relational adjectives in Spanish, e.g., *ambiental* [environmental] > *ambientalización* [*environmentalization]*. Finally, according to Daille (2001), when the head is modified by more than one adjective, the one closest to the head is relational (Mélis-Puchulu, 1991), e.g., *energético* in *sector energético emergente* [emerging energy sector]. Furthermore, she states that relational adjectives cannot precede the noun, e.g., **una eléctrica producción* [an electric production].

In spite of the relevance of this type of variation in CNs, more research is needed to clarify its characteristics and specific uses. To begin with, CNs formed by adjectives have been largely ignored in the literature, with the majority of studies focusing on noun sequences. Some studies have focused on part-of-speech alternation in CNs in other languages. Levi (1978) comments on N+N and Adj+N alternation in English. Maniez (2009) analyzes some semantic criteria of the head noun in a sample of Adj+N and N+prep+N CNs in French. Cartoni (2009) addresses Adj+N and N+prep+N variants in Italian with a view to handling neologisms in a machine translation system. Daille (2001, 2017) describes the modifications to which Adj+N and N+prep+N CNs in French can be subject. Harastani et al. (2013) use Adj+N and N+prep+N variants to facilitate French to English MWT translation. Gledhill & Pecman (2018) explore the use of N+N and N+prep+N CNs in English, based on their cognitive and communicative function. However, this type of variation has received considerably less attention in Spanish. In Section 3 we describe the materials and methods used to conduct our study, which examines these combinations in Spanish.

2. Throughout the text, the most literal English equivalent is given for the examples in Spanish.

3. Materials and methods

This section presents the corpora and software employed (Section 3.1), as well as the methodology used in our study (Section 3.2).

3.1 Materials

3.1.1 The EcoLexicon Spanish corpus

The EcoLexicon Spanish corpus (EcoLexicon ES) was used for the purposes of our study. With a total of 10,667,434 types and 12,824,222 tokens, EcoLexicon ES currently consists of 1,462 specialized documents on branches of science and disciplines pertaining to the environment (see Table 1). Given that the study and modification of the environment is eminently interdisciplinary, each document may be included in several domains. This corpus is available to the public and can be searched on EcoLexicon (http://ecolexicon.ugr.es), a multilingual terminological knowledge base on Environmental Sciences (Faber et al., 2014).

Table 1. Domains and tokens tagged in EcoLexicon ES

Domain	Tokens	Domain	Tokens
Environmental protection	146,266	Chemical Oceanography	180,925
Environmental law	573,753	Meteorology	768,415
Environmental education	70,928	Climatology	776,884
Sustainable tourism	194,093	Ecology	1,089,487
Geography	684,923	Human Ecology	359,117
Biology	785,676	Soil sciences	482,404
Biological Oceanography	437,564	Oceanography	957,329
Botany	965,330	Biological Oceanography	437,564
Zoology	168,373	Physical Oceanography	120,920
Microbiology	357,933	Geological Oceanography	411,478
Molecular biology	275,088	Chemical Oceanography	180,925
Biochemistry	227,520	Marine Engineering	1,023,759
Physics	426,761	Civil Engineering	290,251
Geophysics	170,018	Transport and Infrastructure Engineering	555,668
Physical Oceanography	69,472	Hydraulic Engineering	675,778
Geology	1,584,567	Coastal Engineering	265,728
Hydrogeology	122,072	Mining Engineering	252,308

Table 1. *(continued)*

Domain	Tokens	Domain	Tokens
Geophysics	170,018	Environmental Engineering	489,761
Geochemistry	117,736	Waste Management	503,760
Geological Oceanography	411,478	Water Treatment and Supply	344,078
Geomorphology	411,478	Air Quality Management	452,169
Hydrology	476,515	Soil Quality Management	196,494
Hydrogeology	122,072	Agricultural Engineering	162,198
Hydrometeorology	32,227	Chemical Engineering	134,378
Chemistry	860,806	Energy Engineering	681,052
Geochemistry	117,736	Renewable Energy	290,251
Biochemistry	227,520		

Document types in EcoLexicon ES include journal articles, books, book chapters, doctoral theses, websites, government or industry reports, legislation, lexicographical material and other scientific documents (leaflets, news and newsletters) (see Table 2). Each document has been curated, cleaned and tagged manually.

Table 2. Corpus size by document type

Document type	Tokens
journal articles	4,700,603 (36.65%)
books	2,647,614 (20.65%)
book chapters	443,465 (3.46%)
doctoral theses	3,381,715 (26.37%)
websites	732,861 (5.71%)
reports	503,766 (3.93%)
legislation	314,580 (2.45%)
lexicographical material	10,066 (0.08%)
other	89,552 (0.70%)

3.1.2 The Spanish Web corpus 2018 (esTenTen18)

The Spanish Web corpus 2018, also known as esTenTen18, belongs to the TenTen corpus family, a group of multi-billion-word general language corpora compiled with texts crawled from the web, developed by Lexical Computing Ltd and available on Sketch Engine (Kilgarriff & Renau, 2013). Compiled in 2018, the esTenTen18 contains over 17.5 billion tokens extracted primarily from webs from Argentina, Spain, Mexico and Chile.

3.1.3 Sketch Engine and CQL

Sketch Engine is a browser-based tool that allows the user to build, analyze and query corpora (www.sketchengine.eu) (Kilgarriff et al., 2004). For the purposes of our study, EcoLexicon ES was uploaded to Sketch Engine, lemmatized and morphologically annotated with the Spanish FreeLing 2.0 part-of-speech tagger (Padró et al., 2010). This enables the user to search corpora for complex grammatical and lexical patterns by means of a concordance notation referred to as Corpus Query Language (CQL) in Sketch Engine documentation.

However, it should be noted that Sketch Engine's CQL is greatly based on the Corpus Query Processor query language (or CQP-syntax) implemented in the corpus analysis architecture Corpus Workbench and developed by Christ et al. (1999) at the Institute for Natural Language Processing of the University of Stuttgart (Evert & Hardie, 2011).

3.2 Methods

3.2.1 Extraction of a list of equivalent adjectival and prepositional complex nominals

The first step was to extract a list of the 150 most frequent adjectives from the corpus. Based on the criteria listed in Section 2.2, 33 relational adjectives were identified. A heterogeneous sample was then randomly selected. The sample consisted of the following six adjectives: (i) *ambiental* [environmental] including variants such as *medioambiental* and *medio ambiental*;[3] (ii) *eólico* [aeolian]; (iii)

3. In CQP-syntax, multiple conditions may be juxtaposed by means of Boolean operators. These can be applied both outside and inside querying elements. In the case of the adjective *ambiental* and its derivatives *medio ambiental* and *medioambiental*, it was necessary to use disjunctive (vertical bar) and optional (question mark) operators, as well as nesting with round brackets, to create an expression that would extract all its variants in the same concordance search. The resulting expression was formulated as follows: ([lemma="medio"]? [lemma="ambiente"])|[lemma="medioambiente"]

eléctrico [electric]; (iv) *atmosférico* [atmospheric]; (v) *energético* [energetic]; and (vi) *hídrico* [hydric] (see Table 3).

Table 3. Relational adjectives in our study

Adjective	Frequency
ambiental	8,839
eólico	4,503
eléctrico	4,039
atmosférico	2,999
energético	2,566
hídrico	1,667

The next step was to extract two contrasting lists for each adjective. The first list was of ACNs (N+Adj), such as *demanda hídrica* [water/hydric demand]. The second list was of PCNs (N+prep+N), such as *demanda de agua* [demand of water]. This was done by querying EcoLexicon ES with the following CQL expressions:

1. [tag = "N.*"][lemma = "hídrico"]
   where [tag = "N.*"] is any noun preceding the adjective in question (e.g., *hídrico*). This expression successfully extracted a list of ACNs such as *gestión hídrica* [water management], *flujo hídrico* [water flow], and *escasez hídrica* [water scarcity].
2. [tag = "N.*"] [tag= "SP"] []{0,2}[lemma = "agua"]
   where [tag = "N.*"] is any head noun, [tag = "SP"] is any given preposition, and [lemma = "agua"] is the noun which the relational adjective refers to. Expression (2) includes a span of up to 2 elements in between tags ([]{0,2}) in order to identify variants with determiners e.g., *producción de la electricidad* [production of the electricity]. Thanks to this formulation, numerous PCNs were identified, including *gestión del agua* [water management], *flujo de agua* [water flow], and *escasez de agua* [water scarcity].

Expression 1 generated a list of 100 ACNs, and expression (2), a list of 100 PCNs. By applying this procedure to all six relational adjectives and referential nouns, we obtained 12 contrasting lists, which allowed us to pinpoint the MWTs with the same head noun. The result was a list of equivalent ACNs and PCNs. Finally, two pairs of equivalent CNs for each relational adjective were selected (see Table 4).

Table 4. Set of CNs

ACNs	Freq.	PCNs	Freq.	English
sector energético	78	*sector de energía*	84	energy sector
potencial energético	61	*potencial de energía*	51	energy potential
concentrador eólico	52	*concentrador de viento*	11	wind concentrator
aprovechamiento eólico	33	*aprovechamiento de viento*	19	wind use
producción eléctrica	279	*producción de electricidad*	101	electricity production
demanda eléctrica	55	*demanda de electricidad*	25	electricity demand
disponibilidad hídrica	42	*disponibilidad de agua*	143	water availability
demanda hídrica	32	*demanda de agua*	65	water demand
protección ambiental	138	*protección del ambiente*	228	environmental protection
deterioro ambiental	42	*deterioro del ambiente*	30	environmental deterioration
CO_2 atmosférico	55	*CO_2 de/en la atmósfera*	47	atmospheric CO_2
emisión atmosférica	25	*emisión a la atmósfera*	60	atmospheric/air emission

3.2.2 Co-textual and internal analysis

In our study, EcoLexicon ES was queried with a set of CQL expressions in order to extract a list of equivalent ACNs and PCNs and perform on them the following co-textual and internal analyses:

a. premodification, intermodification and postmodification by adjectives;
b. premodification, intermodification and postmodification by prepositional phrases;
c. coordination with other CNs.

To this end, two sets of 10 CQL expressions were then formulated to query EcoLexicon ES of our selection of CNs. Each of these sets was based on one of the following core basic expressions:

a. [lemma="X"] [lemma="Y"]
   where [lemma="X"] is the head noun and [lemma="Y"] is the relational adjective in ACNs.
b. [lemma="X"] [tag="SP"] [lemma="el"]? [lemma="Z"]
   where [lemma="X"] is the head noun, [lemma="SP"] is the preposition, [lemma="el"]? is an optional definite article and [lemma="Z"] is the referential noun in PCNs. The reader should note that PCNs may contain a definite

article between its preposition and its referential noun. Examples of CN pairs with article-bearing PCNs include *deriva continental* vs. *deriva de los continentes* [continental drift], *célula parenquimática* vs *célula del parénquima* [parenchyma cell] or *absorción férrica* vs. *absorción del hierro* [iron absorption], to name but a few.

The following subsections describe how these two core expressions were modified to create more complex CQL expressions with the aim of analyzing different features in the immediate co-text and in between constituents of PCNs and ACNs.

3.2.2.1 *Extraction of complex nominals modified by adjectives*

In order to extract CNs modified by adjectives, eight CQL expressions (3–8) were used for scenarios of premodification, intermodification and postmodification (see Table 5). These expressions contain a tag for an adjective [tag = "A.*"] that can be:

a. preposed, e.g., ***alta*** *demanda hídrica* [high water demand], ***creciente*** *demanda de agua* [increasing water demand];
b. interposed, e.g., *emisión* ***ácida*** *atmosférica* [acidic air emission], *emisiones* ***contaminantes*** *a la atmósfera* [polluting air emissions]; or
c. postposed, e.g., *deterioro medioambiental* ***urbano*** [urban environmental damage], *deterioro del medio ambiente* ***mundial*** [global environmental damage].

Table 5. CQL expressions for CNs modified by adjectives

	CQL expressions	Type of CN core
Adjectival premodification		
3.	[tag = "A.*"] [lemma = "X"] [lemma = "Y"]	ACN
4.	[tag= "A.*"] [lemma= "X"] [tag= "SP"] [lemma="el"]? [lemma= "Z"]	PCN
Adjectival intermodification		
5.	[lemma = "X"] [tag = "A.*"] [lemma = "Y"]	ACN
6.	[lemma= "X"] [tag= "A.*"] [tag= "SP"] [lemma="el"]? [lemma= "Z"]	PCN
Adjectival postmodification		
7.	[lemma = "X"] [lemma = "Y"] [tag = "A.*"]	ACN
8.	[lemma = "X"] [tag= "SP"] [lemma="el"] [lemma= "Z"] [tag= "A.*"]	PCN

3.2.2.2 *Extraction of complex nominals postmodified by prepositional phrases*

Another set of four CQL expressions were used to obtain CNs postmodified by prepositional phrases. These include a tag for prepositions [tag = "SP"] that allows to extract prepositional phrases that may postmodify CNs (see Table 6). Expressions 9 and 10 generated a set of concordances of prepositional phrases such as *demanda eléctrica* ***de la red*** [electricity demand of the grid] and *emisión atmosférica* ***de las industrias*** [industrial air emissions].

Table 6. CQL expressions for CNs modified by prepositional phrases

	CQL expressions	Type of CN core
9.	[lemma = "X"] [lemma = "Y"] [tag = "SP"]	ACN
10.	[lemma = "X"] [tag = "SP"] [lemma="el"]? [lemma = "Z"] [tag = "SP"]	PCN

3.2.2.3 *Extraction of complex nominals as prepositional modifiers*

Given that CNs can also postmodify other nominal phrases, preceding co-textual elements were analyzed in our study. This was done by devising two CQL expressions containing a tag for prepositions [tag= "SP"] before the CNs. As can be seen in Table 7, expressions 11 and 12 include a span of up to three elements, such as articles or adjectives, which may occur between the preposition and the head noun. These expressions were used to obtain CNs acting as postpositional modifiers of other nominal structures, e.g., ***modelo*** *de aprovechamiento eólico* [model of wind use] and ***superficie frontal*** *del concentrador eólico* [frontal surface area of the wind concentrator].

Table 7. CQL expressions for CNs as prepositional modifiers

	CQL expressions	Type of CN core
11.	[tag = "SP"] []{0,3} [lemma = "X"] [lemma = "Y"]	ACN
12.	[tag = "SP"] []{0,3} [lemma = "X"] [tag = "SP"] [lemma="el"]? [lemma = "Z"]	PCN

3.2.2.4 *Extraction of complex nominals coordinated with other complex nominals*

Another co-textual feature of CNs is coordination. In order to extract all relevant instances, all possible scenarios of CNs coordination were established (see

Table 10 in Section 4.3). To this end, six CQL expressions containing a tag for any coordinating conjunction [tag= "CC"] were formulated, as can be seen in Table 8.

Table 8. CQL expressions for CN coordination

	CQL expressions	Type of CN core
13.	[tag = "CC"] []{0,2} [lemma = "X"] [lemma = "Y"]	ACN
14.	[tag = "CC"] []{0,2} [lemma = "X"] [tag = "SP"] [lemma="el"]? [lemma = "Z"]	PCN
15.	[lemma = "X"] [tag = "CC"] []{1,3} [lemma = "Y"]	ACN
16.	[lemma = "X"] [tag = "CC"] []{1,3} [tag = "SP"] [lemma="el"]? [lemma = "Z"]	PCN
17.	[lemma = "X"] [lemma = "Y"] [tag = "CC"]	ACN
18.	[lemma = "X"] [tag = "SP"] [lemma="el"]? [lemma = "Z"] [tag = "CC"]	PCN

Expressions 13 and 14 extracted instances of CN coordination at the head (e.g., ***fuentes y sectores*** *energéticos* [energy sources and energy sectors] and ***almacenamiento y disponibilidad*** *de agua* [water storage and availability]) as well as other broader scenarios such as *carbonato cálcico y* CO_2 *atmosférico* [calcium carbonate and atmospheric CO_2]. Expressions 17 and 18 also obtained CN coordination in broader scenarios, but they extracted coordination at the modifier instead. For instance, these expressions extracted CNs such as CO_2 ***atmosférico o disuelto en agua*** [CO2 in the atmosphere or diluted in water] as well as *demanda de agua y cambio climático* [water demand and climate change]. On the other hand, expressions 15 and 16 restricted the extraction only to CNs coordinated at the head, e.g., **disponibilidad y almacenamiento** de agua [water availability and storage].

3.2.2.5 *Extraction of other internal elements*

Additionally, two other vaguer CQL expressions were used (see Table 9). These contain a span of up to five elements in between constituent elements. Although they generated a considerable amount of noise, both were useful because they identified examples of CNs with elements that had been deleted in other variants.

Expression 19 extracted lengthier variants of ACNs such as *demanda de recurso hídrico* [water resource demand], which is a synonym of *demanda hídrica* [water demand]. As a counterpart for PCNs, expression 20 extracted PCN variants with additional prepositional phrases between their constituent elements, e.g., *emisión a la atmósfera* [emission into the atmosphere] > *emisión de partículas a la atmósfera* [emission of particles into the atmosphere].

Table 9. CQL expressions for tentative CN variant extraction

	CQL expressions	Type of CN core
19.	[lemma = "X"] []{1,5} [lemma = "Y"]	ACN
20.	[lemma = "X"] [tag = "SP"] []{1,5} [tag = "SP"] [lemma = "el"]? [lemma = "Z"]	PCN

3.2.3 Analysis of preference for adjectival or prepositional complex nominals with esTenTen18

The esTenTen18 corpus was used to compare occurrences of equivalent ACN and PCN. To this end, the corpus was queried with the same CQL expressions (see Section 3.2.2) for each analyzed CN pair. This method proved useful to detect preference for one type of CN over the other, for example, when followed by prepositional postmodification. Ideally, this should have been done with a bigger Spanish environmental specialized corpus comparable to EcoLexicon ES, although such a corpus could not unfortunately be retrieved or accessed.

Even though preference for CN forms in general language corpora cannot be extrapolated to specialized language, it is worth noting that searches in esTenTen18 revealed a strong preference for one CN form (see *emisión a la atmósfera de* vs. *emisión atmosférica de* in Section 4.2.3). An extensive study in this line of work should seek to enlarge EcoLexicon ES in the future.

4. Variation in complex nominals: Adjectival modification vs. prepositional modification

The analysis of the immediate co-texts of CNs, as well as of other interpositional elements, provided insights into the behavior of these MWTs in specialized environmental texts.

4.1 General trends in EcoLexicon ES

The results obtained confirmed that ACNs were more frequent in Spanish specialized texts than PCNs. This coincides with findings for other languages (Maniez, 2014; Daille, 2017; Gledhill & Pecman, 2018). Figure 1 shows the frequency of ACNs and PCNs in the corpus. As can be observed, adjectival modification exceeded prepositional modification in 8 of the 12 CNs. The only exceptions were *demanda* + *agua* [demand + water], *disponibilidad* + *agua* [availability + water],

emisión + *atmósfera* [emission + atmosphere], and *protección* + *ambiente* [protection + environment]. This suggests that MWTs with adjectival modification are more specialized.

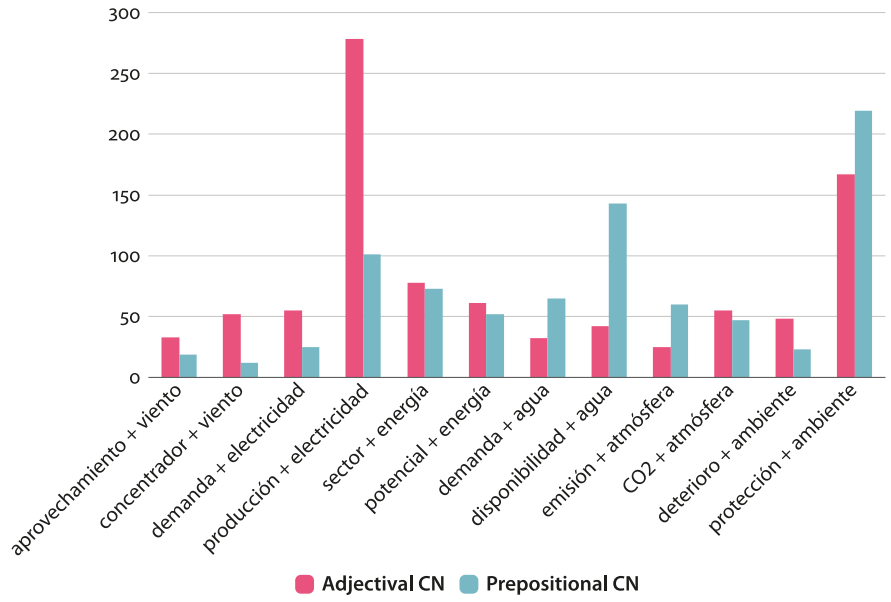

Figure 1. Frequency of ACNs and PCNs

To measure how co-text affected the use of adjectival or prepositional modification, the following variables were analyzed:

a. modification of CNs by additional adjectives, e.g., *aprovechamiento eólico* ***marino*** [marine wind use], and prepositional phrases, e.g., *demanda hídrica* ***para irrigación*** [crop water demand] (see Section 4.2)
b. postpositional modification of other noun phrases by selected CNs, e.g., *modelo de* ***concentrador de viento*** vs. *modelo de* ***concentrador eólico*** [wind concentrator prototype] (see Section 4.3)
c. coordination with other CNs, e.g., *oferta de electricidad y demanda de electricidad* [electricity supply and electricity demand] > *oferta y demanda de electricidad* [electricity supply and demand] (see Section 4.4)
d. length and type of the sentence (see Section 4.5). These elements were the starting point for co-textual analysis.

4.2 Modification by additional adjectives and prepositional phrases

4.2.1 Modification of CNs by additional adjectives

Our results showed that CN modification by other adjectives (in any position) occurred more frequently in PCNs (105) than in ACNs (56), which suggests that PCNs are more unstable. For instance, *disponibilidad + agua* [availability + water] in its PCN form was modified by 19 adjectives: e.g., ***escasa*** *disponibilidad de agua* [limited availability of water]; *disponibilidad* ***mundial*** *de agua* [global availability of water]; *disponibilidad de agua* ***freática*** [availability of groundwater]. For its ACN counterpart, *disponibilidad hídrica*, only 4 antepositional cases were identified, i.e., *alta* [high], *baja* [low], *buena* [good] and *mayor* [greater].

When the adjective was to the left of the CN, e.g., ***alta*** *demanda hídrica* [high water demand], it clearly modified the head. No clear preference for ACNs or PCNs was observed (19 for ACNs vs. 21 for PCNs). However, a wider range of adjectives was found in the case of PCNs, which indicated greater variability.

When the adjective appeared in the middle of the CN, e.g., *producción* ***bruta*** *de electricidad* [gross electricity generation], there was a marked preference for PCNs (40 cases) in comparison to ACNs (1 case), as shown in Figure 2. The only ACN that allowed internal adjectival modification was *emisión atmosférica*.

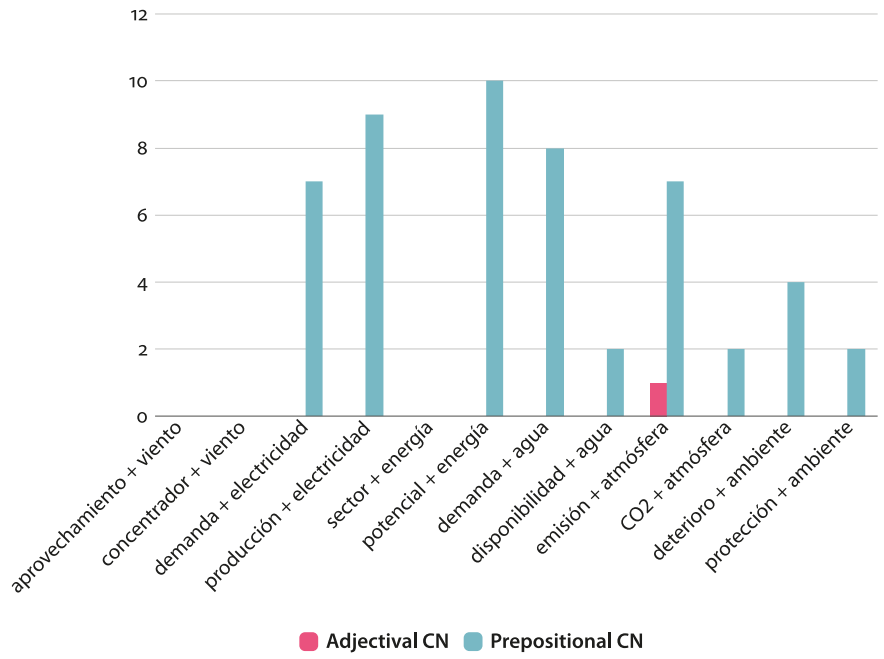

Figure 2. CN modification by the insertion of adjectives

Finally, when an adjective postmodified the CN, e.g., *producción eléctrica* ***eólica*** [wind electricity generation], two dependency trends were observed. When the head and the first modifier were grouped (the adjective complementing thus the entire CN), e.g., [*producción eléctrica*] *global* (global electricity generation), the ACN variant was preferred in 25 cases, whereas the prepositional form was used in only 9. In contrast, when the adjective modified the last constituent of the MWT, e.g., *potencial* [*de energía eólica*] (wind power potential), the prepositional form was used in 23 CNs, and the adjectival variant was used in 11 CNs. The semantic dependencies linked to structure were thus a determining factor in the preference for one variant or another.

4.2.2 Modification of CNs by additional prepositional phrases

Regarding CN modification by prepositional phrases, we focused on prepositional phrases in any position, for example, *demanda hídrica* ***para abastecimiento urbano*** [urban water demand], *emisión* ***de aerosoles*** *a la atmósfera* [emission of aerosols into the atmosphere] or *reducción* ***de*** *la demanda hídrica* [decrease in water demand].

ACNs were found to be more likely to admit postpositional modification by other prepositional phrases than PCNs. This was evident since 8 of the 12 CN variant pairs analyzed displayed a clear preference for such modification in over 60% of the cases. Exceptions were the following: *demanda* + *agua* [demand + water]; *disponibilidad* + *agua* [availability + water]; *emisión* + *atmósfera* [emission + atmosphere]; and CO_2 + *atmósfera* [CO_2 + atmosphere]. These CNs behaved differently since the prepositional variant was preferred in over 60% of all cases, when postmodified by a prepositional phrase.

Interpositional modification of CNs by prepositional phrases was both infrequent and difficult to determine. The extraction process involved identifying the items between the head noun and its modifier by means of a prepositional tag followed by a span. Although this analysis was performed on both ACNs and PCNs, ACNs did not admit this type of modification because of their adjectival nature. Accordingly, any prepositional phrase between the head noun and its modifying adjective could modify an entire ACN.

For instance, *del recurso* [of the resource] in *demanda del recurso hídrico* [demand of the hydric resource] does not modify *demanda* + *agua* [demand + water] in its adjectival instantiation. Instead, it gives rise to a less compact variant, i.e., *demanda del recurso hídrico* [demand of the hydric resource]. The same applies to most PCNs, such as *aprovechamiento de la energía del viento* [use of the energy of the wind] or *potencial de generación de energía* [potential of energy generation]. Both are less compact variants of *aprovechamiento* + *viento* [use + wind] and *potencial* + *energía* [potential + energy], respectively. Nevertheless,

this method was useful for extracting 19 lengthier variants of certain CNs, as in *aprovechamiento de la energía del viento* and *aprovechamiento eólico*. This type of denominative variation, often referred to as 'lexical reduction', implies the deletion of elements with few conceptual content or those not characterizing the concept of the CN (Haralambous and Lavagnino, 2011: 43), e.g., *recurso* in *demanda del recurso hídrico* (thus becoming *demanda hídrica*). They represent a mechanism of linguistic economy and textual link (Haralambous and Lavagnino, 2011). Additionally, we found other PCNs that were not denominative variants, such as *disponibilidad de vapor de agua* [availability of steam]. This MWT is not a synonym of *disponibilidad de agua* [water availability], since water can be available in other states.

4.2.3 Semantic-constrained preference for interpositional or postpositional modification

The interpositional behavior of *emisión a la atmósfera* [emission into the atmosphere] is worth highlighting. This CN accepts a wide range of prepositional phrases between its constituents. Our results showed 25 cases, such as *emisión de partículas a la atmósfera* [emission of particles into the atmosphere], *emisión de N_2O a la atmósfera* [emission of N_2O into the atmosphere], and *emisión de aerosoles terrígenos a la atmósfera* [emission of terrigenous aerosols into the atmosphere]. Though slightly less productive (17 cases), this PCN also admitted equivalent postpositional modification, e.g., *emisión a la atmósfera de contaminantes* [emission into the atmosphere of pollutants]. On the other hand, for the ACN *emisión atmosférica* [atmospheric emission], we only extracted four instances of prepositional postmodification, none of which refers to the PATIENT (i.e., what is being released into the atmosphere), but rather to the AGENT or ORIGIN of such emission, e.g., *emisión atmosférica de las industrias* [atmospheric emission from the industries], *emisiones atmosféricas en los núcleos poblacionales* [atmospheric emissions at populational nuclei].

The esTenTen18 corpus was also used to confirm results and to detect whether there was a preference for a PCN or an ACN when followed by postmodification. For example, there were only 12 hits for the query *"emisión atmosférica de"* in contrast to *"emisión a la atmósfera de"*, which had considerably more (1,247 hits). When specifying the PATIENT of this process, there was an evident preference for postpositional or interpositional modification of the PCN, e.g., *emisión a la atmósfera de contaminantes* [emission into the atmosphere of pollutants] or *emisión de CO_2 a la atmósfera* [emission into the atmosphere of CO_2]. In contrast, when specifying the AGENT or ORIGIN, there was a preference for postpositional modification of the ACN, e.g., *emisiones atmosféricas de las centrales eléctricas* [atmospheric emissions from power plants] or *emisiones atmosféricas de fuentes*

móviles [atmospheric emissions from mobile sources]. As can be seen, argument structure (Martín Arista, 1997; Rodríguez-Juárez, 2017) is also important in the preference for one variant or another. Further research should be carried out to ascertain the influence of argument structure on the different types of denominative variants.

In the results obtained, *emisión* + *atmósfera* was the only conceptual combination that was rendered as a PCN with a preposition other than *de*. This indicates that different prepositions in CNs can reflect a different combinatorial potential of CNs, in terms of conceptual roles and categories of their components. Further research should also investigate the relationship between the type of prepositions in CNs, the type of conceptual propositions that underlie CNs, and its impact on CN formation.

4.2.4 Dependency analysis

When adding other adjectives or prepositional phrases to the two-term CNs in our study, they became longer combinations, thus needing dependency analysis. For instance, it was necessary to know whether the first two elements were grouped, as in [*protección medioambiental*] *general* (general environmental protection), or rather the final constituents, as in *protección* [*del medio ambiente mundial*] (protection of the global environment). This difficulty arose when the new element was placed next to the CN modifier: e.g., *general* in *protección medioambiental general* [general environmental protection], *ácida* in *emisión ácida atmosférica* [acidic air emission] or *de generación* in *potencial de generación de energía* [potential of energy generation]. Knowledge of internal structure was thus essential because these dependencies often determined the preference for adjectival or prepositional variants of CNs, as well as meaning access and production in another language.

In the case of CNs that were difficult to interpret, we applied the indicators in Cabezas-García & León-Araúz (2019) by means of CQL queries in our corpus to disambiguate the CN structure. They propose the following indicators suggesting the dependency of a possible combination inside a CN: (1) the combination appears as an independent CN in the corpus, (2) the preferred combination is more frequent than the other possible groupings, (3) the combination does not allow the insertion of external elements modifying its meaning, (4) the combination forms other CNs, and (5) the combination has synonyms or antonyms.

For example, in *producción eléctrica trifásica* [three-phase electricity generation], *trifásica* was found to modify *eléctrica*, in a structure such as *producción* [*eléctrica trifásica*], because *electricidad trifásica* [three-phase electricity] had 39 occurrences in our corpus while *producción trifásica* [three-phase generation] was not found (indicators 1 and 2).[4] This structure was also confirmed in *potencial de generación de energía* [*potential of energy generation*], where *generación* complemented *energía* because external elements were found between both groups, as in *potencial elevado/eólico/rentable* [*de generación de energía*] (indicator 3). On the contrary, no elements were found between *generación* and *energía*. Other CNs, such as *demanda de electricidad peninsular* [demand of electricity in the (Iberian) Peninsula] had the reverse structure: [*demanda de electricidad*] *peninsular*. This was ascertained after retrieving other CNs formed by *demanda de electricidad* and not by *electricidad peninsular*, such as *demanda de electricidad mundial/nacional/europea* (indicator 4). Additionally, the structure of other CNs, such as *disponibilidad de agua freática* [availability of groundwater], was disambiguated by means of the retrieval of synonyms of one of the possible combinations. In particular, *agua freática* is also referred to as *agua subterránea libre* or *agua subterránea no confinada*, while no synonyms or antonyms of *disponibilidad de agua* were found, thus suggesting the structure *disponibilidad* [*de agua freática*] (indicator 5).

4.3 Complex nominals as postpositional modifiers

Postpositional modification of other noun phrases by our CNs was also observed, e.g., *reducción* ***de la demanda energética*** [decrease in energy demand]. There were three conceptual combinations that showed no evident preference for a specific CN form when acting as a postpositional modifier (all with percentages of less than 60%). Only *disponibilidad + agua* [availability + water] and *emisión + atmósfera* [emission + atmosphere] favored PCN modification. However, the majority of pairs (7 out of 12) appeared as ACNs when acting as postpositional modifiers (>60% of all cases). This indicated a preference for ACNs in the role of postpositional modifiers. Figure 3 shows the number of occurrences of ACNs and PCNs as postpositional modifiers in the corpus.

4. When both forms were possible, the most frequent option was selected.

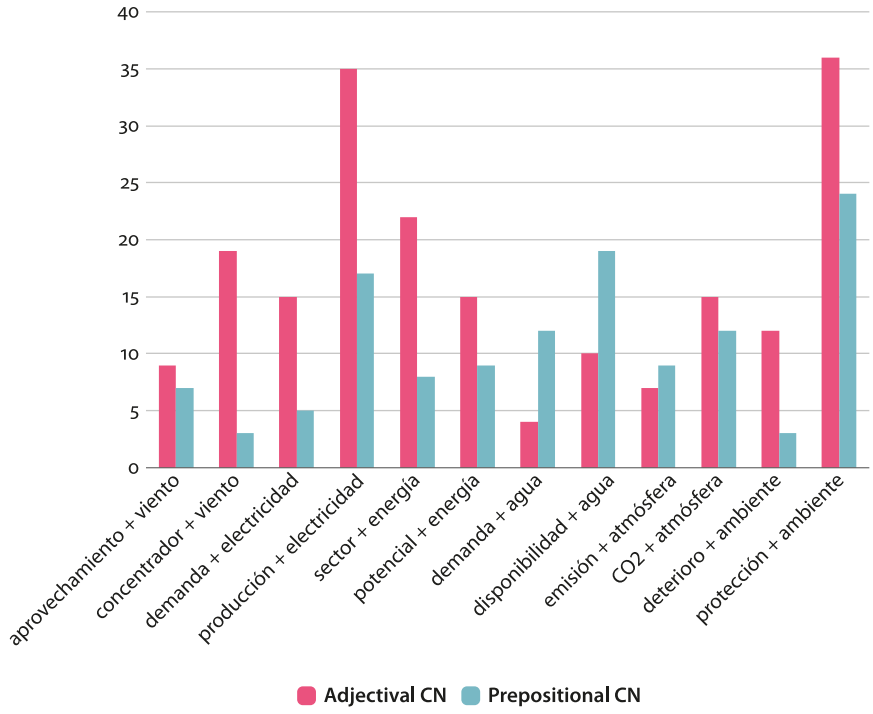

Figure 3. ACNs and PCNs as postpositional modifiers

4.4 Coordination with other complex nominals

Another co-textual feature of CNs was coordination. Recent work on CN variation (Daille, 2017) regards CN coordination as variation. For instance, combining *demanda* + *agua* [demand + water] and *demanda* + *electricidad* [demand + electricity] gives rise to coordinated phrases such as *demanda hídrica y eléctrica* [water and electricity demand].

Our study established all possible combinatorial scenarios for the coordination of two ACNs (ACN+ACN), two PCNs (PCN+PCN), and an ACN and PCN (PCN+ACN). As shown in Table 10, 73 of 114 cases of two-element CN coordination (including prepositions *y* and *o* [*and* and *or*]) were formed by PCN+PCN combinations, whereas only 21 cases resulted from ACN+ACN strategies. Additionally, we found 20 other cases that were produced by mixed PCN+ACN operations. This indicates that there was a preference for PCNs when combined with other CNs, whether another PCN or ACN.

Table 10. All possible scenarios of CN coordination

Case	Sum	Explanation	Example
I	PCN+PCN	Same head noun, different modifiers (only one preposition).	*deterioro de los suelos y el medio ambiente* [soil and environmental degradation]
II	PCN+PCN	Same head noun, different modifiers (two prepositions).	*protección de la salud pública y del medio ambiente* [public health and environmental protection]
III	PCN+PCN	Different head noun, same modifier.	*oferta y demanda de electricidad* [electricity supply and demand]
IV	PCN+PCN	Different head noun, different modifier (or the same but duplicated).	*emisiones a la atmósfera y vertidos al agua* [air and water emissions]
V	ACN+ACN	Same head noun, different modifier.	*sector eólico y energético* [wind and energy sector]
VI	ACN+ACN	Different head noun, same modifier.	*CO_2 y vapor de agua atmosféricos* [atmospheric CO_2 and water vapor]
VII	ACN+ACN	Different head noun, different modifier.	*desarrollo económico y protección medioambiental* [economic development and enviromental protection]
VIII	PCN+ACN	Same head noun, different modifier.	*sector eléctrico y de la energía renovable* [electricity and renewable energy sector]
IX	PCN+ACN	Different head noun, different modifier.	*crisis energética y deterioro del medioambiente* [energy crisis and environmental deterioration]

4.5 Sentence length and type

Finally, our analysis of sentence length did not show any preference for ACNs or PCNs. There was only one case of conceptual combination, *deterioro + ambiental*, which slightly exceeded a 10-word difference in favor of a CN option. For this case, sentences containing the PCN, *deterioro del ambiente* (environmental deterioration), were longer than those with its ACN counterpart *deterioro ambiental* [environmental deterioration]. Conversely, the other pairs analyzed showed an average difference in sentence length no greater than five words. We initially thought that longer sentences would be more likely to have ACNs because of the desire to avoid prepositional clutter, that is to say, long phrases formed by excessive prepositional postmodification with the same preposition, e.g., *aprovechamiento del viento de la zona de estudio* in contrast with

aprovechamiento eólico de la zona de estudio [use of wind power in the area studied]. However, this phenomenon could not be confirmed. Nonetheless, it was observed that ACNs occurred more frequently in titles, footnotes, and other structures containing no verb. The shorter length of this type of phrases without verb may have affected the similar length of sentences with PCNs and ACNs, although this needs to be further explored in future research. Figure 4 shows the number of occurrences of ACNs and PCNs in non-verbal structures.

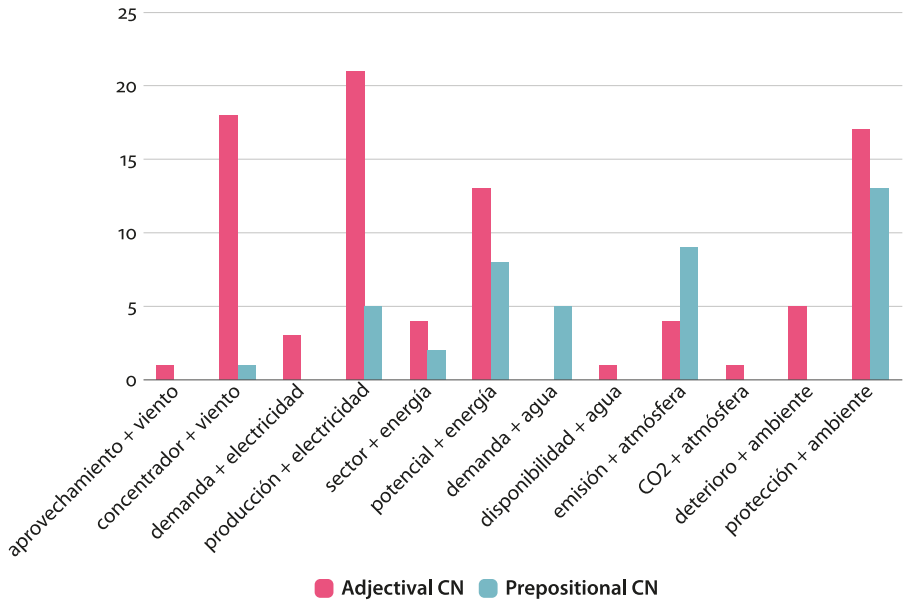

Figure 4. CNs in titles, footnotes and similar non-verbal structures

5. Recommendations of use

Our results reflected that ACNs were the prevalent form of conceptual combination in specialized discourse, which coincided with other research (Maniez, 2014; Dailles, 2017; Gledhill & Pecman, 2018). They are thus the preferred option when the two first CN constituents are grouped, as well as when the ACN is followed by a preposition. Additionally, ACNs are frequent postpositional modifiers in prepositional phrases within other noun clauses. They are also more likely to appear in titles and footnotes, where conciseness is essential. Our study also found that PCNs have a greater combinatorial potential in regard to adjectival modification. Generally, they are the preferred option in combinations with other adjectives, especially if they appear interpositionally. Likewise, we found that there was a

preference for PCNs when the final elements of the MWT were grouped, in contrast to their adjectival counterparts (which suggested the reverse structure).

All too often, specialized texts are riddled with complex CN structures codifying conceptual combinations that may seem obscure to readers. Writers can maximize meaning access to CNs if they combine adjectival and prepositional elements in a way that reflects the dependency structure of a given combination. In view of the results of this study, we make the following recommendations for writers in doubt when faced by the ACN vs. PCN dichotomy:

1. Use a PCN if it requires modification by another adjective and the two final elements of the CN are grouped, e.g., *potencial* [*de energía eólica*] instead of *potencial energético eólico* (wind power potencial).
2. Use an ACN if it requires modification by another adjective and the two first elements of the CN are grouped, e.g., [*producción eléctrica*] *global* instead of *producción de electricidad global*. However, if opting for a prepositional option, use interpositional adjectivation, e.g., *producción global de electricidad* instead of *producción de electricidad global* (global electricity generation).
3. Use an ACN if it modifies another noun as part of a prepositional phrase, e.g., *evolución* [*del CO_2 atmosférico*] instead of *evolución del CO_2 de la atmósfera* (evolution of CO_2 in the atmosphere).
4. Use an ACN when followed by other prepositional phrases, e.g., [*disponibilidad hídrica*] *de la cubierta vegetal* instead of *disponibilidad de agua de la cubierta vegetal* (water availability of the vegetation cover).
5. Use CNs of the same nature when coordination affects their modifiers, e.g., *CO_2 atmosférico y edáfico* or *CO_2 de la atmósfera y del suelo*, but not *CO_2 atmosférico y del suelo* or *CO_2 de la atmósfera y edáfico* (CO_2 in the soil and the atmosphere).

Evidently, these recommendations are based on the analyzed corpus, consisting of specialized scientific texts on the environment. Thus, they may vary in other contexts of communication. For this reason, further research should investigate usage preferences with a view to guiding the production of term variants in different contexts of communication.

6. Conclusions

In specialized texts, multi-word terms (MWTs) usually take the form of complex nominals (CNs). These terms designate a specialized concept by means of different structures, such as the prepositional or adjectival modification of nouns.

However, discourse factors affecting the preference for one form or another are largely context-dependent. This is particularly true of Spanish (Moreno-Fernández, 2003).

In this study, we examined 12 pairs of equivalent ACNs and PCNs, e.g., *producción eléctrica* and *producción de electricidad* [electricity generation], in a corpus of Spanish specialized environmental texts. Their co-text as well as the elements between CN constituents were analyzed to explore the factors determining the use of these variants. More specifically, we examined the modification of CNs by other adjectives or prepositional phrases, their coordination with other CNs, and the length and type of the sentence where they appeared.

Our results showed that, although PCNs and ACNs are apparently interchangeable, there are some co-textual features that determine the preference for a specific variant. In particular, the use of ACNs was favored in the following situations: (i) when adding an adjective (e.g., *global*) to an MWT whose two first elements are grouped, as in [*producción eléctrica*] *global*; (ii) when postmodified by prepositional phrases; (iii) when acting as prepositional modifiers. In contrast, PCNs were preferred in MWTs with a structure such as *disponibilidad* [*de agua freática*] (where an adjective modified the last MWT constituent), and when there was an adjective modifying the head noun. Based on this behavior, we provided some basic CN writing guidelines.

However, MWT variation is a topic that still needs to be further investigated. In addition to a future extensive study in this line of work, other studies should be carried out to ascertain the behavior of CN variants in other domains, as well as in other languages. More concretely, future research will focus on the relationship between the underlying conceptual propositions in CNs (e.g., in terms of argument structure) and other phenomena of morphosyntactic variation in MWTs.

Funding

This research was carried out as part of project FFI2017-89127-P, Translation-Oriented Terminology Tools for Environmental Texts (TOTEM), funded by the Spanish Ministry of Economy and Competitiveness. Funding was also provided by an FPU grant given by the Spanish Ministry of Education to the first author.

References

Baldwin, T., & Kim, S.N. (2010). Multiword Expressions. In N. Indurkhya & F.J. Damerau (Eds.), *Handbook of Natural Language Processing, Second Edition* (pp. 267–292). Boca Raton, FL: CRC Press.

Bally, C. (1909 [1951]). *Traité de stylistique française*. Heidelberg: C. Winter.
Bally, C. (1965). *Linguistique génerale et lingustique française*. Bern: Francke Verlag.
Bocanegra-Valle, A., Lario de Oñate, M.C., & López Torres, E. (2008). *English for Specific Purposes: Studies for Classroom Development and Implementation*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta*, *43*(4), 631–651. https://doi.org/10.7202/002134ar
Cabezas-García, M., & Faber, P. (2017). A Semantic Approach to the Inclusion of Complex Nominals in English Terminographic Resources. In R. Mitkov (Ed.), *Computational and Corpus-Based Phraseology*, Lecture Notes in Computer Science, 10596 (pp. 145–159). Cham: Springer. https://doi.org/10.1007/978-3-319-69805-2_11
Cabezas-García, M., & León-Araúz, P. (2018). Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study. In N. Calzolari et al. (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 2511–2518). Miyazaki, Japan: ELRA.
Cabezas-García, M., & León-Araúz, P. (2019). On the Structural Disambiguation of Multi-word Terms. In G. Corpas Pastor & R. Mitkov (Eds.), *Computational and Corpus-Based Phraseology*, Lecture Notes in Computer Science, 11755 (pp. 46–60). Cham: Springer. https://doi.org/10.1007/978-3-030-30135-4_4
Cabré, M.T. (1999). *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada. https://doi.org/10.1075/tlrp.1
Carrió-Pastor, M.L., & Candel-Mora, M.A. (2012). Corpus analysis: a pragmatic perspective on term variation. *Revista Española de Lingüística Aplicada*, *25*(1), 33–50.
Carrió-Pastor, M.L., & Candel-Mora, M.A. (2013). Variation in the translation patterns of English complex noun phrases into Spanish in a specific domain. *Languages in Contrast*, *13*(1), 28–45. https://doi.org/10.1075/lic.13.1.02car
Cartoni, B. (2008). *De l'incomplétude lexicale en traduction automatique : vers une approche morphosémantique multilingue*. PhD. Université de Genève.
Cartoni, B. (2009). Lexical morphology in machine translation: A feasibility study. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)* (pp. 130–138). Athens: Association for Computational Linguistics. https://doi.org/10.3115/1609067.1609081
Christ, O., Schulze, B.M., Hofmann, A., & König, E. (1999). *The IMS Corpus Workbench: Corpus Query Processor (CQP) – User's Manual*. Institute for Natural Language Processing. Stuttgart: University of Stuttgart.
Collet, T. (2003). A two-level grammar of the reduction processes of French complex terms in discourse. *Terminology*, *9*(1), 1–27. https://doi.org/10.1075/term.9.1.02col
Cowie, A.P. (1981). The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics*, *2*(3), 223–235. https://doi.org/10.1093/applin/2.3.223
Cruse, D.A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
Daille, B. (2001). Qualitative Terminology Extraction: Identifying Relational Adjectives. In D. Bourigault et al. (Eds.), *Recent Advances in Computational Terminology* (pp. 149–166). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/nlp.2.08dai
Daille, B. (2017). *Term Variation in Specialised Corpora: Characterisation, Automatic Discovery and Applications*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.19

Daille, B., Dufour-Kowalski, S., & Morin, E. (2004). French-English Multi-word Term Alignment Based on Lexical Context Analysis. In M.T. Lino et al. (Eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)* (pp. 919–922). Lisbon: ELRA.

Downing, P. (1977). On the Creation and Use of English Compound Nouns. *Language*, *53*(4), 810–842. https://doi.org/10.2307/412913

Escandell Vidall, M.V. (1995). Cortesía, fórmulas convencionales y estrategias indirectas. *Revista Española de Lingüística*, *25*(1), 31–66.

Evert, S., & Hardie, A. (2011). Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In *Proceedings of the Corpus Linguistics 2011 Conference* (pp. 1–21). Birmingham: University of Birmingham.

Faber, P., León-Araúz, P., & Reimerink, A. (2014). Representing Environmental Knowledge in EcoLexicon. In E. Bárcena et al. (Eds.), *Languages for Specific Purposes in the Digital Era. Educational Linguistics*, *vol. 19* (pp. 267–301). Cham: Springer. https://doi.org/10.1007/978-3-319-02222-2_13

Faber, P., & León-Araúz, P. (2016). Specialized knowledge representation and the parameterization of context. *Frontiers in Psychology*, *7*(196). https://doi.org/10.3389/fpsyg.2016.00196

Fernández Fuertes, R., Álvarez de la Fuente, E., Parrado Román, I., & Muñiz Fernández, S. (2008). La "bici pirata" se convierte en "pirate bike" o en "bike pirate": la composición nominal en datos de adquisición de niños. In L. Pérez Ruiz, I. Pizarro Sánchez, & E. González-Cascos Jiménez (Eds.), *Estudios de metodología de la lengua inglesa (IV)* (pp. 421–438). Valladolid: Universidad de Valladolid.

Freixa, J. (2006). Causes of Denominative Variation in Terminology. *Terminology*, *12*(1), 51–77. https://doi.org/10.1075/term.12.1.04fre

Gagné, C.L. (2000). Relational-Based Combinations Versus Property-Based Combinations: A Test of the CARIN Theory and the Dual-Process Theory of Conceptual Combination. *Journal of Memory and Language*, *42*, 365–389. https://doi.org/10.1006/jmla.1999.2683

Gangemi, A., Pisanelli, D.M., & Steve, G. (2000). Understanding Systematic Conceptual Structures in Polysemous Medical Terms. In M.J. Overhage (Ed.), *Proceedings of the AMIA Symposium 2000* (pp. 285–289). Philadelphia, PA: Hanley & Belfus.

Geeraerts, D., Grondelaers, S., & Bakema, P. (1994). *The structure of lexical variation. Meaning, naming, and context*. Berlin/New York: Mouton de Gruyter. https://doi.org/10.1515/9783110873061

Gledhill, C., & Pecman, M. (2018). On Alternating Pre-Modified and Post-Modified Nominals Such As Aspirin Synthesis Versus Synthesis of Aspirin: Rhetorical and Cognitive Packing in English Science Writing. *Fachsprache*, *1*(2), 26–48.

Haralambous, Y., & Lavagnino, E. (2011). La réduction de termes complexes dans les langues de spécialité. *Traitement Automatique des Langues (TAL)*, *52*(1), 37–68.

Harastani, R., Daille, B., & Morin, E. (2013). Identification, alignement, et traductions des adjectifs relationnels en corpus comparables. In *Vingtième conférence du Traitement Automatique du Langage Naturel 2013 (TALN 2013)* (pp. 313–326). Sables d'Olonne, France: ATALA.

Kilgarriff, A., & Renau, I. (2013). EsTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia Social and Behavioral Sciences*, *95*, 12–19. https://doi.org/10.1016/j.sbspro.2013.10.617

Kilgarriff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX International Congress* (pp. 105–115). Lorient: EURALEX.

Langacker, R.W. (2008). *Cognitive Grammar*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Lees, R.B. (1960). *The Grammar of English Nominalizations*. Bloomington, IN: Indiana University Press/The Hague: Mouton.

León-Araúz, P. (2017). Term and Concept Variation in Specialized Knowledge Dynamics. In P. Drouin et al. (Eds.), *Multiple Perspectives on Terminological Variation* (pp. 213–258). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.18.09leo

Levi, J.N. (1978). *The Syntax and Semantics of Complex Nominals*. New York, NY: Academic Press.

Liceras, J.M. (2001). La teoría lingüística y la composición nominal del español y del inglés. In P. Fernández Nistal & J.M. Bravo (Eds.), *Pathways of Translation Studies* (pp. 229–248). Valladolid: S.A.E. University of Valladolid.

Liceras, J.M., Díaz, L., & Salomaa-Robertson, T. (2002). Processing versus representational difficulty in the acquisition of Spanish N-N Compounding. In A.T. Pérez-Leroux & J.M. Liceras (Eds.), *The Acquisition of Spanish Morphosyntax* (pp. 209–237). Dordrecht: Kluwer. https://doi.org/10.1007/978-94-010-0291-2_8

López Hernández, F., & Bernardos Galindo, S. (2018). A relational adjective and a noun semantic binding in the specialized language of Information and Communication Technology. *RESLA*, *31*(1), 197–223. https://doi.org/10.1075/resla.15037.lop

Lorente Casafont, M., Martínez Salom, M.A., Santamaría-Perez, I., & Vargas-Sierra, C. (2017). Specialized collocations in specialized dictionaries. In S. Torner Castells & E. Bernal (Eds.), *Collocations and other lexical combinations in Spanish: theoretical, lexicographical and applied perspectives* (pp. 200–222). London/New York: Routledge.

Lyons, J. (1995). *Linguistic Semantics: An Introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511810213

Mairal Usón, R., & Cortés Rodríguez, F.J. (2000). Semantic packaging and syntactic projections in word formation processes: the case of agent nominalizations. *RESLA*, *14*, 271–294.

Maniez, F. (2009). L'adjectif dénominal en langue de spécialité: étude du domaine de la médicine. *Revue française de linguistique appliquée*, *14*, 117–130. https://doi.org/10.3917/rfla.142.0117

Maniez, F. (2014). Implantation of English Terms Including the -ING Morpheme in French, Spanish and Italian: A Corpus-Based Study of the Debates of the European Parliament. In P. Dury et al. (Eds.), *La néologie en langue de spécialité : détection, implantation et circulation des nouveaux termes* (pp. 189–201). Lyon: Travaux du CRTT.

Martín Arista, J. (1997). La representación subyacente de los compuestos nominales en una gramática funcional del inglés. *Atlantis*, *19*(2), 169–175.

Mélis-Puchulu, A. (1991). Les adjectifs dénominaux: des adjectifs de relation. *Lexique*, *10*, 33–60.

Moreno-Fernández, F. (2003). Anglicismos en el léxico disponible de los adolescentes hispanos de Chicago. In K. Potowski & R. Cameron (Eds.), *Spanish in Contact* (pp. 41–58). Amsterdam: John Benjamins Publishing Company.https://doi.org/10.1075/impact.22.05mor

Murphy, G.L. (1988). Comprehending complex concepts. *Cognitive Science*, *12*, 529–562. https://doi.org/10.1207/s15516709cog1204_2

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of Memory and Language*, *29*, 259–288. https://doi.org/10.1016/0749-596X(90)90001-G

Nakov, P. (2013). On the Interpretation of Noun Compounds: Syntax, Semantics, and Entailment. *Natural Language Engineering*, *19*(3), 291–330. https://doi.org/10.1017/S1351324913000065

Nakov, P., & Hearst, M. (2006). Using Verbs to Characterize Noun-Noun Relations. In J. Euzenat & J. Domingue (Eds.), *Artificial Intelligence: Methodology, Systems, and Applcations. AIMSA 2006* (pp. 233–244). Berlin: Springer.

Nastase, V., & Szpakowicz, S. (2003). Exploring Noun-Modifier Semantic Relations. In *Fifth International Workshop on Computational Semantics* (pp. 285–301). Tilburg: IWCS.

Padró, L., Collado, M., Reese, S., Lloberes, M., & Castellón, I. (2010). FreeLing 2.1: Five Years of Open-Source Language Processing Tools. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odij, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2010)* (pp. 931–936). La Valletta: ELRA.

Palacios Martínez, I. (2014). Variation, development and pragmatic uses of *innit* in the language of British adults and teenagers. *English Language and Linguistics*, *19*(3), 383–405. https://doi.org/10.1017/S1360674314000288

Pecman, M. (2012). Tentativeness in Term Formation. A Study of Neology as a Rhetorical Device in Scientific Papers. *Terminology*, *18*(1), 27–58. https://doi.org/10.1075/term.18.1.03pec

Postal, P. M. (1969). Anaphoric islands. In R. I. Binnick et al. (Eds.), *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society* (pp. 209–239). Chicago: University of Chicago.

Rodríguez-Juárez, C. (2017). Accounting for the Alternating Behaviour of Location Arguments from the Perspective of Role and Reference Grammar. *ATLANTIS, Journal of the Spanish Association of Anglo-American Studies*, *39*(2), 169–189. https://doi.org/10.28914/Atlantis-2017-39.2.09

Rosario, B., Hearst, M., & Fillmore, C. (2002). The Descent of Hierarchy, and Selection in Relational Semantics. In P. Isabelle (Ed.), *ACL '02 Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 247–254). Stroudsburg, PA: Association for Computational Linguistics.

Sager, J. C., Dungworth, D., & McDonald, P. F. (1980). *English Special Languages: Principles and Practice in Science and Technology*. Wiesbaden: Oscar Branstetter Verlag KG.

Sanz Vicente, L. (2012). Approaching Secondary Term Formation through the Analysis of Multiword units: An English–Spanish Contrastive Study. *Terminology*, *18*(1), 105–127. https://doi.org/10.1075/term.18.1.06san

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Smith, J., Durham, M., & Richards, H. (2013). The social and linguistic in the acquisition of sociolinguistic variation. *Linguistics*, *51*(2), 258–324. https://doi.org/10.1515/ling-2013-0012

Štekauer, P., Valera, S., & Kőrtvélyessy, L. (2012). *Word-Formation in the World's Languages*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511895005

Temmerman, R. (2000). *Towards New Ways of Terminology Description: The Sociocognitive Approach*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.3

Vanderwende, L. (1994). Algorithm for Automatic Interpretation of Noun Sequences. In M. Nagao (Ed.), *COLING '94 Proceedings of the 15th International Conference on Computational Linguistics* (pp. 782–788). Stroudsburg, PA: Association for Computational Linguistics. https://doi.org/10.3115/991250.991272

Warren, B. (1978). *Semantic Patterns of Noun-Noun Compounds*. Göteborg: Acta Universitatis Gothoburgensis.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, *35*, 434–453. https://doi.org/10.1006/jmla.1996.0024

La variación en los términos compuestos: Compuestos nominales preposicionales y adjetivales en español

Resumen

Los compuestos nominales (CN) son frecuentes en el discurso especializado en todas las lenguas, ya que se trata de un método productivo de crear términos mediante la combinación de unidades léxicas existentes. En español, estas combinaciones conceptuales pueden adquirir la forma de CN preposicionales (CNP) o los CN adjetivales (CNA) equivalentes, p. ej. *demanda de electricidad* vs. *demanda eléctrica* [electricity demand]. Los adjetivos de los CNA, que suelen proceder de sustantivos, se conocen como 'adjetivos relacionales', pues codifican relaciones semánticas con otros conceptos. Excepto algunos estudios recientes, la investigación se ha centrado en las relaciones semánticas que subyacen en los CN. En el ámbito del procesamiento del lenguaje natural, varios estudios han abordado la detección automática de los adjetivos relacionales en lenguas romances y germánicas. Sin embargo, hasta donde alcanza nuestro conocimiento, no se han realizado estudios discursivos de estos CN encaminados a establecer recomendaciones de redacción. En este estudio se analiza el cotexto de CNP y CNA equivalentes para identificar factores que guíen el uso de una determinada forma. Utilizamos el corpus EcoLexicon ES, formado por textos especializados en español sobre el medio ambiente, para extraer 6 adjetivos relacionales y, a continuación, 12 pares de CN equivalentes. Analizamos su comportamiento en cotexto mediante 20 expresiones en CQP aplicadas en el corpus EcoLexicon ES y un corpus general. Los resultados muestran que el cotexto lingüístico inmediato determina la preferencia por una determinada estructura. Según estos resultados, proponemos pautas de redacción para guiar en la producción de CN.

Palabras clave: término compuesto, compuesto nominal, variación denominativa, adjetivo relacional, contexto, lenguaje especializado

Address for correspondence

Melania Cabezas-García
Departamento de Traducción e Interpretación
Universidad de Granada
C/ Buensuceso, 11
18002 Granada
Spain
melaniacabezas@ugr.es
https://orcid.org/0000-0002-8622-1036

Co-author information

Santiago Chambó
Departamento de Traducción e Interpretación
Universidad de Granada
santiagochambo@ugr.es
https://orcid.org/0000-0002-8219-3621

Publication history

Date received: 21 January 2019
Date accepted: 30 September 2019