

Construcción de un diccionario combinatorio de inglés biomédico, *SciE-Lex*

Building up a combinatorial dictionary of biomedical English, *SciE-Lex*

ISABEL VERDAGUER CLAVERA
Universidad de Barcelona
i.verdaguer@ub.edu

NATALIA JUDITH LASO MARTÍN
Universidad de Barcelona
njlaso@ub.edu

Resumen: El objetivo de este artículo es presentar el proceso de gestación de *SciE-Lex*, un diccionario combinatorio de términos generales frecuentemente utilizados en inglés biomédico, que nació con el propósito de ayudar a los científicos españoles a redactar artículos científicos en inglés. Como proyecto lexicográfico se integra plenamente en la corriente léxico-gramatical que subraya la estrecha relación entre el léxico y la gramática. Su desarrollo ha venido marcado por tres etapas. En la primera, *SciE-Lex* proporciona, además del equivalente en español de cada entrada, información morfosintáctica y combinatoria, ilustrada con ejemplos y con notas de uso. En una segunda fase *añadimos* expresiones prefabricadas e información explícita acerca de su variabilidad, composición, funcionamiento y distribución en el artículo biomédico. Finalmente, en una tercera etapa, aun en desarrollo, nos proponemos introducir la semántica de marcos para establecer relaciones sintáctico-semánticas entre los distintos términos que pertenecen al mismo marco.

Palabras clave: diccionario combinatorio; inglés biomédico; estudios léxico-gramaticales; expresiones prefabricadas, semántica de marcos.

Abstract: The aim of this article is to present the development of *SciE-Lex*, a lexical database of non-specialised terms used in biomedical English. *SciE-Lex* has been designed with the purpose of helping Spanish scientists to write scientific articles in English. As a lexicographic project, it is fully integrated into the lexico-grammatical approach that underlines the close relationship between the lexicon and grammar. There have been three stages in its development: In the first stage, *SciE-Lex* provides, in addition to the Spanish equivalents of each entry, morphosyntactic and combinatorial information, illustrated with examples and usage notes. In a second phase, we added prefabricated expressions and explicit information about their variability, composition, discourse function and distribution in the biomedical research article. Finally, in a third stage we aim to introduce frame semantics in order to establish syntactic and semantic relations between the terms that belong to the same frame.

Key words: combinatorial dictionary; biomedical English; lexico-grammatical studies; prefabricated expressions; frame semantics.

Fecha de presentación: 24/02/2020 *Fecha de aceptación:* 23/11/2020

ISABEL VERDAGUER CLAVERA Y NATALIA JUDITH LASO MARTÍN,
«Construcción de un diccionario combinatorio de inglés biomédico, *SciE-Lex*», *Revista de
Lexicografía*, XXVI (2020), pp. 159-174
ISSN: 1134-4539, e-ISSN: 2603-667. doi: <https://doi.org/10.17979/rlex.2020.26.0.6049>

INTRODUCCIÓN

Con el propósito de ayudar a los científicos españoles a publicar en lengua inglesa en un mundo globalizado, en el que están obligados a dominar las particularidades del discurso biomédico en inglés, el grupo de investigación GreLic nos embarcamos en un proyecto lexicográfico para confeccionar un diccionario combinatorio del inglés biomédico, *SciE-Lex*. *SciE-Lex* empezó, pues, a gestarse teniendo en cuenta unos destinatarios específicos, la comunidad biomédica española, principalmente investigadores y profesionales del área de ciencias de la salud que tienen que redactar artículos científicos en inglés.

Al escribir un artículo académico es necesario expresar las ideas, la metodología y los resultados de una forma muy precisa y clara. Para ello, además de un buen conocimiento de la terminología especializada, hay que saber organizar correctamente el artículo, utilizar correctamente las estructuras sintácticas prototípicas del discurso biomédico, así como usar los patrones combinatorios de las palabras propios de cada registro. Si no se tienen esos conocimientos o se tienen de una forma parcial, es importante disponer de herramientas que proporcionen la información necesaria.

Tal y como han constatado Laso y John (2017: 150), la mayoría de bases de datos léxicas en formato electrónico se centran en inglés general, como *WordNet* (Miller *et al.*, 1990), que organiza información léxica en función de los diferentes sentidos de las palabras; *EuroWordNet* (Vossen, 2004), que proporciona un análisis semántico de las relaciones semánticas entre *synsets*; *Open Multilingual WordNet* (Bond, Vossen, McCrae y Fellbaum, 2016), que conecta *WordNets* de distintas lenguas; *SIMuLLDA* (Janssen, 2004), base de datos léxica multilingüe que utiliza «structured interlingua» (una «interlengua estructurada»); y bases de datos que proporcionan una descripción semántica de unidades léxicas basadas en marcos semánticos (*FrameNet*). No obstante, a pesar del creciente desarrollo de bases de datos que proporcionan información léxico-gramatical y discursiva de las lenguas, se requieren más recursos lingüísticos que den cuenta de las especificidades de los discursos de especialidad.

En este sentido, no hemos encontrado diccionarios biomédicos que ayuden a la producción escrita de forma global. Aunque en estas últimas décadas han salido a la luz diccionarios generales que proporcionan información sintáctica y combinatoria, la mayoría de los diccionarios científicos existentes incluyen solamente información enciclopédica y/o terminológica, y no incorporan información sobre la fraseología y las posibles combinaciones de palabras, esencial para la correcta construcción del discurso científico. Aunque existe una corriente que aboga por la inclusión de información lingüística en los diccionarios terminológicos (L'Homme, 2005, 2008, 2010, 2012, 2019; Faber, 2012, 2015; Faber y L'Homme, 2014) y defiende la combinación de información general y especializada (L'Homme, Subirats y Robichaud, 2016), hay muy pocos diccionarios especializados que proporcionen información semántica, sintáctica y colocal para una correcta construcción de la frase y, menos todavía, que tengan en

cuenta las características globales del discurso científico en inglés para una correcta construcción de todo el texto. Entre los recursos en los que se puede encontrar la información lingüística necesaria para poder solucionar problemas comunicativos cabe mencionar las siguientes bases de datos especializadas: *DiCoInfo* de términos informáticos (L'Homme, 2005, 2008), que se ha adaptado a un diccionario especializado de aprendices (Alipour, Robichaud, L'Homme, 2015); *DicoEnviro* (L'Homme, 2012); *EcoLexicon* (Reimerink y Faber, 2009), sobre términos medioambientales; o los diccionarios del proyecto *Diccionarios Valladolid-UVa* (Fuertes-Olivera, 2018; Fuertes-Olivera, Niño y Sastre, 2019). Estos recursos, sin embargo, o se centran en la terminología especializada y no en los términos generales usados en los textos científicos o no se centran en el inglés biomédico. Hay, pues, un vacío, que *SciE-Lex* se propone llenar.

Partiendo de las necesidades de los usuarios, tal como sugiere la teoría funcional de la lexicografía (Bergenholtz y Tarp, 2003; Fuertes-Olivera, 2012; Fuertes-Olivera y Tarp, 2014; Niño y Fuertes-Olivera, 2017), nuestro objetivo es proporcionar un recurso que solucione «the set of specific problems that pop up for a specific group of users with specific characteristics in specific users situations» (Bergenholtz y Tarp, 2003: 172). En nuestro caso, que proporcione toda clase de información —fonológica, morfológica, sintáctica, semántica y colocacional— sobre el uso de términos no especializados utilizados en el registro biomédico, información necesaria para satisfacer las necesidades del usuario en una situación comunicativa.

MARCO TEÓRICO: *SCIE-LEX* EN LA TRADICIÓN DE ESTUDIOS LÉXICO-GRAMATICALES

Debido a la creciente concienciación de la importancia de analizar el uso real de la lengua en diversas situaciones comunicativas, la lingüística de corpus ha tenido un gran impacto en la creación de obras lexicográficas (Renouf y Sinclair, 1991; Bogaards, 1996; Biber *et al.*, 1999; Cowie 1999; Moon, 2007a/b/c; Rundell y Fox, 2007; Granger y Paquot, 2012; Hanks, 2012; Domínguez Vázquez *et al.*, 2015). Según Moon (2007c), la creación de diccionarios basados en corpus sobresale como una de las aplicaciones más destacadas de la lingüística de corpus. Desde la primera edición del *Collins Cobuild English Dictionary (CCELD)*, publicado por primera vez en 1987, muchos proyectos lexicográficos se han basado en datos extraídos de un corpus textual. Tanto la investigación llevada a cabo desde el proyecto *Cobuild* (Sinclair *et al.*, 2003) como los estudios lexicográficos basados en corpus desarrollados en la Universidad de Birmingham han contribuido enormemente a la creación de obras lexicográficas muy relevantes (Carter, 1998; Cowie, 1999; Rundell y Fox, 2007; Granger y Paquot, 2012).

Sin duda, tanto el uso de datos de corpus como las herramientas de software empleadas para su análisis han acelerado el proceso de elaboración de una obra lexicográfica. En palabras de Meyer (2002: 16), «because so much text is now available in computer-readable form, many stages of dictionary creation can be automated». Los avances en la tecnología de corpus han sido un factor determinante en la proliferación de numerosos diccionarios basados en corpus, como por ejemplo el *Cambridge International Dictionary of English* (1995), el *Longman Dictionary of Contemporary English* (2005), el *Macmillan English Dictionary* (2007), o el *Collins Robert French-English/English-French Dictionary* (2010), que han confiado en los siguientes repositorios de datos basados en corpus: el

Cambridge International Corpus y el *Cambridge Learners' Corpus*, el *British National Corpus* (BNC), el *World English Corpus* y la base de datos Le Robert's, respectivamente.

En las últimas décadas, esta área de investigación de rápida expansión ha enfatizado la interacción entre el léxico y la gramática, así como la importancia de los patrones colocacionales y coligacionales, es decir, de combinaciones recurrentes de palabras y de categorías gramaticales, en la construcción del significado. Este enfoque léxico-gramatical aplicado al análisis lingüístico se desvía claramente de otras descripciones lingüísticas que tradicionalmente mostraban una clara separación entre los dominios lingüísticos ya mencionados. Por el contrario, tanto la lingüística teórica como la lingüística aplicada han centrado últimamente su atención en el hecho de que el léxico y la gramática no deben verse como niveles de análisis independientes el uno del otro. Dentro de este marco, la gramática se considera una «heavily constrained and abstract form of vocabulary rather than a separate linguistic level» (Gledhill, 2000: 74); es decir, tanto la gramática como el vocabulario se entienden como dos extremos del mismo espectro.

SciE-Lex, como proyecto lexicográfico, se integra plenamente en esta corriente léxico-gramatical que subraya la estrecha relación entre el léxico y la gramática y que parte de los estudios fundacionales de Hornby (1954), Halliday (1992) y cuyas raíces teóricas se retrotraen a las observaciones de Sinclair (1991) sobre el hecho de que las unidades léxicas suelen asociarse a fraseologías y significados típicos que a menudo se expresan mediante una variedad de patrones combinatorios. Esta concepción neo-firthiana del análisis de la lengua parte de la base, pues, de que la gramática y el léxico deben tratarse como dos niveles de descripción y análisis lingüísticos que son difícilmente separables.

Teniendo en cuenta que «any identification of a recurring pattern rests on observation» (Hunston, 2007: 259), el trabajo con corpus textuales se ha convertido en las últimas décadas en un lugar común para los estudios léxico-gramaticales, ya que la utilización de ejemplos reales de uso de la lengua proporciona descripciones de significado que muestran de los patrones léxico-gramaticales en los que aparece frecuentemente una determinada unidad léxica y que, desde esta perspectiva, dan cuenta de la creación de unidades extendidas de significado («extended units of meaning», Tognini-Bonelli, 2001).

No cabe duda de que una de las áreas en las que la evolución de la tecnología de corpus aplicada al análisis lingüístico ha tenido un mayor impacto es en el estudio del léxico. Singleton (2000) observa que el enfoque generativo evolucionó gradualmente desde principios de la década de 1950 hacia una nueva perspectiva lingüística que prioriza el estudio del léxico como principio fundamental del eje vertebrador del análisis lingüístico (Willis, 1990; Lewis, 1993; Hunston, 2002). Respecto a este punto, es importante señalar además que la mencionada reactivación del léxico está estrechamente relacionada con el advenimiento de la lingüística de corpus, la cual ha jugado un papel fundamental en el desarrollo de los marcos lingüísticos que ponen de relieve las relaciones sintagmáticas del léxico así como la noción de que la elección léxica es anterior a la selección de patrones gramaticales (Sinclair, 1991; Francis, 1993; Hunston y Francis, 2000):

Lexis is communicatively prior to syntax. As communicators we do not proceed by selecting syntactic structures and independently choosing lexical items to slot into them. Instead, we have concepts to convey and communicative choices to make which re-

quire central lexical items, and these choices find themselves syntactic structures in which they can be said comfortably and grammatically (Francis, 1993: 142).

Esta aproximación léxica a la descripción lingüística ha provocado una amplia gama de estudios que entienden el léxico como «systematically structured through patterns of use» (Hyland, 2008: 6). En opinión de Sinclair (1991), los significados no son selecciones independientes. Por el contrario, se considera que la estrecha interrelación entre elementos léxicos y su contexto de ocurrencia, a menudo denominada ‘colocación’, desempeña un papel importante en la interpretación de los textos. En este sentido, la construcción del diccionario combinatorio *SciE-Lex*, que describiremos en el siguiente apartado, se basa en la noción de ‘léxico-gramática’, la cual aborda el estudio de la lengua como un sistema unitario donde la gramática y el léxico son dos perspectivas complementarias del mismo continuo, en lugar de dos sistemas de análisis independientes.

DESCRIPCIÓN DE *SCIE-LEX*

Tres fases han marcado la gestación y posterior elaboración de *SciE-Lex*. En una primera fase, incluimos en el diccionario, además del equivalente en español de cada entrada, información morfosintáctica y combinatoria, ilustrada con ejemplos y con notas de uso. En una segunda fase, y en consonancia con las nuevas tendencias de estudios fraseológicos basados en corpus, añadimos expresiones prefabricadas (*lexical bundles*) e información explícita acerca de su variabilidad, composición, funcionamiento y distribución en el discurso biomédico. Finalmente, en una tercera etapa nos proponemos introducir la semántica de marcos para poder establecer relaciones entre los distintos términos que pertenecen al mismo marco.

Primera fase

El primer paso fue la compilación de un corpus, el *Health Science Corpus (HSC)*, que fuera representativo y reflejara el uso real de los recursos de la lengua en los textos científicos en el área de la biomedicina. Este corpus, del cual hemos extraído la información que incorporamos en el diccionario, consta aproximadamente de cuatro millones de palabras y se basa en artículos de reconocido prestigio de las áreas de la medicina, la biología, la bioquímica y la biomedicina. Una vez compilado el corpus, que incluye solo los textos propiamente dichos, utilizamos el programa de análisis textual *WordSmith Tools* para extraer la lista de palabras (para una descripción detallada de la metodología utilizada, véase Verdaguer *et al.*, 2013), a partir de los cuales seleccionamos los términos a incluir en el diccionario. Siendo conscientes de que ya hay en el mercado distintos diccionarios especializados que proporcionan el equivalente terminológico de los términos científicos y además la terminología específica en inglés no representa un problema para la comunidad biomédica española, ya que suelen utilizar la mayoría de esos términos en inglés (Laso, 2009), decidimos abordar las palabras generales usadas en el registro científico, ya que presentan más dificultades combinatorias (Laso, 2009) tanto a nivel sintáctico (construcción gramatical) como a nivel léxico (colocacional). A partir del análisis de las concordancias, *collocates* (palabras que se combinan frecuentemente) y *clusters* (grupos de palabras que aparecen recurrentemente y consisten en repeticiones exactas de la misma secuencia léxica) obtenidos con el programa, extrajimos toda clase de informa-

ción que ayudase a la producción de un texto científico, la cual fue almacenada en una base de datos y posteriormente volcada en el diccionario combinatorio *SciE-Lex*.

A continuación mostramos la información proporcionada en la primera fase de *SciE-Lex*, ejemplificada con la entrada *condition*

The screenshot shows the SciE-Lex interface. On the left is an alphabetical index. The main content area is titled 'Condition' and includes a search bar with 'lexical bundles'. Below this, it lists grammatical categories and their associated words and phrases:

- N** [condición] condition, conditions
- Adj ~** adverse ~, poor ~ | aerobic ~, anaerobic ~ | artificial ~ | certain ~, defined ~ | climactic ~, environmental ~ | different ~ | dry ~ | favourable ~, good ~, ideal ~, optimal ~ | identical ~, same ~, similar ~ | medical ~, physiological ~ | mild ~ | natural ~ | normal ~, standard ~ | neutral ~ | social ~, socioeconomic ~ | stringent ~
This hypothesis has been supported under certain conditions. | The highly variable results were attributed to the widely varying environmental conditions. | We assessed its export ability under different experimental conditions. | It survived in good condition for a considerable time. | This was followed by an extensive wash under stringent conditions.
- Prep ~** in ~, under ~
It is predicted by ecological theory under controlled conditions.
- N ~** assay ~ | baseline ~, control ~ | culture ~ | field ~, greenhouse ~, laboratory ~ | growth ~
performed in continued laboratory conditions | under restrictive growth conditions
- ~ Adj** ~ necessary, ~ required
We determined the conditions necessary for growth.

Figura 1. Entrada de condition en SciE-Lex

Como puede verse, a la izquierda aparece un listado de palabras incluidas en el diccionario, en este caso que empiezan por la letra c entre las que se encuentra *condition*. En la parte derecha vemos la entrada, con el lema destacado en negrita, seguido de su pronunciación en formato audio para ayudar a los usuarios en las presentaciones orales. A continuación, encontramos la categoría gramatical, nombre en este caso, su equivalente en español (*condición*) y su flexión morfológica. En este caso, el nombre es contable y por lo tanto proporcionamos información sobre su forma singular y plural. Además, aparecen los distintos patrones combinatorios, agrupados por categorías gramaticales (adjetivo, nombre y preposición y adverbio) y con la tilde en posición anterior o posterior reflejando la posición de la palabra *condition* en la combinación. Seguidamente aparecen las combinaciones léxicas más frecuentes, ordenadas alfabéticamente y según su campo semántico. En esta entrada vemos que los adjetivos que preceden *condition* más frecuentemente son *adverse ~, poor ~ | aerobic ~, anaerobic ~ | artificial ~ | certain ~, defined ~ | climactic ~, environmental...*, mientras *~ necessary, ~ required* siguen al nombre. Finalmente hay ejemplos que ilustran y completan la información proporcionada en la entrada. Estos ejemplos están inspirados en el corpus, aunque no son transcripciones exactas de las frases encontradas, que a menudo son de gran complejidad. En las palabras polisémicas hay también una aclaración del significado por medio de una glosa o términos sinónimos y en algunas entradas se proporcionan además notas aclaratorias para destacar usos especiales y/o para ayudar al usuario a utilizar correctamente el término de búsqueda.

Segunda fase

En la segunda fase ampliamos la información fraseológica, ya que el análisis del *Health Science Corpus* puso de manifiesto la tendencia fraseológica del lenguaje científico. No solo es en la lengua oral donde los hablantes utilizan frecuentemente combinaciones recurrentes de palabras, ya que disminuyen el tiempo de procesamiento del lenguaje (Nattinger y DeCarrico, 1992; Wray, 2002, 2008), sino también, como viene demostrando la investigación en lingüística de corpus (Sinclair, 1991, 2004; Stubbs, 2001, Biber y Barbieri, 2007; Römer y Schulze, 2009) están muy presentes en la lengua escrita. Los llamados *lexical bundles* (paquetes léxicos) (Biber *et al.*, 1999), es decir, secuencias de tres o cuatro palabras que aparecen repetidamente en el discurso analizado, son, según Biber, «bloques de construcción del discurso» y desempeñan funciones fundamentales. Además, en los lenguajes de especialidad, el conocimiento de la fraseología académica propia del registro es fundamental para determinar la pertenencia del autor a la comunidad científica, puesto que muestran su familiarización con las convenciones típicas del registro (Laso y John, 2013; Vincent, 2013; Carrió-Pastor, 2017; Navarro y Martínez, 2019). Así pues, decidimos ampliar *SciE-Lex* con la inclusión de *lexical bundles*, que contribuyen a la caracterización de la fraseología prototípica del discurso científico y a la organización estructural del mismo.

Consideramos que la información que debía incluirse en la segunda fase del diccionario era la siguiente:

1) Unidad fraseológica (*Lexical bundle*), conectada con las entradas del diccionario combinatorio (1ª fase de *SciE-Lex*) mediante hipervínculos. Estas unidades fraseológicas no son necesariamente estructuralmente completas, sino que tomamos la decisión de incluir secuencias que, sin constituir estructuras completas, como por ejemplo *it's noted that*, o *in agreement with*, son muy frecuentes.

2) Función discursiva. Con el objetivo de estudiar la función de las estructuras prefabricadas en el discurso biomédico, establecimos una taxonomía de funciones discursivas que nos ha permitido analizar de forma sistemática las diversas funciones de las expresiones identificadas como elementos organizadores del discurso, así como estudiar la interrelación y frecuencia entre las tipologías sintácticas y las discursivas. También analizamos qué expresiones fraseológicas presentan mayor variabilidad y qué tipo de variabilidad permiten, relacionando su variabilidad con su función discursiva y su distribución en el artículo biomédico. La lista de funciones del discurso se construyó teniendo en cuenta la clasificación funcional de Biber *et al.* (2004) y Hyland (2008), que clasifica las unidades fraseológicas en tres grupos: 1) las que describen el proceso de investigación; 2) las que organizan el discurso; y 3) las que establecen la posición del autor y su interacción con el lector. También se tuvieron en cuenta las funciones discursivas para aprendices de inglés (Evans, 1998) y el diccionario *Macmillan English for Advanced Learners*. Asimismo, incorporamos a nuestro análisis los «rhetorical moves» de Swales (1990, 2004).

3) Distribución en el discurso. Puesto que estas expresiones prefabricadas pueden ocupar distintas posiciones en el artículo biomédico (Hoey, 2005), es importante saber dónde aparecen, lo cual es especialmente relevante en las secuencias que tienen como función la estructuración del discurso. Para ello, dividimos el corpus inicial en cuatro

subcorpus, basándonos en las cuatro secciones típicas: (IMRD: Introducción, Método, Resultados, Discusión). Es importante recalcar, además, que una misma unidad fraseológica puede tener distintas funciones según su posición. Además, basándonos en razones pedagógicas, decidimos incluir un campo para ejemplos extraídos del corpus y otro para notas aclaratorias de utilidad para el usuario.

El aplicativo permite realizar distintos tipos de búsqueda, como por ejemplo, consultar el listado de funciones y ver todas las unidades fraseológicas asociadas a cada función. Igualmente, en cada entrada de unidad fraseológica se pueden ver las funciones que ésta desempeña. La figura 2 muestra la información sobre los *lexical bundles* asociados a *condition*: las propias expresiones, su función discursiva y distribución textual, así como ejemplos:

Figura 2. Lexical bundles asociados a la palabra *condition*

Como puede verse, el lema *condition* aparece en diversas (aunque parecidas) unidades fraseológicas (*under these conditions, under the conditions used...*) con la función (*Discourse Function*) de expresar condiciones y enmarcarlas, y se encuentra en la sección de resultados (*Text Distribution*). Después del ejemplo que ilustra su uso, tenemos una nota que puede ayudar al usuario aclarando que «under these conditions» normalmente aparece en posición inicial.

Por otra parte, en *SciE-Lex* se pueden buscar también los paquetes léxicos que expresen una determinada función. La figura 3 muestra a la izquierda la lista de funciones discursivas (*Acknowledging funding; Asking for information...*) y a la derecha los «lexical bundles» que se han encontrado en el corpus que expresan la función «Expressing conditions and framing»:

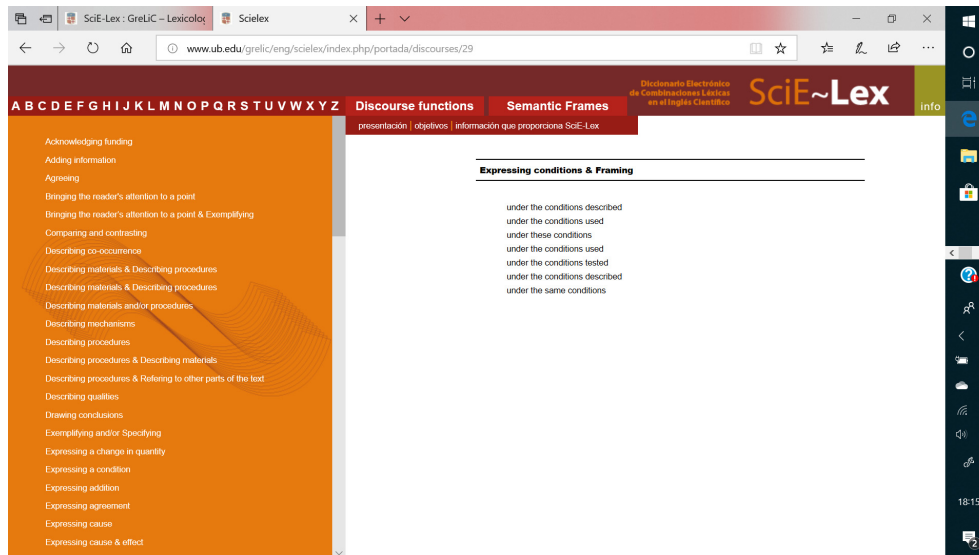


Figura 3. Lexical bundles asociados a la función «Expressing conditions and framing»

Tercera fase

Finalmente, nos propusimos poner de manifiesto las conexiones entre palabras que tienen patrones sintáctico-semánticos similares en el discurso biomédico en lengua inglesa. Aunque somos conscientes de que es más útil como herramienta metodológica para lingüistas y docentes de inglés para fines específicos que para los usuarios inicialmente previstos para el diccionario, creemos que estos también pueden beneficiarse del conocimiento de las redes semántico-sintácticas y colocacionales. Para ello hemos recurrido a *FrameNet*, una base de datos léxica en línea basada en la semántica de marcos, que vincula los significados de las palabras con sus marcos semánticos y sus contextos sintácticos, sistematizando las conexiones entre unidades léxicas relacionadas. *FrameNet* centra su atención en los marcos semánticos y permite conectar unidades léxicas que muestran propiedades sintácticas y semánticas similares; es decir, las unidades léxicas que pertenecen al mismo marco.

La idea básica de la semántica de marcos es que las palabras activan (o evocan) marcos en la mente de los hablantes (Fillmore, 1976, 1985; Fillmore *et al.*, 2003; Fillmore y Baker, 2010). La semántica de marcos parte de la base de que el significado de las unidades léxicas ha de describirse en relación con los marcos semánticos que evocan —es decir, representaciones esquemáticas de una situación concreta en la que están implicados varios participantes. Cada sentido de una palabra evoca un determinado marco semántico o estructura conceptual, incluyendo a los distintos participantes o elementos del marco. Así, por ejemplo, el verbo *concern*, al ser polisémico, evoca diversos marcos semánticos. En algunas oraciones, como «These questions concern rhetorical issues» evoca el marco *Topic*, definido en *FrameNet* como «A stretch of linguistic discourse or a Text that a Communicator produces has a topic that it is about». En otras, como «It concerns me that people are not getting enough help» evoca el marco *Cause_emotion*, definido como «An

Agent acts to cause an Experiencer a certain Emotion». Dentro de cada marco semántico se identifican sus participantes, algunos necesarios (*core*) aunque no necesariamente explícitos y otros periféricos u opcionales (*peripheral*), como modo, grado, tiempo o lugar. En el marco *Topic* «These questions» es el *Texto* y «rhetorical questions» es el *Topic*, que son elementos necesarios de este marco.

Como puede verse en estos ejemplos, *FrameNet* trata el lenguaje general. Su objetivo es identificar y definir todos los marcos posibles evocados por las unidades léxicas en un idioma, y anotar las oraciones extraídas de un corpus lingüístico para mostrar todas sus realizaciones sintáctico-semánticas. Nuestro objetivo ha sido aplicarlo al registro biomédico, conforme a la tendencia actual de aplicar la semántica de marcos a registros específicos: Dolbey (2009) lo ha aplicado a la biología molecular; Venturi (2013) al lenguaje legal; L'Homme (2008, 2010, 2014, 2019), L'Homme y Robichaud (2014), L'Homme, Subirats y Robichaud (2016) a la ecología y la informática; y Reimerink y Faber (2009), Faber y Buendía Castro (2014) y Faber, León-Araúz y Reimerink (2016) a la ecología.

En los registros específicos las palabras tienen sentidos más determinados que en el lenguaje general y son menos polisémicas. Así pues, el verbo *concern* en el registro biomédico se usa evocando el marco *Topic*, como es lógico en los registros científicos: «A key question concerns whether active genes are differentially localized». Nuestro objetivo ha sido identificar los diversos marcos a los que pertenecen los verbos tratados en *SciE-Lex* para que el usuario pueda tener una perspectiva onomasiológica y pueda agrupar las unidades léxicas por sus características semánticas. Así pues, además de *concern*, verbos como *address*, *discuss* y *refer* evocan el mismo marco.

La figura 4 muestra una pantalla en la que se ven verbos que empiezan con la letra c (en los que *concern* está incluido) con los marcos que evocan

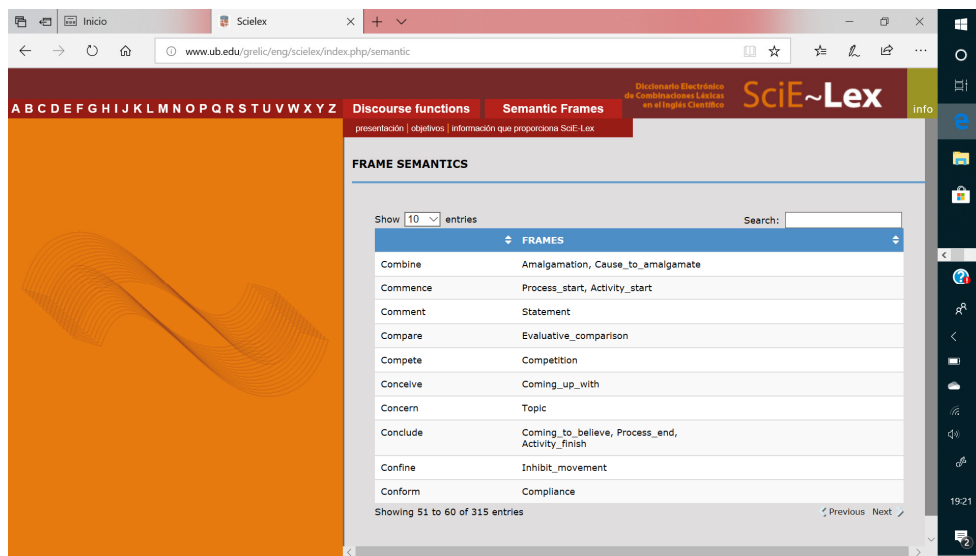


Figura 4. Verbos y marcos semánticos en SciE-Lex

La figura 5 muestra los verbos que pertenecen al marco semántico *Topic*

The screenshot shows the SciE-Lex website interface. The main content area is titled 'FRAME SEMANTICS' and features a search bar with the text 'Topic'. Below the search bar, there is a table listing verbs associated with the 'Topic' frame. The table has two columns: the first column lists the verbs, and the second column lists the associated frames. The verbs listed are 'Speak_on_topic', 'Resolve_problem', 'Concern', 'Discuss', and 'Refer'. The frames listed are 'Speak_on_topic, Resolve_problem', 'Topic', 'Discussion, Topic; Speak_on_a_topic', and 'Topic'. The table also indicates that there are 315 total entries, and the current view shows 1 to 4 of 4 entries.

Verb	Associated Frames
Address	Speak_on_topic, Resolve_problem
Concern	Topic
Discuss	Discussion, Topic; Speak_on_a_topic
Refer	Topic

Figura 5. Palabras asociadas al marco semántico *Topic* en *SciE-Lex*

El hecho de haber identificado los marcos semánticos asociados a las unidades léxicas incluidas en *SciE-Lex* permite al usuario tener fácil acceso a toda la información que proporciona *FrameNet* e identificar y agrupar los verbos que están asociados con el mismo marco. Además, su aplicación permite poner de manifiesto las conexiones semánticas, sintácticas y colocacionales de los verbos que pertenecen al mismo marco y ayudar a descubrir similitudes y diferencias sutiles entre palabras relacionadas, cuyo dominio contribuye a mejorar la producción escrita.

CONCLUSIONES

En conclusión, *SciE-Lex*, que fue inicialmente ideado para proporcionar información léxico-gramatical y colocacional del lenguaje general usado en el discurso biomédico en inglés, ha ido evolucionando con la introducción de *lexical bundles*, así como con información sobre su función y distribución en las distintas secciones del discurso. La evaluación que se ha llevado a cabo de *SciE-Lex* por parte de expertos, así como los distintos talleres que Laso y John han realizado con usuarios reales en la Universidad de Barcelona en los años 2014, 2018, 2019 y 2020 confirman la validez y utilidad de *SciE-Lex*. Finalmente, la introducción de redes semántico-sintácticas basadas en la semántica de marcos, aparte del interés teórico y metodológico para lingüistas y profesores de inglés para fines específicos, creemos que ayudará al usuario del diccionario a ser consciente de los patrones semánticos, sintácticos y combinatorios que los términos agrupados comparten, lo cual contribuirá a mejorar la escritura de sus textos profesionales.

REFERENCIAS BIBLIOGRÁFICAS

- ALIPOUR, Maryan, Benoit ROBICHAUD y Marie-Claude L'HOMME (2015): «Towards an Electronic Specialized Dictionary for Learners», en I. Kosem, M. Jakubíček, J. Kallas, y S. Krek, eds., *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*, Ljubljana-Brighton, Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd, pp. 51-69.
- ATKINS, Beryl T., Alain DUVAL y Rosemary C. MILNE (2010^o): *Collins Robert French-English/English-French Dictionary*, Glasgow-London, Harper Collins.
- BERGENHOLTZ, Henning y Sven TARP (2003): «Two opposing theories: On H. E. Wiegand's recent discovery of lexicographic functions», *Hermes. Journal of Linguistics*, 31, pp. 171-196.
- BIBER, Douglas, Stig JOHANSSON, Geoffrey LEECH, Susan CONRAD y Edward FINEGAN (1999): *Longman Grammar of Spoken and Written English*, Harlow, Pearson.
- BIBER, Douglas y Federica BARBIERI (2007): «Lexical bundles in university spoken and written registers», *English for Specific Purposes*, 26, 3, pp. 263-286.
- BIBER, Douglas, Susan CONRAD y Vivianne CORTÉS (2004): «If you look at...: Lexical bundles in university teaching and textbooks», *Applied Linguistics*, 25, 3, pp. 371-405.
- BOGAARDS, Paul (1996): «Dictionaries for learners of English», *International Journal of Lexicography*, 9, 4, pp.277-320.
- BOND, Francis, Piek VOSSEN, John P. MCCRAE y Christiane FELLBAUM (2016): «CILI: the collaborative interlingual index», en V. Barbu Mititelu, C. Forăscu, P. Vossen, C. Fellbaum, eds., *Proceedings of the 8th Global WordNet Conference*.
- CARRIÓ-PASTOR, M. Luisa (2017): «Verbal phraseology: An analysis of cognitive verbs in linguistics, engineering and medicine academic papers», en R. Mitkov, ed., *Cognitive and Cultural Aspects of Phraseology*, Springer International Publishing, pp. 325-336.
- CARTER, Ronald (1998²): *Vocabulary: Applied Linguistic Perspectives*, London, Routledge.
- COWIE, Anthony P. (1999): *English Dictionaries for Foreign Learners*, Oxford, Oxford University Press.
- DOLBEY, Andrew (2009): *BioFramenet*, Berkeley, Department of Linguistics, University of California Berkeley.
- DOMÍNGUEZ VÁZQUEZ, M. José Xavier GÓMEZ GUINOVARTE y Carlos VALCÁRCEL RIVEIRO (2015): *Lexicografía de las lenguas románicas: Aproximaciones a la lexicografía moderna y contrastiva*, Berlin-Boston, Mouton de Gruyter.
- EVANS, Virginia (1998): *Successful Writing Proficiency*, Tübingen, Gunter Narr.
- FABER, Pamela, ed. (2012): *A Cognitive Linguistics View of Terminology and Specialized Language*, Berlin-Boston, Mouton de Gruyter.
- FABER, Pamela (2015): «Frames as a framework for terminology», en H. J. Kockaert y F. Steurs, eds., *Handbook of Terminology*, Amsterdam, John Benjamins, pp. 14-33.
- FABER, Pamela y Miriam BUENDÍA CASTRO (2014): «EcoLexicon», en C. V. Andrea Abel y N. Ralli, eds., *Proceedings of the XVI EURALEX International Congress*, Bolzano, Institute for Specialised Communication, pp. 601-607.
- FABER Pamela y Marie-Claude L'HOMME (2014): «Lexical Semantic Approaches to Terminology: An Introduction», *Terminology*, 21, 2, pp. 143-150.

- FABER, Pamela, Pilar LEÓN-ARAÚZ, y Arianne REIMERINK (2016): «EcoLexicon: new features and challenges», en I. Kernerman, I. Kosem Trojina, S. Krek, y L. Trap-Jensen, eds., *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, Portorož, pp.73-80.
- FILLMORE, Charles. J. (1976): «Frame semantics and the nature of language», *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280, pp. 20-32.
- FILLMORE, Charles. J. (1985): «Frames and the semantics of understanding», *Quaderni di Semantica*, 6, pp. 222-254.
- FILLMORE, Charles. J., Christopher JOHNSON y Miriam R. PETRUCK (2003): «Background to FrameNet», *International Journal of Lexicography*, 16, 3, pp. 235-250.
- FILLMORE, Charles. J. y Collin BAKER (2010): «A frames approach to semantic analysis», en B. Heine y H. Narrog, eds., *The Oxford Handbook of Linguistic Analysis*, Oxford, Oxford University Press, pp. 313-339.
- FRANCIS, Gill (1993): «A Corpus-Driven Approach to Grammar. Principles, Methods and Examples», en M. Baker, G. Francis, y E. Tognini-Bonelli, eds., *Text and Technology. In Honour of John Sinclair*, Amsterdam, John Benjamins, pp. 137-156.
- FUERTES-OLIVERA, Pedro (2012): «La elaboración de diccionarios especializados para el traductor: Teoría y práctica», *Revista de lexicografía*, XVIII, pp. 7-23.
- FUERTES-OLIVERA, Pedro, ed. (2018): *The Routledge Handbook of Lexicography*, London-New York, Routledge.
- FUERTES-OLIVERA, Pedro y Sven TARP (2014): *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*, Berlin, Mouton de Gruyter
- FUERTES-OLIVERA, Pedro, Marta NIÑO y Ángeles SASTRE (2019): «Tecnología con fines lexicográficos: su aplicación en los Diccionarios Valladolid-UVa», *Revista Internacional de Lenguas Extranjeras*, Monográfico, 10, pp. 75-100.
- GLEDHILL, Christopher (2000): *Collocations in science writing*, Tübingen, Gunter Narr.
- GRANGER, Sylviane y Magali PAQUOT (2012): *Electronic Lexicography*, Oxford, Oxford University Press.
- HALLIDAY, Michael A. K. (1992): «Language as system and language as instance: the corpus as a theoretical construct», en J. Svartvik, ed., *Directions in Corpus Linguistics*, Berlin, Mouton, pp. 61-77.
- HANKS, Patrick (2012): «The corpus revolution in lexicography», *International Journal of Lexicography*, 25, 4, pp. 398-436.
- HOEY, Michael (2005): *Lexical priming: A new theory of words and language*, London, Routledge.
- HORNBY, Albert S. (1954): *A Guide to Patterns and Usage in English*, London, Oxford University Press.
- HUNSTON, Susan (2007): «Semantic prosody revisited», *International Journal of Corpus Linguistics*, 12, 2, pp. 249-268.
- HUNSTON, Susan (2002): «Pattern grammar, language teaching, and linguistic variation. Applications of a corpus-driven grammar», en R. Reppen, S. M. Fitzmaurice y D.Biber, eds., *Using Corpora to Explore Linguistic Variation*, Amsterdam, John Benjamins, pp.167-183.
- HUNSTON, Susan y Gill FRANCIS (2000): *Pattern Grammar*, Amsterdam, John Benjamins.

- HYLAND, Ken (2008): «As can be seen: Lexical bundles and disciplinary variation», *English for Specific Purposes*, 27, pp. 4-21.
- JANSSEN, Maarten (2004): «Multilingual lexical databases, lexical gaps and SIMuLLDA», *International Journal of Lexicography*, 17, 2, pp. 137-154.
- LASO, Natalia Judith (2009): *A Corpus-Based Study of the Phraseological Behaviour of Abstract Nouns in Medical English: A Needs Analysis of a Spanish Medical Community*. Ph.D. dissertation, University of Barcelona (accesible en: http://www.tdx.cat/bitstream/handle/10803/1671/NJLM_THESIS.pdf;jsessionid=2B6BE5199B0F415FD7380A1FF2495AFC.tdx2?sequence=1).
- LASO, Natalia Judith y Suganthi JOHN (2013): «An exploratory study of NNS medical writers' awareness of the collocational patterning of abstract nouns in medical discourse», *Revista Española de Lingüística Aplicada*, 26, pp. 307-331.
- LASO, Natalia Judith y Suganthi JOHN (2017): «The pedagogical benefits of a lexical database (*SciE-Lex*) to assist the production of publishable biomedical texts by EAL writers», *Iberica*, 33, pp. 147-172.
- LEWIS, Michael (1993): *The Lexical Approach: The State of ELT and a Way Forward*, Hove, LTP.
- L'HOMME, Marie-Claude (2005): «Conception d'un dictionnaire fondamental de l'informatique et de l'Internet: sélection des entrées», *Le langage et l'homme*, 40, 1, pp. 137-154.
- L'HOMME, Marie-Claude (2008): «Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés», *Traduire*, 217, pp. 78-103.
- L'HOMME, Marie-Claude (2010): «Designing Terminological Dictionaries for Learners based on Lexical Semantics: The representation of actants», en P. Fuertes-Olivera, ed., *Specialised Dictionaries for Learners*, Berlin, Mouton de Gruyter, pp. 141-153.
- L'HOMME, Marie-Claude (2012): «Compiling specialized dictionaries differently: a brief overview of terminological projects at the Observatoire de linguistique Sens-Texte (OSLT)», *K Dictionary News*, 20, pp. 2-7.
- L'HOMME, Marie-Claude (2014): «Why Lexical Semantics is important for e-lexicography», *International Journal of Lexicography*, 27, 4, pp. 360-377.
- L'HOMME, Marie-Claude (2019): *Lexical Semantics for Terminology*, Amsterdam, John Benjamins.
- L'HOMME, Marie-Claude y Benoît ROBICHAUD (2014): «Frames and terminology: representing predicative units in the field of the environment», en Michael Zock, Reinhard Rapp y Chu-Ren Huang, eds., *Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex 2014)*, Dublin, Association for Computational Linguistics and Dublin City University, pp. 186-197.
- L'HOMME, Marie-Claude, Carlos SUBIRATS y Benoît ROBICHAUD (2016): «A proposal for combining "general" and specialized frames», en *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex V)*, Osaka, pp. 156-165.
- Longman Dictionary of Contemporary English* (2005): Essex, Pearson Education.
- MEYER, Charles (2002): *English Corpus Linguistics: An Introduction*, Cambridge, Cambridge University Press.
- MILLER George A., Richard BECKWITH, Christiane FELLBAUM, Derek GROSS, Katherine J. MILLER (1990): «Introduction to WordNet: An On-line Lexical Database», *International Journal of Lexicography*, 3, 4, pp. 235-244.

- MOON, Rosamund (2007a): «Phraseology in general monolingual dictionaries», en H. Burger, D. Dobrovól'skij, P. Kühn y N.R. Norrick, *An International Handbook of Contemporary Research*, Berlin-New York, Walter de Gruyter, pp. 909-918.
- MOON, Rosamund (2007b): «Corpus linguistic approaches with English corpora», en H. Burger, D. Dobrovól'skij, P. Kühn y N.R. Norrick, *An International Handbook of Contemporary Research*, Berlin-New York, Walter de Gruyter, pp. 1045-1059.
- MOON, Rosamund (2007c): «Sinclair, lexicography, and the Cobuild Project. The application of theory», *International Journal of Corpus Linguistics*, 12, 2, pp. 159-181.
- NATTINGER, James R. y Jeanette DE CARRICO (1992): *Lexical Phrases and Language Teaching*, Oxford, Oxford University Press.
- NIÑO AMO, Marta y Pedro FUERTES-OLIVERA (2017): «La Introducción Sistemática en el Diccionario Especializado», *Revista de Lexicografía*, XXIII, pp. 137-156
- NAVARRO GIL, Noelia y Elena MARTÍNEZ CARO (2019): «Lexical bundles in learner and expert academic writing», *Bellaterra Journal of Teaching & Learning Language & Literature*, 12, 1, pp. 65-90.
- PROCTER, Paul (1995): *Cambridge International Dictionary of English*, Cambridge, Cambridge University Press.
- REIMERINK, Arianne y Pamela FABER (2009): «Ecolexicon: A frame-based knowledge base for the environment», en J. Hřebíček, J. H. Mirovsky, W. Pillmann, I. Holoubek y T. Bandholtz, eds., *European conference of the Czech Presidency of the Council of the EU TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*, Brno, Masaryk University, pp. 25-27.
- RENOUF, Antoinette y John SINCLAIR (1991): «Collocational frameworks in English», en K. Aijmer y B. Altenberg, *English Corpus Linguistics*, London, Longman, pp. 128-144.
- RUNDELL, Michael y Gwineth FOX (2007²): *Macmillan English Dictionary. For Advanced Learners*, Oxford, Macmillan.
- RÖMER, Ute y Rainer SCHULZE (2009): *Exploring the Lexis-Grammar Interface*, Amsterdam, John Benjamins.
- SINCLAIR, John (1991): *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- SINCLAIR, John, ed. (2003⁴): *Collins COBUILD English Dictionary*, Glasgow, Harper Collins Publishers.
- SINCLAIR, John (2004): *Trust the Text: Language, Corpus and Discourse*, London, Routledge.
- SINGLETON, DAVID (2000): *Language and the Lexicon. An Introduction*, London, Arnold.
- STUBBS, Michael (2001): *Words and Phrases: Corpus Studies of Lexical Semantics*, Oxford, Blackwell.
- SWALES, John M. (1990): *Genre Analysis: English in Academia and Research Settings*, Cambridge, Cambridge University Press.
- SWALES, John M. (2004): *Research Genres: Explorations and Applications*, Cambridge, Cambridge University Press.
- TOGNINI-BONELLI, Elena (2001): *Corpus linguistics at work*, Amsterdam, John Benjamins.
- VENTURI, Giulia (2013): «A semantic annotation of Italian legal texts. A FrameNet-based approach», en M. Fried y K. Nikiforidou, eds., *Advances in Frame Semantics*, Amsterdam, John Benjamins, pp. 51-84.

- VERDAGUER, Isabel, Natalia Judith LASO y Danica SALAZAR, eds. (2013): *Biomedical English*. Amsterdam-Philadelphia, John Benjamins.
- VINCENT, Benet (2013): «Investigating academic phraseology through combination of very frequent words: a methodological exploration», *Journal of English for Specific Purposes*, 12, pp. 44-56.
- VOSSEN, Piek (2004): «EuroWordNet: A Multilingual Database of Autonomous and Language-specific Wordnets Connected via an Inter-Lingual-Index», *International Journal of Lexicography*, 17, 2, pp. 161-173.
- WILLIS, Dave (1990): *The Lexical Syllabus. A New Approach to Language Teaching*, London-Glasgow, Collins.
- WRAY, Alison (2002): *Formulaic Language and the lexicon*, Cambridge, Cambridge University Press.
- WRAY, Alison (2008): *Formulaic Language: Pushing the Boundaries*, Oxford, Oxford University Press.