



**UNIVERSIDAD
DE GRANADA**

**Facultad de Traducción e
Interpretación**

**GRADO EN TRADUCCIÓN E
INTERPRETACIÓN**

TRABAJO FIN DE GRADO

**CREACIÓN DE UNA BASE DE DATOS
BILINGÜE (INGLÉS-ESPAÑOL) PARA
LA ASIGNATURA DE CIENCIAS DE LA
EDUCACIÓN PRIMARIA**

Presentado por:

D. Federico Ortega Riba

Tutor:

Dra. Míriam Buendía Castro

Curso académico 2022/2023

ÍNDICE DE CONTENIDOS

ABSTRACT	4
RESUMEN	5
1. INTRODUCCIÓN	6
2. ESTADO DE LA CUESTIÓN	7
2.1. EL DISCURSO GENERAL Y EL ESPECIALIZADO: EL CASO DE LA EDUCACIÓN PRIMARIA	7
2.2. RECURSOS TERMINOGRÁFICOS ESPECIALIZADOS EN ESPAÑA	11
2.3. COMPETENCIAS LINGÜÍSTICAS Y DIGITALES EN EDUCACIÓN PRIMARIA	13
3. METODOLOGÍA	14
4. ANÁLISIS DE LOS CORPUS CON SKETCH ENGINE Y EXCEL	16
5. RESULTADOS	19
6. CREACIÓN DE LA BASE DE DATOS Y PROPUESTA DE ENTRADA	20
7. CONCLUSIONES	25
8. BIBLIOGRAFÍA	26

Abstract

Terminographic resources in Spain such as databases, glossaries or multilingual dictionaries created by Royal Academies or other organizations are designed by specialists and aimed at experts. With regard to education, the Ministry of Education and Vocational Training proposes to consolidate multilingual and technological skills at the end of Primary Education, and to promote open educational resources. Nevertheless, there is currently a lack of linguistic infrastructures for students enrolled in bilingual subjects (Spanish-English) at this stage. The aim of this work is to compile a corpus to serve as a basis for analyzing the most frequent monolexical and polylexical units in the subdomains of Natural Science. Based on the results, we present the creation of a database which meets the terminological needs of students between the ages of six and twelve. This resource is proposed as a future reference work for teachers and students of Primary Education.

Keywords: terminology, corpus linguistics, linguistic analysis, terminographic resource, Natural Science, Primary Education

Resumen

Los recursos terminográficos en España como bases de datos, glosarios o diccionarios multilingües creados por Reales Academias u otros organismos están diseñados por especialistas y dirigidos a expertos. En lo que respecta a la enseñanza, el Ministerio de Educación y Formación Profesional propone consolidar las competencias plurilingües y tecnológicas al acabar la Educación Primaria y promover los recursos educativos abiertos. Sin embargo, en la actualidad no existen infraestructuras lingüísticas para estudiantes de esta etapa que cursen asignaturas en modalidad bilingüe (español-inglés). Mediante este trabajo, se pretende compilar un corpus que sirva de base para analizar las unidades monoléxicas y poliléxicas más frecuentes en los subdominios de Ciencias de la Naturaleza. Partiendo de los resultados, se propone la creación de una base de datos que se ajuste a las necesidades terminológicas del estudiantado de entre seis y doce años. Este recurso se plantea como una futura obra de referencia para docentes y alumnos de Educación Primaria.

Palabras clave: terminología, lingüística de corpus, análisis lingüístico, recurso terminográfico, Ciencias de la Naturaleza, Educación Primaria

1. Introducción

Los recursos terminográficos son herramientas que nos permiten homogeneizar el conocimiento lingüístico y resolver problemas relacionados con la terminología, como contextos de uso o colocaciones. Asimismo, son utilizados en la enseñanza de idiomas en forma de obras de consulta (base de datos, diccionarios o glosarios). Para crear y ampliar recursos terminográficos de calidad, es necesario poseer conocimientos especializados en dos ámbitos: lingüística y tecnología. Las humanidades digitales y la lingüística computacional nos permiten llevar a cabo tareas específicas en el ámbito del procesamiento del lenguaje natural mediante programas como Sketch Engine (Kilgarriff et al. 2004), mientras que otras aplicaciones están diseñadas para trabajar con datos que, de otra forma, programas como Sketch Engine no nos permitirían, sobre todo en lo referente a análisis estadísticos. Es el caso de las funciones de Excel (Microsoft Corporation, 2018) o las de Python (Van Rossum y Drake, 2009).

En lo que respecta a de bases de datos lingüísticas, el punto de partida es el corpus. El *Diccionario de la lengua española* de la RAE define *corpus* como «conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación»¹, mientras que John Sinclair (1991, p. 171), el padre de la lingüística de corpus, sostiene que «a corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language». La lingüística de corpus clasifica estos conjuntos ateniéndose a diversos parámetros como su objetivo, accesibilidad o formato; si bien este último punto es cada vez menos relevante, ya que encontraríamos limitaciones para acceder a ellos de otra forma que no fuese la digital. Además de esta clasificación, las características de los corpus los diferencian en anotados (etiquetas morfológicas, sintácticas, etc.), de referencia (recogen la lengua general), de aprendices de segundas lenguas extranjeras, entre otras categorías.

Desde el punto de vista de un traductor, la importancia del corpus reside en el idioma o idiomas en los que se configure, así como su acceso. Este trabajo está centrado en la creación de corpus monolingües, comparables y escritos, que son las características básicas de los corpus compilados para la creación de la base de datos. En primer lugar, se establece que son monolingües porque han sido compilados individualmente en inglés y

¹ Real Academia Española. (s.f.). Corpus. En *Diccionario de la lengua española*. Recuperado el 26 de mayo de 2023, de <https://dle.rae.es/corpus>

español. En segundo lugar, son escritos, a diferencia de los orales. En última instancia, son comparables. Peters et al. (1996) define los corpus comparables como:

Un conjunto de textos en más de una lengua que, sin ser traducciones, por coincidir en el tema y en la función comunicativa proporcionan una excelente base para la comparación de dos o más lenguas. (citado en López Rodríguez, 2001, p. 175)

Las ventajas de la creación y utilización de corpus para traductores se resumen en los siguientes puntos, como sostiene Bowker (2002): extracción de frecuencias de palabras para saber el uso de una unidad y contextos de uso sobre cómo se expresa una unidad en la lengua de estudio. A partir del conocimiento que nos proporciona el análisis de corpus, podemos crear herramientas representativas que sirvan como obra de referencia a un público concreto. En los siguientes apartados de este Trabajo de Fin de Grado, se exponen las bases teórico-prácticas que se deben consolidar para que el desarrollo y análisis de un corpus sean óptimos. Una vez definidas, se propone la creación de una base de datos bilingüe (español-inglés) para la asignatura Ciencias de la Naturaleza de Educación Primaria a partir de los corpus compilados.

2. Estado de la cuestión

2.1. El discurso general y el especializado: el caso de la Educación Primaria

Uno de los principales problemas que conlleva la creación de diccionarios, sobre todo aquellos enfocados al estudiantado de entre seis y doce años, es la dicotomía entre discurso general y especializado, entre término y palabra. Si estudiamos el enfoque normalizador de la terminología, de acuerdo con la norma ISO 1087 *Terminology work and terminology science – Vocabulary* (2019) el término es «a designation that represents a general concept by linguistic means». Del mismo modo, se explicita que puede consistir en una o más palabras.

Aunque este organismo de estandarización sea reconocido de manera internacional, la definición no es suficientemente limitante y nos lleva a preguntarnos cuáles son las diferencias entre el discurso general y el especializado. A esta problemática le sumamos otra: a la hora de extraer las palabras cuyas frecuencias son más elevadas en un corpus, las herramientas informáticas solo nos devuelven las unidades monoléxicas y poliléxicas más empleadas en un campo de especialidad. No obstante, cuando se analiza un corpus con información destinada a estudiantes de primaria, encontraremos entre las frecuencias más altas palabras que a primera vista no pasarían por discurso especializado.

Los lenguajes de especialidad se entienden como lenguas naturales que sirven como instrumento para compartir conocimiento especializado. En el proceso de la comunicación entre especialistas, Jakobson (1963) enumera factores como el emisor, el canal, el código, el receptor, el mensaje y el referente. Mientras que el discurso general, ya sea oral u escrito, se define como aquel usado en la vida cotidiana para producir una reacción en el receptor (Alcaraz y Martínez, 1997). El discurso general también se caracteriza por ser propio de una comunidad lingüística determinada. Sus rasgos están delimitados por su contexto cotidiano, lo que lleva a una variación a nivel dialectal, en geolecto, sociolecto y cronolecto, y funcional, que se manifiesta en: discurso normalmente oral, temática no especializada, tenor referencial y registro informal con bajo grado de abstracción.

A diferencia del lenguaje general, el especializado trata temáticas, ámbitos y experiencias especializadas, no son fenómenos aislados y también mantiene el tenor referencial (Cabré, 1993, p. 135). Desde el punto de vista de la traducción, el discurso especializado destaca por su idiosincrasia léxica sin estar separado del lenguaje general, es decir, se puede entender como un subconjunto de este, usado por especialistas que comparten códigos y tenor. Sager (1990, p. 18) afirma: «the linguistic subsystem selected by an individual whose discourse is to be centred on a particular subfield».

Además, dentro del propio discurso especializado pueden desarrollarse otros subdiscursos especializados según el conocimiento de la materia entre los interlocutores. No es lo mismo el nivel de lengua de dos oncólogos que están al mismo nivel, al de un oncólogo y un niño de doce años. Debido a estas casuísticas, que dependen estrechamente de la pragmática del discurso especializado, Cabré considera que existe una intersección entre lengua general y lenguajes especializados (Figura 1).

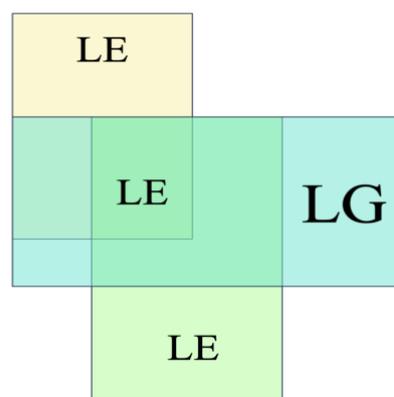


Figura 1. Los lenguajes de especialidad (LE) y la lengua general (LG). Adaptado de Gómez (2005, 46).

Para ilustrar esto, tomemos como ejemplo un tema de Ciencias de la Naturaleza sobre el cuerpo humano en Educación Primaria. Cuando se habla del cuerpo, la voz *mano* es aquí una unidad del discurso especializado orientada a niños de esta etapa educativa. Sin embargo, en un contexto especializado para anatomía del tronco superior, la voz *mano* pasaría a ser *parte distal del miembro superior*. Al igual que decía Gutiérrez (1998) en la siguiente afirmación:

Lo que «playa» significa para cada uno de nosotros tiene poco que ver con un «depósito arenoso en forma de suave talud inclinado hacia el mar, formado por partículas procedentes de la erosión costera y situado sobre la plataforma litoral». (p. 105)

Al existir un *continuum* del conocimiento y no estar dividido en compartimentos estancos, si se observa un término desde diferentes perspectivas, se asociará al discurso especializado o al discurso general. El momento en el que pasa al discurso especializado se denomina terminologización. Para ilustrar esto, la voz *resistencia*, según el *DRAE*, se define como «acción y efecto de resistir o resistirse»; pero también en el ámbito de la electricidad «elemento que se intercala en un circuito para modificar el paso de la corriente o para producir calor»².

Al igual que existe la terminologización, también se da el caso contrario en el proceso denominado banalización, cuyo origen atribuye Cabré a los medios de comunicación de masas y al acceso a la enseñanza universal. Es el caso de la voz *pélet* (adaptación de *pellet*), según la Fundéu: «para aludir a las piezas cilíndricas prensadas de materiales variados y que sirven como combustible para calderas o estufas»³.

Por otra parte, Cabré también describe las variaciones del lenguaje especializado a través de dos ejes, el horizontal y el vertical. En cuanto a la primera división, se basa en la multidimensionalidad de las perspectivas utilizadas para tratar una temática concreta. La temática define el discurso especializado, pero la perspectiva es definitoria para conseguir el objetivo que se persigue; se puede tratar un discurso especializado de manera trivial como si fuese conocimiento general o se puede abordar desde el punto de vista especializado.

En la misma línea del ejemplo anterior, la voz *mano* también se podría utilizar en un temario sobre anatomía del tronco inferior para estudiantes universitarios de Medicina,

² Real Academia Española. (s.f.). Resistencia. En *Diccionario de la lengua española*. Recuperado en 26 de mayo de 2023, de <https://dle.rae.es/resistencia?m=form>

³ FundéuRAE. (25 noviembre de 2022). «*pélet*», adaptación de «*pellet*». Recuperado el 27 de mayo de 2023. <https://www.fundeu.es/recomendacion/pelet-adaptacion-de-pellet/#:~:text=La%20forma%20p%C3%A9let%20es%20una,combustible%20para%20calderas%20o%20estufas.>

pues perdería importancia en ese contexto. En lo referente a la variación vertical, son los participantes los que determinan el nivel de especialización del discurso, de ahí que el emisor transmita un mensaje en función del grado de familiaridad del receptor con la disciplina tratada. En el caso de Ciencias de la Naturaleza, podemos tomar las cuatro variedades que Paepcke (1975) propone en el discurso técnico, con ejemplos propios relacionados con la voz *mano*:

1. científico: digital palmar común de las ramas del nervio mediano;
2. nivel de taller: nervio mediano de la mano;
3. nivel de uso cotidiano: nervio que conecta con el pulgar, índice y cordial;
4. nivel de publicidad/ventas: uno de los nervios de la mano.

En función de los participantes de la situación comunicativa también se distinguen cuatro tipos de discursos especializados:

1. de especialista a especialista,
2. de especialista a semi-especialista,
3. de especialista a aprendiz de especialista,
4. de especialista a público general.

Cuando hablamos de terminología sobre la asignatura Ciencias de la Naturaleza para Educación Primaria, los límites de las variedades del discurso técnico y las situaciones comunicativas son difusas. Con el fin de establecer una categoría, podríamos suponer un discurso de taller y una situación de especialista a público general. Ahora bien, teniendo en cuenta que la Educación Primaria está dirigida al estudiantado de entre seis y doce años, esta división es susceptible a cambios.

En definitiva, con el objetivo de suplir las carencias y establecer normas en el debate sobre qué se considera término y qué no se considera término, tomaremos como referencia la Teoría Comunicativa de la Terminología (TCT) de Teresa Cabré, que sostiene la no distinción entre el conocimiento general y el especializado. En nuestra base de datos estarán incluidas aquellas unidades de un léxico fronterizo entre la lengua común y la especializada, al no haber una delimitación exacta entre los dos. A través de criterios pragmáticos y comunicativos, determinamos la presencia de términos y vocabulario general. Investigaremos el contexto y el conocimiento que comparten los alumnos de Educación Primaria en una misma situación comunicativa o subdominio dentro de Ciencias de la Naturaleza.

2.2. Recursos terminográficos especializados en España

Incluso limitando las búsquedas con operadores booleanos, se desaconseja a los traductores el uso de un buscador web como corpus. Dada la cantidad de información que nos proporciona internet y el hecho de que no exista una organización encargada de filtrar la relevancia de los textos publicados, se hace necesario que los traductores profesionales compilen corpus a partir de recursos fiables o los extraigan de manera automática utilizando programas adecuados para ello: el denominado *web scraping* en el ámbito de la lingüística computacional.

Los parámetros que deben observarse a la hora de seleccionar información para el corpus deben ser la autoridad, el contenido y el diseño (Buendía y Ureña, 2010). En España existen numerosos recursos terminográficos monolingües que se acogen a dichos parámetros y cuyo uso es frecuente entre terminólogos y traductores. Entre los corpus de referencia más relevantes, destacan el Corpus del Español del siglo XXII (CORPES), el Corpus de Referencia del Español Actual (CREA) y el Corpus Diacrónico del Español (CORDE), todos ellos accesibles a través del banco de datos de la RAE. Dichos corpus son óptimos para estudiar los problemas planteados en los objetivos de su creación, ya sea la representación textual del conjunto de países hispanohablantes desde 2001 hasta la actualidad, como es el caso del CORPES; ya sea la representación textual de todas las épocas y lugares donde se habló el idioma hasta 1974, como es el caso del CORDE. Dicho de otro modo, tienen un alto grado de representatividad (Bieber, 1993), puesto que reflejan la lengua de estudio y su variabilidad.

En general, y como se puede inferir de los ejemplos anteriores, las Reales Academias se encargan de recoger sus propios términos para permitir la homogeneidad, favorecer la inclusión de neologismos o préstamos nuevos, así como equilibrar y actualizar las distintas áreas del conocimiento. Estos recursos suelen estar creados por especialistas y lingüistas computacionales que respetan los criterios de la Comisión Lingüística para la Terminología Española (COLTE). Entre los recursos disponibles para áreas especializadas, podemos encontrar los siguientes: el Corpus Científico-Técnico (CCT) de la Real Academia Española, el *Diccionario panhispánico del español jurídico* de la misma institución, el *Diccionario Español de Energía* de la Real Academia de Ingeniería, el *Diccionario de términos médicos* (DTM) de la Real Academia Nacional de Medicina, el Vocabulario científico y técnico de la Real Academia de Ciencias Exactas, Físicas y Naturales, así como el *Diccionario Terminológico de las Ciencias*

Farmacéuticas de la Real Academia Nacional de Farmacia. La mayoría de ellos dispone de términos con equivalentes en inglés e incluso otras lenguas.

Existen otros recursos terminográficos bilingües y multilingües designados como bancos de datos y glosarios específicos del ámbito científico-técnico. Algunas recomendaciones de la Dirección General de Traducción de la Comisión Europea incluyen recursos como IATE (*Inter-Active Terminology for Europe*)⁴, desarrollado por la Unión Europea en todas sus lenguas oficiales, que ayuda a divulgar, recopilar y gestionar terminología; UNTerm⁵, el mismo sistema que IATE pero desarrollado por las Naciones Unidas; TERMCAT⁶, terminología catalana con equivalencias en inglés, francés y español; o TERMIUM Plus⁷, por el Gobierno de Canadá. Entre otros de dominio público pero menos frecuentados por traductores encontramos: glosario de medio ambiente (AEMA), Electropedia (CEI) o Estudio de los Lenguajes Especializados en Español (ESLEE).

Al margen de estas herramientas, podemos destacar otros materiales como EcoLexicon⁸, base de datos desarrollada por el grupo LexiCon de la Universidad de Granada; Cosnautas⁹, por Fernando A. Navarro, y Varimed¹⁰, diccionario sobre variaciones terminológicas médicas en inglés y español.

Debemos recordar que los miembros e instituciones que colaboran en la creación y desarrollo de estas infraestructuras lingüísticas y favorecen el uso común de las lenguas especializadas suelen ser miembros de las diversas asociaciones terminológicas que existen a nivel nacional e internacional. En España, la Asociación Española de Terminología (AETER) es la más reconocida, formando parte a su vez de la Asociación Europea de Terminología (AET). Desde el 1997 organiza jornadas y se encarga de coordinar las Reales Academias.

Teniendo en cuenta lo expuesto anteriormente, para un alumno o alumna de Educación Primaria estos recursos resultan de difícil comprensión, al ser lingüísticamente más complejos. El único recurso terminográfico monolingüe del ámbito de Ciencias de la Naturaleza dirigido a estudiantes de Educación Primaria que existe en España es el *Primer diccionario de medicina ilustrado* (Estopà, 2018). En él, se explica el significado

⁴ <https://iate.europa.eu/home>

⁵ <https://unterm.un.org/unterm2/en>

⁶ <https://www.termcat.cat/es>

⁷ <https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng>

⁸ http://ecolexicon.ugr.es/visual/index_en.html

⁹ <https://www.cosnautas.com/es>

¹⁰ <https://varimed.ugr.es/>

de las unidades, se proporciona su categoría gramatical, palabras que tienen una entrada en alguna otra parte del diccionario, audio con explicación y un dibujo que ayuda a entender los términos. A pesar de ser un recurso monolingüe muy completo, la realidad bilingüe en España obliga a la creación de una base de datos como la que se plantea en este trabajo.

2.3. Competencias lingüísticas y digitales en Educación Primaria

El Ministerio de Educación y Formación Profesional dispone en el portal del sistema educativo español una serie de competencias clave para los currículos básicos por enseñanzas, asentadas en el Real Decreto 157/2022 por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. A la hora de evaluar el bilingüismo en España y la forma en que los alumnos pueden acceder a recursos terminográficos, es interesante conocer tres de estas competencias propuestas: la competencia en comunicación lingüística, la competencia plurilingüe y la competencia digital.

Por una parte, la competencia en comunicación lingüística es aquella por la que se espera que los alumnos puedan interactuar de forma multimodal en diferentes eventos comunicativos. Los centros de Educación Primaria deben instar a sus alumnos a evitar la desinformación digital y a comunicarse de manera exitosa a través de las destrezas y actitudes que les permiten entender, interpretar y evaluar críticamente un mensaje en sus diferentes canales.

Por otra parte, la competencia plurilingüe propone el uso de lenguas destinado al aprendizaje y la comunicación. Su enfoque incluye el respeto y reconocimiento de los perfiles lingüísticos con el fin de promover planes que impulsen la mediación lingüística y ayuden a la transferencia de conocimiento lingüístico, incluyendo lenguas clásicas, familiares y oficiales. Al completar la Educación Primaria, esta competencia permitirá el uso de más de un idioma, además del materno, para responder a situaciones en contextos cotidianos, hacer transferencias sencillas entre idiomas para entenderse en diferentes eventos lingüísticos básicos, ampliar el léxico individual en otros idiomas y respetar la diversidad lingüística que existe en España. Dicha competencia está estrechamente relacionada con la competencia tecnológica, por la cual se espera que los alumnos y alumnas sean capaces de realizar búsquedas en internet para construir nuevo conocimiento con una actitud crítica.

Con la base de datos que proponemos, también se incluyen otras competencias como la alfabetización digital en información y datos, para que el alumnado sepa utilizar de la manera más adecuada la página web; integrar contenidos digitales en formatos de texto, imagen o audio para comunicar ideas mientras respetan los derechos de propiedad intelectual; y la iniciación en el desarrollo de soluciones.

El Trabajo de Fin de Grado que se plantea se ha pensado con base en dos planes impulsados por el Gobierno: el Plan de Impulso de las Tecnologías del Lenguaje, que promueve el Ministerio de Asuntos Económicos y Transformación Digital, así como el Plan de Digitalización y Competencias Digitales del Sistema Educativo, promovido por el Ministerio de Educación y Formación Profesional. Si tomamos las competencias descritas anteriormente, existe la necesidad de desarrollar Recursos Educativos Abiertos (REA) y, en este caso, en lengua inglesa. Según las Guías de la BUH (2023):

Los REA, conocidos en inglés como Open Educational Resources (OER) son documentos, materiales y contenidos educativos ofrecidos libre y abiertamente para que cualquiera los pueda utilizar para enseñar, aprender o investigar. Frecuentemente, estos Recursos Educativos Abiertos están distribuidos bajo una licencia Creative Commons que concede permisos y derechos al usuario bajo el marco de las cinco R.

En la misma línea, el Informe sobre el estado de las tecnologías del lenguaje en España dentro de la Agenda Digital para España (Ministerio de Industria, Energía y Turismo y Ministerio de Hacienda y Administraciones Públicas, 2013) sugiere garantizar la disponibilidad de aplicaciones avanzadas en español e inglés a través de un incremento en el número, calidad y variedad de recursos lingüísticos. Desde el propio Gobierno de España se insiste en la necesidad de trabajo conjunto entre lexicógrafos, terminólogos y lingüistas computacionales para crear plataformas tecnológicas de calidad.

Una vez analizada la infinidad de recursos especializados con los que contamos en español, también debemos considerar los problemas lingüísticos en estudiantes y otro público no especializado, ya que en ocasiones el uso de un diccionario no consigue eliminar las carencias de información de un estudiante de Educación Primaria. Por tanto, comprobamos de nuevo la necesidad imperiosa de crear un recurso con tales características.

3. Metodología

Para llevar a cabo un estudio exhaustivo, se ha compilado un corpus comparable para cada curso del conjunto de la edición de Anaya (2019) para la asignatura Ciencias de la

naturaleza y su versión inglesa, llamada *In focus*. De esta manera, se analizan las unidades de todos los subdominios, que son seis: la función de relación, la nutrición y la dieta, los seres vivos, los ecosistemas, la materia y la energía, y las máquinas. Quedan fuera del corpus los recursos electrónicos anexos: índices, repasos trimestrales, propuestas didácticas, así como la programación y documentación del proyecto, pues no demuestran gran especialización y tienen un enfoque más metodológico que descriptivo. En total, se han compilado 36 temas en español y 36 temas en inglés.

Por otra parte, cada tema compilado sigue el siguiente etiquetado: tema_editorial_idioma_curso_subdominio_extensión. Si tomamos como ejemplo el tema 1 de primero de primaria en español, nos quedaría 01ANES1RELTXT.

Una vez extraídos los temas de cada uno de los cursos, se han subido a la herramienta Sketch Engine¹¹. Cada corpus sigue el siguiente etiquetado: editorial_curso_idioma, y se han extraído tanto unidades monoléxicas como poliléxicas en español e inglés. La descripción de los corpus ha seguido el siguiente esquema: Anaya. Ciencias de la Naturaleza. [Curso].º Primaria para el español; y Anaya. Ciencias de la naturaleza In focus. [Curso].º Primaria para el inglés.

La elección de estas unidades se ha realizado según los criterios de relevancia expuestos anteriormente en el dominio de ciencias de la naturaleza, de ahí que se hayan eliminado unidades no especializadas. Recordemos a Cabré (2009): «el sentido especializado que hemos descrito como valor asociado a las unidades del léxico no es un conjunto predefinido y encapsulado de información, sino una selección específica de características semánticas según las condiciones de cada situación de uso».

Una particularidad de la compilación del corpus ha sido la dificultad de extracción de información. El formato digital de los libros no permite copiar ni descargar el contenido para que sea fácilmente extraíble en Sketch Engine. Para superar estas dificultades, primero se ha extraído el texto plano haciendo uso de un reconocimiento óptico de caracteres, a través del plugin *OCR - Image Reader* (Girko, 2023) disponible para Google Chrome, así como *Blackbox Code Search Autocomplete Extraction* (Blackbox, 2023). Se han ajustado sus parámetros en función del idioma y precisión. Una vez estos dos programas devolvían texto plano, se ha copiado en un documento .txt que ha sido revisado posteriormente y donde se han eliminado todos aquellos problemas de

¹¹ Sketch Engine (<https://www.sketchengine.eu/>) es una herramienta de gestión de corpus y análisis del procesamiento del lenguaje con funcionalidades diversas. Debido a su fiabilidad y rapidez, será la herramienta que se utilice de manera prioritaria a lo largo de este Trabajo de Fin de Grado.

reconocimiento como espacios en blanco, saltos de línea, caracteres ASCII no reconocidos y signos de puntuación incorrectos.

A través de un análisis exhaustivo, el planteamiento inicial incluye 100 unidades monoléxicas y 100 unidades poliléxicas, aunque en apartados posteriores comprobaremos que varían en función de las unidades que nos proporcione Sketch Engine. Las estadísticas de palabras extraídas para los corpus son las siguientes:

☰ AN_1_EN	English	2,020
☰ AN_1_ES	Spanish	5,533
☰ AN_2_EN	English	3,076
☰ AN_2_ES	Spanish	8,325
☰ AN_3_EN	English	5,534
☰ AN_3_ES	Spanish	16,742
☰ AN_4_EN	English	6,914
☰ AN_4_ES	Spanish	24,697
☰ AN_5_EN	English	7,228
☰ AN_5_ES	Spanish	23,663
☰ AN_6_EN	English	9,192
☰ AN_6_ES	Spanish	28,241

Figura 2. Codificación, idioma y número de palabras de cada corpus.

Los corpus donde se recogen las unidades de todos los cursos siguen el siguiente etiquetado: editorial_ALL_idioma. Se han recogido estos dos corpus adicionales para poder estudiar las concordancias y diseñar las entradas. Las estadísticas de palabras en este caso son:

☰ AN_ALL_EN	English	33,964
☰ AN_ALL_ES	Spanish	107,201

Figura 3. Codificación y número de palabras de los corpus en inglés y en español.

4. Análisis de los corpus con Sketch Engine y Excel

Una vez compilados los corpus y subidos a la plataforma (Figuras 2 y 3), se ha utilizado la función *Keywords* para extraer las palabras clave de cada uno de los temas compilados. Al realizar una comparación automática con corpus de referencia, podremos comprobar

qué unidades monoléxicas y poliléxicas (Figura 4) son más frecuentes en nuestros corpus especializados.



Figura 4. Extracción de unidades monoléxicas y poliléxicas. *Single-words* y *multi-word terms* en inglés, respectivamente.

Tomemos dos cursos como ejemplo. El primer nivel equivale a primero de primaria. El problema que se plantea es el hecho de que no hay suficientes palabras como para que el programa devuelva un máximo de 1000 (Figura 5). El segundo problema que se plantea es la frecuencia de aparición de unidades de la lengua general frente a unidades de significación especializada.

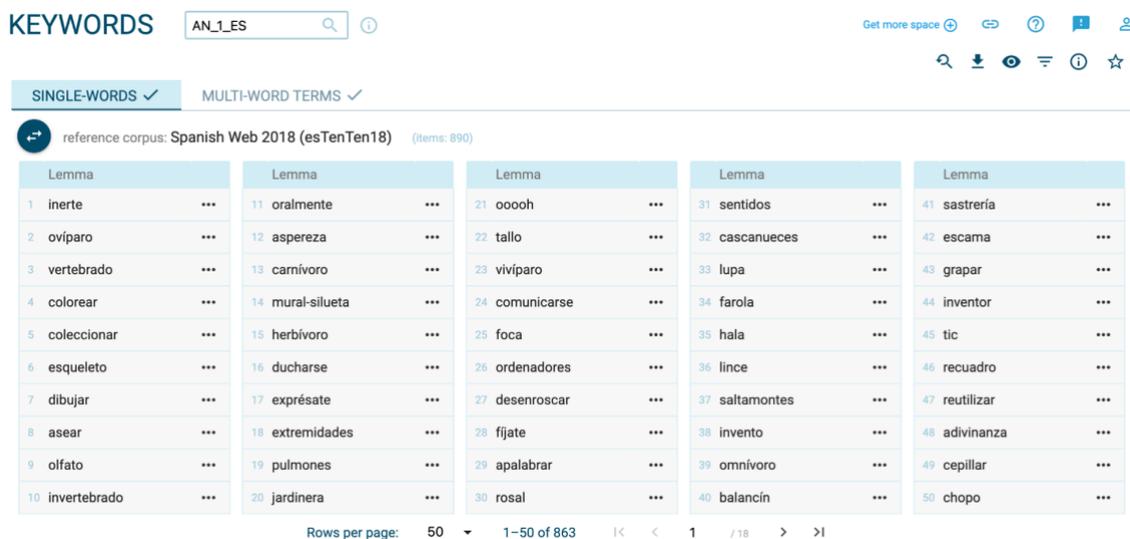


Figura 5. Ejemplo de extracción terminológica con la función *keywords* para primero de primaria en español.

Con el fin de asegurar la homogeneidad y respetar las pautas que hemos asentado en el marco teórico del trabajo, se han eliminado aquellas unidades no especializadas o no pertenecientes al subdominio de estudio. Entre ellas encontramos: *colorear*, *dibujar*, *asear*, *mural-silueta*, *exprésate* o *farola* (Figura 5). El hecho de que aparezcan como palabras clave se debe a que aparecen en ejercicios de las unidades o en las introducciones y páginas finales que hay en cada tema de la asignatura. Al revisar manualmente todas las unidades, limitamos los términos válidos a 99 en cuanto a unidades monoléxicas y 21 en unidades poliléxicas. Si calculamos el porcentaje de unidades que son adecuadas para incluirlas en nuestra base de datos, seguimos la siguiente fórmula:

$$\text{Porcentaje de unidades finales} = \frac{\text{Unidades seleccionadas} * 100}{\text{Total de keywords}}$$

Al seguir esta fórmula vemos que el 11,04 % de las *keywords* son aptas para incluirlas en nuestra base de datos. Ahora bien, si comparamos el número de unidades para sexto de primaria, limitamos los términos válidos a 263 unidades monoléxicas y 71 unidades poliléxicas.

The screenshot shows a web interface for keyword extraction. At the top, there is a search bar with 'AN_6_ES' and a search icon. Below the search bar, there are tabs for 'SINGLE-WORDS' (selected) and 'MULTI-WORD TERMS'. The main content area displays a table of 50 terms, each with a number, a lemma, and a three-dot menu icon. The terms are: 1. protozoo, 2. pluricelular, 3. heterótrofo, 4. autótrofo, 5. unicelular, 6. eucarioto, 7. imán, 8. protoctistas, 9. excretor, 10. biotopo, 11. ovíparo, 12. alvéolo, 13. biocenosis, 14. electroimán, 15. trófico, 16. invertebrado, 17. artrópodo, 18. vertebrado, 19. magnetismo, 20. fotosíntesis, 21. moneras, 22. alga, 23. efector, 24. branquia, 25. locomotor, 26. procariota, 27. órgano, 28. asexual, 29. aurícula, 30. citoplasma, 31. quimo, 32. excreción, 33. pubertad, 34. molusco, 35. porífero, 36. pseudópodo, 37. dióxido, 38. cigoto, 39. vivíparo, 40. refracción, 41. circulatorio, 42. espermatozoide, 43. salmonelosis, 44. quilo, 45. brújula, 46. hongo, 47. cnidario, 48. descomponedor, 49. oxígeno, 50. electromagnetismo. At the bottom of the table, there is a pagination control showing 'Rows per page: 50' and '1-50 of 1,000'.

Lemma	Lemma	Lemma	Lemma	Lemma
1 protozoo	11 ovíparo	21 moneras	31 quimo	41 circulatorio
2 pluricelular	12 alvéolo	22 alga	32 excreción	42 espermatozoide
3 heterótrofo	13 biocenosis	23 efector	33 pubertad	43 salmonelosis
4 autótrofo	14 electroimán	24 branquia	34 molusco	44 quilo
5 unicelular	15 trófico	25 locomotor	35 porífero	45 brújula
6 eucarioto	16 invertebrado	26 procariota	36 pseudópodo	46 hongo
7 imán	17 artrópodo	27 órgano	37 dióxido	47 cnidario
8 protoctistas	18 vertebrado	28 asexual	38 cigoto	48 descomponedor
9 excretor	19 magnetismo	29 aurícula	39 vivíparo	49 oxígeno
10 biotopo	20 fotosíntesis	30 citoplasma	40 refracción	50 electromagnetismo

Figura 6. Ejemplo de extracción terminológica con la función *keywords* para sexto de primaria en español.

Si calculamos el porcentaje de unidades que son adecuadas para incluirlas en nuestra base de datos siguiendo la fórmula anterior, obtenemos un 16,7 %. Todavía es un porcentaje bajo, pero comprobamos la tendencia a un incremento de la dificultad terminológica conforme pasan los cursos (Figura 6). Una vez hecho esto, se han acotado el número de palabras clave en todos los cursos para intentar recoger el mismo número de unidades y se han filtrado hasta alcanzar 120 para primero de primaria y 200 para el resto de los niveles, tanto en los corpus en inglés como en español.

Las unidades resultantes se han comparado en Excel. Para ello, primero se han vuelto a copiar todas ellas en una columna titulada *All corpus words* y se han eliminado duplicados a través de la opción *Herramientas de tabla > Quitar duplicados*. Las palabras resultantes se han dispuesto en una segunda columna titulada *Single words*. Más tarde, se ha empleado la función estadística *CONTAR.SI*, que cuenta el número de celdas, dentro del rango, que cumplen el criterio especificado. Con el objetivo de contar las frecuencias, la función quedaría formulada como sigue: `=CONTAR.SI(A$2:A$879;B2)`. Una vez aplicada a la celda C2, se arrastra con el cursor para aplicarla al resto de celdas de la columna titulada *Frequency*. De este modo, se ha extraído el número de veces que se repite cada palabra (Figura 7).

All corpus words	Single words	Frequency
inerte	ovíparo	6
ovíparo	vertebrado	6
vertebrado	invertebrado	6
esqueleto	carnívoro	6
olfato	vivíparo	6
invertebrado	anfibio	6
carnívoro	esqueleto	5
herbívoro	herbívoro	5
extremidades	escama	5
pulmones	máquina	5
vivíparo	oído	5
sentidos	electricidad	5
omnívoro	músculo	5
escama	hueso	5
tic	locomotor	5
tacto	circulatorio	5

Figura 7. Ejemplo de los resultados de la función estadística *CONTAR.SI* con las unidades monoléxicas en español.

Una vez hecho esto para las unidades monoléxicas en ambos idiomas, se ha establecido el límite de frecuencias en 3, ya que uno menor no sería suficientemente representativo. El número de unidades que se repiten 3 o más veces es el siguiente:

- Unidades monoléxicas en español: 95;
- Unidades monoléxicas en inglés: 94;
- Unidades poliléxicas en español: 27;
- Unidades poliléxicas en inglés: 25.

Ahora que conocemos estos datos, se ha creado otra hoja de cálculo donde se han comparado manualmente las unidades coincidentes en inglés y español (Figura 8).

Spanish words	English words	Bilingual matches
ovíparo	viviparous	ovíparo
vertebrado	oviparous	vertebrado
invertebrado	omnivore	invertebrado
carnívoro	vertebrate	carnívoro
vivíparo	invertebrate	vivíparo
anfibio	amphibian	anfibio
esqueleto	reptile	esqueleto
herbívoro	herbivore	herbívoro
escama	carnivore	máquina
máquina	pollute	oído
oído	intestine	electricidad
electricidad	skeleton	músculo

Figura 8. Comparación manual de las unidades monolingües coincidentes en inglés y en español con frecuencia superior o igual a 3.

5. Resultados

Al tener estos datos de los cuatro tipos de unidades, monoléxicas y poliléxicas en inglés y en español, se han comparado manualmente las coincidencias en ambos idiomas para

aquellas unidades que coincidían 3 o más veces. Los resultados del análisis indican que coinciden 57 unidades monoléxicas y 18 unidades poliléxicas, por lo tanto contamos con 150 unidades para poblar nuestra base de datos.

Para extraer el porcentaje de unidades coincidentes en ambos idiomas, se ha utilizado la siguiente fórmula:

$$\text{Porcentaje de coincidencias} = \frac{\text{Unidades coincidentes} * 100}{\text{Total de unidades}}$$

Los resultados son los siguientes: 60 % y 60,63 % para unidades monoléxicas en inglés y español, respectivamente. 66,66 % y 72 % en unidades poliléxicas en inglés y español, respectivamente.

Una vez extraído el porcentaje para cada idioma, se ha hecho la media de los porcentajes de coincidencias. El análisis demuestra que existe más de un 50 % de relación entre la frecuencia de unidades en el corpus en español y su equivalente en el corpus en inglés, en concreto: 60,31 % en unidades monoléxicas y 69,33 % en unidades poliléxicas.

La lista de términos seleccionados se acota, por tanto, a 150 unidades monoléxicas y poliléxicas, es decir 75 en español y 75 en inglés.

6. Creación de la base de datos y propuesta de entrada

Para crear la base de datos se ha utilizado un dominio público en la plataforma WordPress.com. En cuanto a la macroestructura de la base de datos, el menú principal se estructura en:

- Página de inicio: con el nombre de la base de datos y una interfaz de búsqueda para introducir el término. Más abajo se explican los motivos de creación, la manera de hacer las búsquedas, así como la información pertinente de cada una de las entradas.
- Acerca de: menú fijo donde se explica el motivo de creación del blog y los objetivos del trabajo.
- Guía de usuario: menú fijo donde se explica la información que contiene cada entrada y cómo han sido diseñadas.
- Glosario: contiene un desplegable según el idioma. Una vez se accede a él, se visualiza una lista ordenada alfabéticamente de todos los términos en inglés y en español
- Contacto: formulario de contacto para consultas o peticiones de los usuarios.

El diseño de las entradas, el cual está explicado tanto en la página de inicio como en el menú Guía de usuario, contará con los puntos propuestos por terminólogos y lingüistas. Entre ellos se enumeran, como mínimo, definición, ejemplo de uso y principales colocaciones (véase Buendía 2013; Buendía y Faber 2014; Buendía, Montero y Faber 2014; Buendía 2019, *inter alia*). La configuración final sería la siguiente:

1. Definición y ejemplo de uso: se extraerá la definición para cada una de las entradas y se acompañará de un ejemplo de uso. Con esto en mente, se empleará la funcionalidad *Concordance* y el parámetro avanzado Corpus Query Language (CQL), un tipo de código usado en Sketch Engine para hacer buscar patrones gramaticales o léxicos complejos o emplear criterios de búsqueda que no están disponibles desde la interfaz básica. Como base a las búsquedas en CQL se ha seguido la propuesta de León-Araúz y San Martín (2018).
2. Equivalente: en inglés, enlazado con su respectiva entrada.
3. Colocaciones y ejemplo de uso: priorizando las colocaciones verbales por frecuencia. Para ello, se ha utilizado la funcionalidad *Word Sketch*, que nos permite estudiar la categorías o categorías gramaticales de una palabra y sus relaciones sintácticas.
4. Imagen: con licencia Creative Commons.

Los corpus de referencia serán AN_ALL_ES y AN_ALL_EN para español e inglés, respectivamente (Figura 3). A continuación se propone el análisis lingüístico para el término *fotosíntesis* y su equivalente en inglés *photosynthesis*.

En primer lugar, se realiza una búsqueda en CQL con la siguiente ecuación: `[lemma="fotosíntesis"][]{0,5}[lemma="ser|generar|transformar|obtener|utilizar|mantener"] within <s/>`¹²



Figura 9. Ejemplo de búsqueda en CQL mediante la funcionalidad *Concordance* de Sketch Engine.

¹² De esta manera, acotamos la búsqueda a aquellos verbos especificados en cualquier de sus formas que estén hasta cinco *tokens* a la derecha de *fotosíntesis*.

Una vez comprobados los resultados, de las concordancias 4 y 6 (Figura 9) se extrae la definición y el ejemplo de uso: *La fotosíntesis es el proceso por el que la planta utiliza las sustancias que absorbe y la energía de la luz del sol para fabricar nutrientes. La fotosíntesis mantiene el nivel de dióxido de carbono en la atmósfera, ya que permite reducir el producido por las industrias, los motores de los vehículos y los animales.*

El equivalente es fácilmente extraíble: *photosynthesis*, como ya se había estudiado en la tabla de coincidencias anterior (Figura 7). Las colocaciones extraídas del *word sketch* (Figura 10) son las siguientes:

- fotosíntesis + hacer: *En las hojas, las plantas hacen la fotosíntesis; utilizan la energía del Sol para combinar el dióxido de carbono y el agua y fabricar sus propios alimentos.*
- fotosíntesis + producir: *La fotosíntesis produce oxígeno como desecho.*
- realizar + fotosíntesis: *Para fabricar sus alimentos, las plantas obtienen sustancias del suelo y del aire, y realizan la fotosíntesis.*

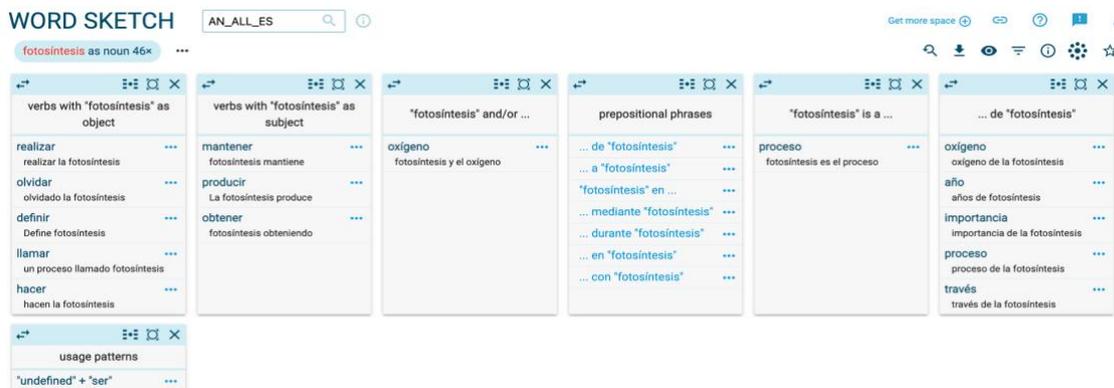


Figura 10. Ejemplo de la funcionalidad *Word Sketch* de *Sketch Engine* para *fotosíntesis*.

La imagen seleccionada cumple con la licencia Creative Commons para su distribución y es citada según las recomendaciones de la organización:

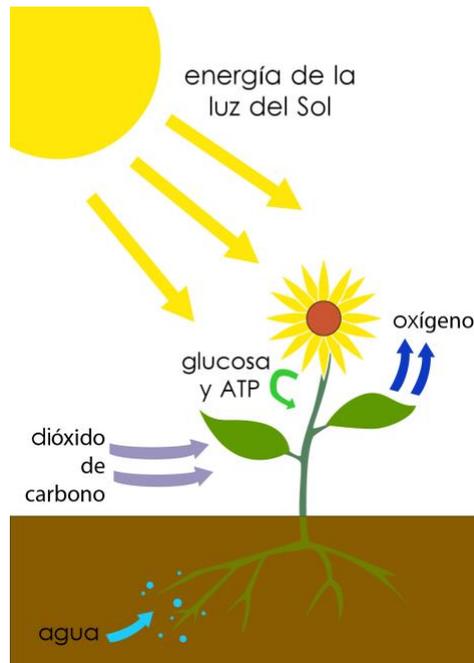


Figura 11. “Química de la fotosíntesis: Reactivos y productos” cuyo original es “The Scientific Process of Photosynthesis” de [At09kg](#) con licencia [CC BY-SA 3.0](#). Traducido y adaptado por [RoRo](#), con licencia [CC BY-SA 3.0](#).

Para el término en inglés, *photosynthesis*, la búsqueda en CQL quedaría como sigue: `[lemma="photosynthesis"][]{0,5}[lemma="be|perform|obtain|absorb|take"]` within <s/>. De esta manera, se acota la búsqueda a aquellos verbos especificados que se encuentren hasta cinco *tokens* después del lema.

Para la definición, se ha escogido la siguiente: *Plants take in water, minerals and carbon dioxide, and they perform photosynthesis to obtain nutrients and oxygen.* Su equivalente es *fotosíntesis* y las colocaciones verbales en inglés son:

- perform + photosynthesis: *Plants exhibit autotrophic nutrition; in other words, they produce nutrients from substances that they take from the environment, and perform photosynthesis.*
- photosynthesis + take: *The thin, flat green organs, joined to the stem, where photosynthesis takes place.*

En los casos en los que existe una adaptación traducida de la imagen original y es posible respetar el diseño, se ha optado por escoger la imagen más similar. De este modo, a los niños de Educación Primaria les resultará más fácil e intuitivo reconocer los elementos.

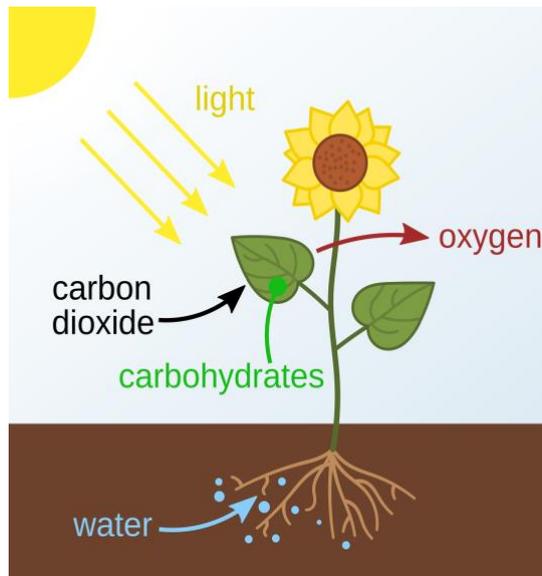


Figura 12. “[The overall process of photosynthesis](#)”, whose original is “[The Scientific Process of Photosynthesis](#)” by [At09kg](#), used under [CC BY-SA 3.0](#). This work has been translated and adapted by [Wattle](#) and [Nefronus](#), and it is licensed under [CC BY-SA 4.0](#).

Después de recoger todos estos datos, podríamos publicarlo en nuestra base de datos. El resultado final para la primera entrada sería el siguiente:

fotosíntesis

▲ cienciaseducacion ● 27 de mayo de 2023 ■ Sin categoría ✎ Editar

photosynthesis

«La fotosíntesis es el proceso por el que la planta utiliza las sustancias que absorbe y la energía de la luz del sol para fabricar nutrientes. La fotosíntesis mantiene el nivel de dióxido de carbono en la atmósfera, ya que permite reducir el producido por las industrias, los motores de los vehículos y los animales.»

Colocaciones:

- fotosíntesis + hacer: «En las hojas, las plantas hacen la fotosíntesis ; utilizan la energía del Sol para combinar el dióxido de carbono y el agua y fabricar sus propios alimentos.»
- fotosíntesis + producir: «La fotosíntesis produce oxígeno como desecho.»
- realizar + fotosíntesis: «Para fabricar sus alimentos, las plantas obtienen sustancias del suelo y del aire, y realizan la fotosíntesis.»



Figura 13. Ejemplo del término *fotosíntesis* en la base de datos.

En general, las entradas propuestas como en la Figura 13 están configuradas con definiciones y ejemplos de uso sencillos y colocaciones con frecuencias altas, lo que facilitará la mejor comprensión de los conceptos por parte del alumnado (véase Norman y Rumelhart 1975 y Rumelhart 1980). Cuando se haya poblado la base de datos con todas las unidades recogidas, se plantea pasar al plan personal de WordPress.com con el fin de abrir al público este recurso.

7. Conclusiones

Los objetivos principales de este Trabajo de Fin de Grado han sido la extracción y análisis terminológico de unidades monoléxicas y poliléxicas a partir de un corpus compilado en función de los subdominios de Ciencias de la Naturaleza de Educación Primaria y la creación de un recurso terminográfico bilingüe para el estudiantado de esta etapa educativa. Se ha constatado que, ante falta materiales con estas características, en la actualidad es necesaria la creación de recursos educativos abiertos para impulsar las competencias digitales y lingüísticas del alumnado en la modalidad de enseñanza bilingüe (español-inglés). Asimismo, es tarea de expertos, terminólogos y lingüistas computacionales desarrollar medios y permitir la digitalización en todos los niveles.

La extracción terminológica dio lugar a un corpus en español de 107 201 palabras y un corpus en inglés de 33 964. Tras la selección de 150 unidades, se creó una base de datos. Gracias a la representatividad y adecuación de los resultados obtenidos, esta base de datos se propone como una obra de referencia que conseguirá superar las lagunas de conocimiento y las limitaciones de la modalidad bilingüe en Educación Primaria.

Esta investigación se entiende como la antesala a un Trabajo de Fin de Máster e incluso una tesis doctoral en la que se aumente el número de palabras recogidas en el corpus, se ahonde en el aspecto computacional y se resuelvan dificultades de escalabilidad.

8. Bibliografía

- Agencia Europea del Medio Ambiente. (2022). *Glosario de medio ambiente*.
<https://www.greenfacts.org/glossary/index.htm>
- Alcaraz Varó, E. y Martínez Linares, M.^a A. (1997). *Diccionario de lingüística moderna*. Ariel.
- Ballester Carrillo, A. (2004). Corpus científico-técnico de la Real Academia Española. En González, L. y Hernández, P. (Coords.), *Las palabras del traductor: actas del II Congreso "El Español, Lengua de Traducción"*. (129-136). Bruselas: ESLEtRA. https://cvc.cervantes.es/lengua/esletra/pdf/02/016_ballester.pdf
- Bieber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257.
- Blackbox. (2023). *Blackbox Code Search Autocomplete Extraction*.
<https://chrome.google.com/webstore/detail/blackbox-code-search-auto/mcgbeeipkmelnpldkobichboakdfaeon>
- Bowker, L. (2002). *Computer-aided translation: a practical introduction*. University of Ottawa Press.
- Buendía Castro, M. (2013). *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources* [Tesis doctoral, Universidad de Granada]. <http://hdl.handle.net/10481/29527>
- Buendía Castro, M. (2020). Bilingual and multilingual online environmental knowledge resources: a comparative study for translation purposes. *International Journal of Lexicography*, 33(1), 40-72. doi.org/10.1093/ijl/ecz022
- Buendía Castro, M. y Faber, P. B. (2014). Collocation Dictionaries: A Comparative Analysis. *MonTI. Monografías de Traducción e Interpretación*, (6), 203-235. doi.org/10.6035/MonTI.2014.6.7
- Buendía Castro, M., Montero Martínez, S. y Faber, P. B. (2014). Verb collocations and phraseology in EcoLexicon. *Yearbook of Phraseology*, 5(1), 57-94. doi.org/10.1515/phras-2014-0004
- Buendía Castro, M., y Gómez-Moreno, J. M. U. (2010). ¿Cómo diseñar un corpus de calidad? Parámetros de evaluación. *Sendeban*, 21(1), 165-180. doi.org/10.30827/sendeban.v21i0.372

- Buendía Castro, M. (2021). *Verb collocations in dictionaries and corpus: an integrated approach for translation purposes*. Peter Lang GmbH, Internationaler Verlag der Wissenschaften.
- Cabré, M. T. (1993). *La terminología*. Antártida / Empúries.
- Cabré, M. T. (2009). La Teoría Comunicativa de la Terminología, una aproximación lingüística a los términos. *Revue Francaise De Linguistique Appliquee*, 14(2), 9-15. doi.org/10.3917/rfla.142.0009
- Cendrero, A., Zazo, C., Alonso, B., y Pérez-Estaún, A. (2009). *Vocabulario científico y técnico*. Real Academia de Ciencias Exactas, Físicas y Naturales.
- Estopà, R. (2023). *Mi primer diccionario de medicina ilustrado*. <https://www.diccionaridemedicina.app/web/es/>
- European Union. (s.f.). *IATE terminology database*. Accedido el 20 de mayo de 2023. <https://iate.europa.eu/entry/result/930970>
- Faber, P. B. (Ed.). (2012). *A cognitive linguistics view of terminology and specialized language*. De Gruyter Mouton.
- Faber, P., B., León-Araúz, P., Prieto Velasco, J. A., y Reimerink, A. (2007). Linking images and words: the description of specialized concepts. *International Journal of Lexicography*, 20(1), 39-65. doi.org/10.1093/ijl/eci038
- Faber, P., León-Araúz, P., y Prieto Velasco, J. A. (2009). Semantic relations, dynamicity, and terminological knowledge bases. *Current Issues in Language Studies*, 1(1), 1-23.
- FundéuRAE. (25 noviembre de 2022). «pélet», adaptación de «pellet». Recuperado el 27 de mayo de 2023. <https://www.fundeu.es/recomendacion/pelet-adaptacion-de-pellet/#:~:text=La%20forma%20p%C3%A9let%20es%20una,combustible%20para%20calderas%20o%20estufas.>
- Generalitat de Catalunya y Departament de Cultura. (2023). *TERMCAT, Centre de Terminologia*. <https://www.termcat.cat/es>
- Girko, B. (2023). *OCR – Image Reader*. <https://chrome.google.com/webstore/detail/ocr-image-reader/bhbhjjkcoghhibegcmbombkakkpdbo>
- Gómez de Enterría Sánchez, J. y Martí Sánchez, M. (2006). Estudio de los Lenguajes Especializados en Español. En Ahumada Lara, I. (Ed.), *Lenguas de especialidad y lenguajes documentales: VII Jornada de la Asociación Española de Terminología*. (67-80). Asociación Española de Terminología. <https://dialnet.unirioja.es/servlet/articulo?codigo=4170987>

- Gómez González-Jover, A. (2005) *Terminografía, lenguajes profesionales y mediación interlingüística: aplicación metodológica al léxico especializado de la industria del calzado y las industrias afines* [Tesis doctoral, Universidad de Alicante]. <http://hdl.handle.net/10045/760>
- Government of Canada. (2023). *TERMIUM Plus*. <https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng>
- Guías de la BUH. (2023). *Recursos Educativos Abiertos (REA): ¿QUÉ SON LOS RECURSOS EDUCATIVOS ABIERTOS?* <https://guiasbuh.uhu.es/RecursosEducativosAbiertos>
- Gutiérrez Rodilla, B. M. (1998). *La ciencia empieza en la palabra. Análisis e historia del lenguaje científico*. Península.
- Hurlé, A. D., Varó, E. A., y Motos, R. M. (2007). *Diccionario terminológico de las ciencias farmacéuticas*. Ariel.
- International Electrotechnical Commission. (2023). *Electropedia*. <https://www.electropedia.org/>
- International Standard Organization. (2019). *Terminology work and terminology science – Vocabulary* (ISO Standard n.º 1087:2019). <https://www.iso.org/standard/62330.html>
- Jakobson, R. (1963). *Essais de Linguistique*. Éditions de Minuit.
- Kilgarrif, A., Rychly, P., Smrz, P. y Tugwell, A. (2004). The Sketch Engine. *Information Technology Research Institute Technical Report Series, ITRI-04-08*. Universidad de Brighton.
- León-Araúz, P., y Martín, A. S. (2018). The EcoLexicon semantic sketch grammar: from knowledge patterns to word sketches. En Kernerman, I. y Krek, S. (Eds.) *Proceedings of the LREC 2018 Workshop «Globalex 2018 – Lexicography & WordNets»*. doi.org/10.48550/arXiv.1804.05294
- López Rodríguez, C. L. (2001). *Tipología textual y cohesión en la traducción biomédica inglés-español: un estudio de corpus* [Tesis doctoral, Universidad de Granada]. <http://hdl.handle.net/10481/27313>
- Martínez, S. M., y Benítez, P. F. (2009). Terminological competence in translation. Terminology. *International Journal of Theoretical and Applied Issues in Specialized Communication*, 15(1), 88-104. doi.org/10.1075/term.15.1.05mon
- Microsoft Corporation. (2018). *Microsoft Excel*. <https://office.microsoft.com/excel>

- Ministerio de Asuntos Económicos y Transformación Digital. (2015). *Plan de Impulso de las Tecnologías del Lenguaje*.
- Ministerio de Educación y Formación Profesional. (2022). *Plan de Digitalización y Competencias Digitales del Sistema Educativo*.
- Ministerio de Industria, Energía y Turismo y Ministerio de Hacienda y Administraciones Públicas. (2013). *Agenda Digital para España*.
- Montero-Martínez, S., y Benítez, P. F. (2011). Terminología para traductores e intérpretes. Tragacanto.
- Municio, Á. M., y Martínez, A. C. (2003). *Diccionario español de la energía: con vocabulario inglés-español*. Doce calles.
- Norman, D.A. y Rumelhart, D.E. (1975). *Explorations in cognition*. Freeman.
- Paepcke, F. (1975). Gemeinsprache, Fachsprachen und Übersetzung. En K. Berger y H. M. Sëller (eds.), *Übersetzen, Lebem*. Narr.
- Real Academia Española: Banco de datos. (2021). *Corpus de referencia del español actual (CREA)*. Recuperado el 25 de mayo de 2023. <https://apps2.rae.es/CREA/view/inicioExterno.view;jsessionid=209A57324BEAE2A5D07EC8210D451BDF>
- Real Academia Española: Banco de datos. (s.f.). *Corpus del Español del Siglo XXI (CORPES)*. Recuperado el 25 de mayo de 2023. <https://apps2.rae.es/CORPES/view/inicioExterno.view;jsessionid=7247D735C1E6BCFB0A1F24F5A8ECB15B>
- Real Academia Española: Banco de datos. (s.f.). *Corpus diacrónico del español (CORDE)*. Recuperado el 25 de mayo de 2023. <https://corpus.rae.es/cordenet.html>
- Real Academia Española. (2022). *Diccionario de la lengua española*. <https://dle.rae.es>
- Real Academia Española. (s.f.). Corpus. En *Diccionario de la lengua española*. Recuperado el 26 de mayo de 2023, de <https://dle.rae.es/corpus?m=form>
- Real Academia Española. (s.f.). *Diccionario panhispánico del español jurídico (DPEJ)*. Recuperado el 25 de mayo de 2023. <https://dpej.rae.es/>
- Real Academia Española. (s.f.). Resistencia. En *Diccionario de la lengua española*. Recuperado el 26 de mayo de 2023, de <https://dle.rae.es/resistencia?m=form>
- Real Academia Nacional de Medicina de España. (s.f.) *Diccionario de términos médicos*. Recuperado el 25 de mayo de 2023. <https://dtme.ranm.es/index.aspx>

Real Decreto 157 de 2022 [Ministerio de Educación y Formación Profesional]. Por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. 1 de marzo de 2022.

Rumelhart, D.E. (1980). Schemata: the building blocks of cognition. En R.J. Spiro et al. (Eds.), *Theoretical Issues in Reading Comprehension*. Lawrence Erlbaum.

Sager, J. C. (1990). *A practical course in terminology processing*. John Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

United Nations Department for General Assembly and Conference Management. (s.f.).

UNTerm. Recuperado el 25 de mayo de 2023. <https://unterm.un.org/unterm2/en/>

Van Rossum, G., y Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.