



RILEX
REVISTA SOBRE INVESTIGACIONES LÉXICAS

VOLUMEN VIII - NÚMERO 1
ENERO, 2025

Gerda Haßler
Carlos Ynduráin Pardo de Santayana
Marco Antonio Pérez Durán
Ma. Guadalupe Rojas Corona
Elodie Weber
Filip Zeman
María Victoria Cantero Romero

REVISTAS CIENTÍFICAS DE LA UNIVERSIDAD DE JAÉN

<https://doi.org/10.17561/rilex.8.1>

Los estudios e investigaciones que se recogen en esta revista están sujetos a una licencia de reconocimiento de *Creative Commons*. Esta licencia permite **compartir** (copiar y redistribuir el material en cualquier medio o formato) y **adaptar** (remezclar, transformar y construir a partir del material para cualquier propósito, incluso comercialmente) el material siempre que se indique adecuadamente el origen y los cambios.

CONSEJO EDITORIAL

EDITORA

Dr.^a M.^a Águeda Moreno Moreno (Universidad de Jaén)

DIRECTOR EDITORIAL

Dr. Jesús Camacho Niño (Universidad de Jaén)

SECRETARÍA

Dr.^a Marta Torres Martínez (Universidad de Jaén)

CONSEJO DE REDACCIÓN

DIRECCIÓN

Dr.^a M.^a Águeda Moreno Moreno (Universidad de Jaén)

SUBDIRECCIÓN/SECRETARÍA

Dr.^a Marta Torres Martínez (Universidad de Jaén)

VOCALES

Dr.^a Eleni Leontaridi (Aristotle University of Thessaloniki)

Dr.^a Elisabeth Fernández Martín (Universidad de Almería)

Dr. Francisco Pedro Pla Colomer (Universidad de Jaén)

Dr. Jesús Camacho Niño (Universidad de Jaén)

Dr. Matías Hidalgo Gallardo (Università degli Studi di Bergamo)

Dr. Narciso Contreras Izquierdo (Universidad de Jaén)

Dr. Tibor Berta (Universidad de Szeged)

Dr.^a Victoria Rodrigo (Georgia State University)

EQUIPO TÉCNICO

EDITOR TÉCNICO

Dr. Jesús Camacho Niño

ASISTENCIA TÉCNICA

Alicia Arjonilla Sampedro (Universidad de Jaén)

Inmaculada Ruiz Sánchez (Universidad de Jaén)

COMITÉ CIENTÍFICO

Ángel López García-Molins, Universidad de Valencia, España

Cecilio Garriga Escribano, Universidad Autónoma de Barcelona, España

Concepción Maldonado González, Universidad Complutense de Madrid, España

Dolores Azorín Fernández, Universidad de Alicante, España

Giuseppe Trovato, Universidad de Venecia, Italia

Gloria Clavería Nadal, Universidad Autónoma de Barcelona, España

Humberto Hernández Hernández, Universidad de La Laguna, España

Josefina Prado Aragonés, Universidad de Huelva, España

José Ignacio Pérez Pascual, Universidad de A Coruña, España

José Ramón Carriazo Ruiz, Universidad Nacional del Educación a Distancia, España

Mar Campos Souto, Universidad de Santiago de Compostela, España

Mar Cruz Piñol, Universidad de Barcelona, España

M.^a Luisa Calero Vaquera, Universidad de Córdoba, España

Marta Higuera García, Instituto Cervantes, España

Matteo de Beni, Universidad de Verona, Italia

Pedro Fuertes-Olivera, Universidad de Valladolid, España

Stefan Ruhstaller, Universidad Pablo de Olavide, España

Sven Tarp, Universidad de Aarhus, Dinamarca

ÍNDICE

Gerda Haßler

¿El principio de ordenación onomasiológica como perturbación del orden?

Caminos de la lexicografía a la historiografía de la lingüística.....7

Carlos Ynduráin Pardo de Santayana

Aproximación a los sentidos metafóricos de *grande* y *pequeño*: propuesta lexicográfica.....37

Marco Antonio Pérez Durán y Ma. Guadalupe Rojas Corona

Análisis de las frecuencias léxicas nominales, adjetivales y adverbiales presentes en escritos de estudiantes universitarios de la Clínica de Escritura de la Facultad de Psicología de la UASLP.....69

Elodie Weber

La variación diatópica del léxico: el caso de *agregar* y *añadir* en España

e Hispanoamérica.....93

Filip Zeman

Uso de corpus propios en la enseñanza del léxico formulaico en el aula de Español

como Lengua Extranjera.....123

María Victoria Cantero Romero

El léxico ELE en los modelos de lenguaje.....155



¿EL PRINCIPIO DE ORDENACIÓN ONOMASIOLÓGICA COMO PERTURBACIÓN DEL ORDEN? CAMINOS DE LA LEXICOGRAFÍA A LA HISTORIOGRAFÍA DE LA LINGÜÍSTICA

THE ONOMASIOLOGICAL ARRANGEMENT PRINCIPLE AS A DISTURBANCE OF ORDER? PATHS FROM LEXICOGRAPHY TO THE HISTORIOGRAPHY OF LINGUISTICS

Gerda Haßler
Universität Potsdam
gerda.hassler@uni-potsdam.de

RESUMEN

Esta contribución pretende ser un alegato a favor de la aplicación del método onomasiológico en la historiografía de la lingüística. Dado que en aquellas épocas aún no se habían consolidado los conceptos lingüísticos se produjeron desarrollos conceptuales en los textos que dieron lugar a variantes, una ordenación alfabética, como la habitual en los diccionarios lingüísticos, resulta poco adecuada. Se presentará un primer resultado de la aplicación del método onomasiológico a la historia de la lingüística con el lexicon de los fundamentos teórico-lingüísticos de los siglos XVII y XVIII (Haßler & Neis, 2009). En esta contribución, explicaremos el enfoque metodológico de los estudios onomasiológicos sobre la historia de los conceptos lingüísticos, utilizando ejemplos que ilustran la aplicabilidad del método onomasiológico en futuros proyectos. Dada la enorme variedad de términos y denominaciones, no podemos limitarnos a buscar una palabra o un grupo de palabras en los textos. Existe una asimetría entre la vertiente conceptual y la denominativa que hace imposible establecer correspondencias directas.

Palabras clave: método onomasiológico, conceptos metalingüísticos, diccionarios de términos lingüísticos, arbitrariedad del signo.

ABSTRACT

This contribution is intended as a plea for the application of the onomasiological method in the historiography of linguistics. For periods in which the consolidation of linguistic concepts had not yet been completed and in which conceptual developments occurred in the texts that gave rise to variants, an alphabetical arrangement, as is usual in linguistic dictionaries, is not very appropriate. A first result of the application of the onomasiological method to the history of linguistics will be presented with the lexicon of the linguistic-theoretical foundations of the 17th and 18th centuries (Haßler & Neis, 2009). In this contribution, we will explain the methodological approach of onomasiological studies on the history of linguistic concepts, using examples that illustrate the applicability of the onomasiological method in future projects. Given the enormous variety of terms and designations, we cannot limit ourselves to looking for a word or a group of words in texts. There is an asymmetry between the conceptual and denominational sides that makes it impossible to establish direct correspondences.

Keywords: onomasiological method, metalinguistic concepts, dictionaries of linguistic terms, arbitrariness of the sign.

Recibido: 02-09-2024
Aceptado: 16-10-2024

DOI: <https://doi.org/10.17561/rilex.8.1.9160>



1. INTRODUCCIÓN

Esta contribución pretende ser un alegato a favor de la aplicación de cuestiones onomasiológicas en la historiografía de la lingüística, especialmente, para las épocas en las que la consolidación de los conceptos lingüísticos aún no había concluido y en las que se produjeron desarrollos conceptuales en los textos que dieron lugar a variantes. Por tanto, se trata de una contribución metodológica.

En primer lugar, se ilustrará la inadecuación del principio de ordenación utilizado en los diccionarios de términos lingüísticos que hacen referencia a contextos historiográficos. A continuación, abordaremos la aparición del principio onomasiológico que ha alterado el orden alfabético. Este principio onomasiológico también se ha aplicado en el léxico de términos lingüísticos básicos de los siglos XVII y XVIII (Haßler & Neis, 2009), cuya estructura y principios se presentarán brevemente. Finalmente, mediante los ejemplos de las denominaciones de tiempos verbales en las primeras gramáticas del español y en un manual de lenguas del siglo XVII, se mostrarán las ventajas del principio onomasiológico.

2. INADECUACIÓN DEL PRINCIPIO SEMASIOLÓGICO EN LOS CONTEXTOS HISTORIOGRÁFICOS

Una forma de presentación de los conocimientos conceptuales lingüísticos son los diccionarios de términos lingüísticos que también se han publicado en español, sobre todo, a partir de los años ochenta, y que se utilizan en la enseñanza en las universidades. Algunos ejemplos son los siguientes:

Abad Nebot, Francisco (1986). *Diccionario de lingüística de la escuela española*. Gredos.

Abraham, Werner (1981). *Diccionario de terminología lingüística actual*. Gredos.

Alcaraz Varó, Enrique (2004). *Diccionario de lingüística moderna*. Ariel.

Benito Lobo, José A. (1982). *Lengua española: glosario de términos lingüísticos*. Servicio de Publicaciones del Ministerio de Educación y Ciencia (ediciones 1983, 1986, 1987, 1990, 1991, 1992, 1993).

Cardona, Giorgio Raimondo (1991). *Diccionario de lingüística*. Edición española a cargo de María Teresa Cabello. Ariel.

Cerdà Massó, Ramón (coordinador) (1986). *Diccionario de lingüística*. Anaya.

Crystal, David (2000). *Diccionario de lingüística y fonética*. Traducción y adaptación de Xavier Villalba. Octaedro.

Dubois, Jean (1979). *Diccionario de lingüística*; versión... [del francés] de Inés Ortega y Antonio Domínguez; dirección y adaptación de Alicia Yllera. Alianza Editorial, D.L.

Instituto Nacional de Bachillerato a Distancia (1984). *Glosario de términos lingüísticos*. Servicio de Publicaciones del Ministerio de Educación y Ciencia.

Lewandowski, Theodor (1982). *Diccionario de lingüística*. Traductores, María Luz García-Denche Navarro, Enrique Bernárdez. Cátedra.

Mounin, Georges (1982). *Diccionario de lingüística*, dirigido por Georges Mounin, traductor Ricardo Pochtar; adaptador Eugenio Martínez Celadrán. Labor.

En estos diccionarios, los términos se presentan por orden alfabético. La microestructura de los artículos individuales incluye explicaciones del significado y, en algunos casos, ejemplos de uso de autores destacados. También se pueden encontrar en Internet diccionarios de lingüística organizados alfabéticamente, como el de Fernández López (2018). Este diccionario va más allá del ordenamiento habitual en el sentido de que la organización de la microestructura es muy variable. Debajo de cada palabra hay extractos de obras sobre el término correspondiente, como en el ejemplo siguiente, cuando se habla de la arbitrariedad del signo en un libro de introducción a la lengua española, donde se invita al lector a deducir una definición del concepto a partir de los extractos:

La arbitrariedad del signo

La relación que reúne en un signo lingüístico a un concepto y a una imagen acústica es arbitraria. Si observamos un signo lingüístico, la palabra *mujer*, por ejemplo, la sucesión de sonidos que forma su significante no tiene ninguna relación intrínseca con el concepto que evoca; le hecho de que pudiera haberse elegido cualquier otra secuencia sonora queda de manifiesto si comprobamos que en diversos códigos lingüísticos existen significantes diferentes para ese mismo concepto:

español: *mujer*

catalán: *dona*

francés: *femme*

inglés: *woman*

Hemos de detenernos un instante a reflexionar sobre lo que entendemos por *arbitrariedad*. No sugiere que el hablante pueda alterar el signo lingüístico a su capricho: una vez adoptado por la comunidad, ningún miembro de ella es libre de cambiarlo. Si se dan cambios en un signo, tanto en el plano fónico como en el significado, es por obra de la evolución histórica.

Lo que significa arbitrariedad a este respecto es que, en el momento en que una comunidad asoció una secuencia sonora a un concepto, lo acordó sin que mediara una razón que motivara esa unión; ésta se acepta convencionalmente en todas las lenguas. Por ello decimos que el signo lingüístico es *inmotivado* o *arbitrario*.

Algunos lingüistas consideran las onomatopeyas como signos lingüísticos motivados por una voluntad de imitación: Los sonidos de una palabra sugieren acústicamente el objeto o la acción que significan (*tic-tac*, *miau*, etcétera). También en las familias de palabras formadas por el procedimiento de derivación señalan mayor motivación que en las palabras primitivas. (Así, parecería más motivado *panadero* o *empanada* que *pan*). [Quilis, A. / Esgueva, M. / Gutiérrez, M. L. / Ruiz-Va, P.: *Lengua española*. Madrid: Centro de Estudios Ramón Areces, 1991, p. 34] (Fernández López, 2018, <https://n9.cl/1fd90>).

Aunque en Internet se permite incluir diversos textos, tablas y representaciones gráficas de los distintos términos, el orden alfabético excluye el establecimiento de relaciones entre ellos, como muestra el comienzo del orden alfabético en el diccionario de lingüística en línea antes mencionado:

Aoristo	Apóstrofe
Apelativo	Aprendizaje ostensivo
Apelativo o nombre común	Apriorismo
Apéndice	Aprosdóqueton
Apercepción	Arabismos en el español
Apical	Aragonés
Aplología - Aplografía	Arbitrariedad del signo lingüístico
Apócope	Árbol de Porfirio
Apódosis	Archifonema
Apofántico	Archilexema
Apokoinu	(Fernández López, 2018, https://n9.cl/2eke0o)
Aposición	
Aposiopsis o reticencia	

Aunque en los artículos se contrastan a veces diferentes puntos de vista, el contenido de los conceptos se presenta como algo predeterminado y no se

clasifica en función de su evolución histórica. Además, el orden alfabético no puede reflejar la relación entre términos y conceptos individuales.

Por otro lado, aunque los diccionarios de terminología lingüística que se centran en enfoques lingüísticos contemporáneos ofrecen a veces referencias a la historia de términos concretos, suelen situar el inicio de su desarrollo en función de aspectos problemáticos de la historia de la ciencia. El limitado horizonte retrospectivo de estos diccionarios resulta especialmente evidente cuando los argumentos se basan en la teoría de una determinada escuela o corriente de pensamiento y se ignoran otras consideraciones sobre el mismo término.

Esto se aplica también al concepto de ‘arbitrariedad del signo lingüístico’. Desde hace casi sesenta años, el estudio de Coseriu (1967) ofrece una visión histórica bien fundada de este concepto. El desdoblamiento del concepto aristotélico en *non natura sed ad placitum* ‘no de naturaleza, sino a voluntad’ y la consiguiente posibilidad de enfatizar e incluso aislar uno de los dos lados parece ser el punto de partida de posteriores desarrollos conceptuales. Por ello, a primera vista resulta aún más sorprendente que en las formulaciones de los diccionarios de terminología lingüística bajo la palabra clave *arbitrario/arbitrariedad* se mencione por lo general a Saussure como punto de partida a pesar de que este afirmaba explícitamente que se basaba en una tradición. Además, la definición varía mucho en cuanto al contenido y puede orientarse en distintas direcciones. Las siguientes citas de diccionarios lingüísticos sirven para ilustrarlo:

Arbitrariness: The absence of any physical correspondence between linguistic signals (such as words) and the entities in the world to which they refer. There is nothing in the way the word *table* is pronounced or written which physically resembles the thing ‘table’. The opposite view is sometimes maintained, with evidence adduced from onomatopoeic and other symbolic uses of sound. *See also* Nominalism; Onomatopoeia; Sound Symbolism (Crystal, 1992, p. 26).

Dans la théorie saussurienne, l'*arbitraire* caractérise le rapport qui existe entre le signifiant et le signifié. La langue est arbitraire dans la mesure où elle est une convention implicite entre les membres de la société qui l'utilisent; c'est dans ce sens qu'elle n'est pas “naturelle”. Le concept qu'exprime un mot comme *corde* n'a aucun rapport de nécessité avec la suite des sons [kord] ou la graphie *corde*. La preuve en est que les langues aussi voisines que le français et l'italien ont

pour désigner des objets identiques des mots entièrement différents [...] *Arbitraire* exclut dans cette acception la possibilité pour le sujet parlant de faire dépendre de sa volonté personnelle le choix de la forme exprimant tel signifié ou le choix d'un signifié pour telle forme [...] (Dubois, 1994, pp. 46-47).

Arbitrarität (lat. arbitrium 'Willkür'. Auch: Unmotiviertheit, Willkürlichkeit) Auf F. de Saussure (1857–1913) zurückgehende Bez. für die Beliebigkeit des sprachl. Zeichens. Das sprachl. Zeichen ist willkür. geschaffen, es gibt keinen naturgegebenen Zusammenhang zwischen dem Lautkörper des Zeichens und dessen Inhalt. [...] (Glück, 1993, p. 52) 'Arbitrariedad (del latín arbitrium 'arbitrariedad'. También: inmotivación, a voluntad) Término utilizado por F. de Saussure (1857-1913) para describir la arbitrariedad del signo lingüístico. El signo lingüístico se crea arbitrariamente, no hay conexión natural entre el cuerpo fonético del signo y su contenido.'

Además de la falta de profundidad del horizonte retrospectivo, la disposición alfabética de los diccionarios lingüísticos también impide considerar la aparición de nuevos términos lingüísticos. En la historiografía de la lingüística, hasta ahora solo se ha postulado de manera esporádica y global que los términos socializan, posibilitando así la comunicación consensual y disensual y no solo nombran, sino también crean objetos científicos en el contexto de los discursos e instituciones científicas (cf. Knobloch, 1996, p. 261).

Dentro de la historia de los conceptos, los conceptos básicos se entienden como indicadores de relaciones sociales y profesionales del pasado. Al mismo tiempo, se consideran factores constitutivos y confirmatorios de estas conexiones (Koselleck, 1979, pp. 107-109). En la investigación sobre la historia de los conceptos en la historia de la ciencia se intenta "espiar" las relaciones entre los conceptos en retrospectiva (Knobloch, 1996, p. 262), reconstruyendo su funcionamiento en la economía lingüística y simbólica del respectivo discurso disciplinar o general.

Las siguientes ideas elementales de la historia de la lingüística ya ponen en duda que la continuidad de un cuerpo de palabra esté relacionada con la constancia conceptual y que el cambio terminológico esté vinculado a la innovación conceptual:

1. No se puede dar por sentado que cuando se crea un nuevo concepto exista inmediatamente un nombre generalmente aceptado

que lo transmita sin dificultad. Un ejemplo de ello es el concepto de ‘estructura de la información’, para el que se han utilizado y se utilizan diferentes términos: *tema*, *rema*; *soporte*, *aporte*, *tema psicológico*, *enfoque* y *fondo*. En alemán, la palabra *Nachdruck*, que significa ‘firmeza, vehemencia’, se remonta incluso a finales del siglo XVIII como término para designar hechos comparables (Haßler, 2019).

2. Si un término está establecido y aceptado para un concepto, esto no es razón para suponer su uso coherente. La diferenciación de escuelas lingüísticas y campos de investigación conduce a menudo a la polisemia, incluso en el ámbito conceptual. Un ejemplo de ello es el uso del término *modalidad* para denotar la actitud del hablante ante el contenido y para las diversas realizaciones materiales del lenguaje.
3. También se puede suponer un espectro diversificado de denominaciones para conceptos lingüísticos establecidos. Esto también se aplica al concepto de ‘arbitrariedad del signo lingüístico’, para el que se pueden encontrar los siguientes términos en textos de algunas lenguas europeas de los siglos XVII y XVIII, que pueden acentuar diferentes aspectos del concepto, como la independencia de la forma sonora del significado, la configuración institucional del signo, la configuración según la voluntad de las personas que determinan la denominación y la artificialidad del signo: lat. *arbitrarium*, *signum arbitrium*, *arbitrarie impositiones*, *ad placitum nomina significare*, *signum artificiale*; al. *willkürliche Zeichen*, *willkührliche Zeichen*; *künstlichste[n]* und *willkührlichste[n]* Zeichen, *willkürliche Zeichen der Gegenstände*, *zufällige Zeichen der Gegenstände*, *Eine durch Fleiß und Uebung errichtete so genaue Verknüpfung zwischen den zu denkenden Begriffen und den hörbaren Tönen*; ingl. *arbitrary*, *arbitrarily*, *a perfect arbitrary imposition*; *not natural, but only imposed and arbitrarily*; *arbitrary marks*, *arbitrary characters*, *voluntary imposition*, *voluntary signs*; fr. *arbitraire*, *caractère arbitraire*, *signe arbitraire*, *Signes arbitraires*, *les signes arbitraires*

des idées, mots libres et arbitraires, l'arbitraire dans les langues, signes d'institution, d'institution arbitraire, signes artificiels, reflexion ou choix arbitraire, des significations accidentelles, signification arbitraire; it. arbitrario, segni arbitrarii, non per sua natura, mà pel consenso, & patto degli huomini trà di loro; segno non naturale, ma arbitrario; esp. Determinacion á significar tal objeto arbitraria ó dependiente de la libre voluntad de los hombres (Haßler & Neis, 2009, p. 206).

Incluso en el ámbito de términos gramaticales, las mismas formas verbales recibían nombres diferentes en las gramáticas españolas. Por ejemplo, el pretérito perfecto simple se denominaba *tiempo passado acabado* (Nebrija, 1980 [1492], 185, pp. 31-39), *passado* (Villalón, 1558), *passado cumplido i acabado* (Correas, 1626, p. 156), *preterito perfeto* (Jiménez Patón, 1614; Villar, 1651), *pretérito perfecto* (Martínez Gómez Gayoso, 1769, San Pedro, 1769), *pretérito perfecto remoto* (RAE, 1771; RAE, 1796), *pretérito perfecto remoto simple* (Calleja, 1818, p. 25), *pretérito definido* (Noboa, 1839, p. 84), *pretérito absoluto* (Salvá, 1852, p. 172) (cf. Haßler, 2016, pp. 170-176). Como se verá en la sección 5 de este artículo, hay diferentes enfoques conceptuales detrás de estos términos, aunque se refieran a la misma forma.

3. EL PRINCIPIO ONOMASIOLÓGICO

Si queremos analizar la aparición y el desarrollo de los términos lingüísticos, es evidente que resulta más conveniente una ordenación distinta de la alfabética. Para esta otra ordenación, recordemos, en primer lugar, el triángulo semiótico según Ogden y Richards en su obra *The Meaning of Meaning* (1923) (Imagen 1).

En el ejemplo de las denominaciones del pretérito perfecto simple, la referencia es una forma verbal y el concepto es la idea que el gramático respectivo tiene de sus propiedades, por lo que puede centrarse en la simplicidad morfológica, la completitud de la situación designada o la distancia del momento del habla. El símbolo o significante es el nombre elegido para la forma verbal.

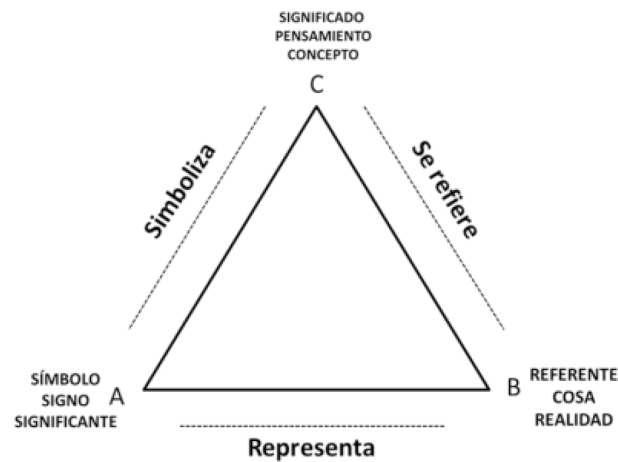


IMAGEN 1: El triángulo semiótico según Ogden y Richards

Mientras que los diccionarios de lingüística ofrecen una representación semasiológica y explican los términos ordenados alfabéticamente, en una representación onomasiológica se ofrece la posibilidad de partir de un hecho de la realidad lingüística o de un concepto y preguntarse por sus realizaciones terminológicas (Gráfico 1):

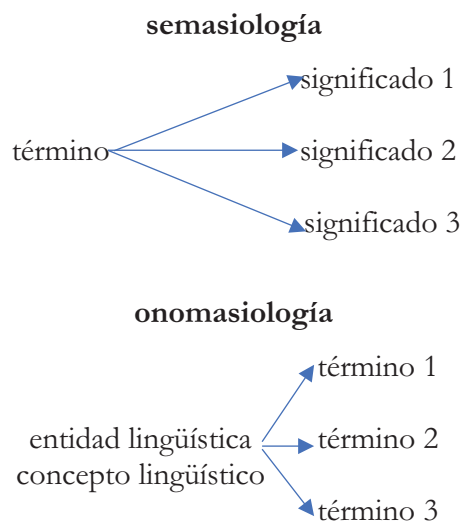


GRÁFICO 1: Semasiología y onomasiología

El término *onomasiología*, del griego ὄνομα ‘nombre’, fue acuñado por primera vez por Zauner (1902) y se refiere a la rama de la lingüística que se

ocupa de la asignación de un nombre a un concepto o referente. La investigación onomasiológica dio lugar a los “diccionarios ideológicos” en lexicografía que se organizan según relaciones conceptuales y pretenden ayudar en la redacción y mejorar el estilo (por ejemplo, Gómez Carillo y Sola, 1925). La aplicación del método onomasiológico perturba naturalmente el orden alfabético de los términos.

Cuando Vossler describió en 1919 la onomasiología o teoría de la designación como “uno de los mayores avances de la lingüística, [...] cuyo alcance aún apenas puede estimarse” (Vossler, 1919, p. 43), se refería a una definición metodológica precisa que surgió en el transcurso del siglo XIX y que había dado lugar a numerosos estudios individuales en torno al cambio de siglo. Estos estudios lexicológicos partían inicialmente de las cosas y luego se preguntaban por su denominación (Quadri, 1952). Los puntos de partida para una renovación de las cuestiones onomasiológicas en la lexicografía surgieron, sobre todo, con el sistema conceptual diseñado por Hallig y von Wartburg (1952) como base de la lexicografía. Con el creciente desarrollo de los métodos cognitivos en lingüística, el método onomasiológico experimentó una nueva renovación. Sobre esta base se desarrolló una teoría coherente de los cambios semánticos que también puede aplicarse lexicográficamente (cf. Blank & Koch, 2003; Grzega, 2015; Fernández-Domínguez, 2019). La práctica de los esfuerzos históricos conceptuales debe tener en cuenta el hecho de que existen conceptos de lógica y alcance diferentes en cada materia, que las formas de palabras persistentes (o idénticas) señalan poca continuidad o identidad, así como las nuevas formas de palabras señalan discontinuidad (Knobloch, 1996, p. 262).

4. LA APLICACIÓN DEL PRINCIPIO ONOMASIOLÓGICO EN EL *LEXIKON SPRACHTHEORETISCHER GRUNDBEGRIFFE DES 17. UND 18. JAHRHUNDERTS*

En el *Lexikon sprachtheoretischer Grundbegriffe des 17. und 18. Jahrhunderts* (‘Léxico de los fundamentos teórico-lingüísticos de los siglos XVII

y XVIII', Haßler & Neis, 2009) se ha aplicado sistemáticamente el método onomasiológico, es decir, se ha abandonado el orden alfabético y se ha optado por un orden conceptual multinivel. A continuación, se muestra la traducción al español de la primera parte¹ (Gráfico 2):

ONTOLOGÍA / SIGNOS / LENGUAJE

La naturaleza de la lengua
lenguaje natural
lenguaje hablado humano (frente a otros
signos)

Arbitrariedad
Convención
Naturalidad
Linearidad

Signo vs. idea (concepto)

Adquisición del lenguaje

Adquisición deficiente del lenguaje

Funciones del lenguaje

Función comunicativa de la lengua
Función cognitiva del lenguaje
Función social de la lengua

ORIGEN/CREACIÓN/DESARROLLO

Origen del lenguaje

Lengua original
Confusión lingüística
Cambio de idioma
Corrupción

Lengua de signos frente a lengua hablada

Etimología

Analogía

Normalización

Apología

UNIDAD vs. DIVERSIDAD

Universalidad y diversidad

Carácter especial de una lengua

Lenguaje universal

Comparación y tipo de lengua

Dialecto

Méritos / perfección / defectos

Claridad (clarté)

Eufonía

Riqueza

Validez universal

Perspicuitas, energeia, abundancia y armonía

Uso de la lengua

Uso (Usus, usage)

Estilo

Abuso

Traducción

Descripción gramatical

Gramática

Sintaxis

Oración

Orden de las palabras/ordo naturalis/inversión

Partes de la oración

etc.

GRÁFICO 2: Primera parte de la grilla de conceptos en el léxico de los fundamentos teórico-lingüísticos de los siglos XVII y XVIII

El término *Grundbegriffe* ('conceptos básicos') pretende abarcar un tipo de conceptos anclados tanto en la conciencia lingüística de la época como en la teoría lingüística y las diversas formas de uso práctico de la lengua. Para

¹ La numeración discontinua se debe a que no se incluyeron en el léxico todas las entradas de la tabla de términos creada anteriormente.

algunos de los conceptos incluidos aún no había surgido un término fijo en aquella época, por lo que hemos optado por expresiones descriptivas, como ‘carácter especial de una lengua’.

Los conceptos se constituyen en textos y en interacción con series de textos. El objetivo de este léxico era presentar las conceptualizaciones del pensamiento lingüístico anteriores a la institucionalización de la lingüística, observarlas en su forma específica en los siglos XVII y XVIII y situarlas en desarrollos más amplios desde la Antigüedad hasta nuestros días.

Por supuesto, no podíamos partir de un término para cada concepto y analizar su uso en los corpus disponibles electrónicamente. Esto no es posible porque en el siglo XVII no existía un término fijo para muchos conceptos. La búsqueda de los términos para conceptos lingüísticos que apenas se estaban constituyendo solo era posible, por supuesto, mediante la lectura atenta, es decir, una lectura precisa que rastrea detalles, matices de significado y características lingüísticas y que sitúa el texto como objeto en el centro de interés. Este método es similar a la lectura lingüística minuciosa (*linguistic close reading*, Hallet 2010), que también se practicó en el proyecto “Libros de texto de lenguas extranjeras en el siglo XVII” del Centro de Investigación en Colaboración (*Sonderforschungsreich*) 980 “Episteme en movimiento. Transferencia de conocimientos del mundo antiguo a la Edad Moderna”². Este enfoque pone gran énfasis en lo específico en comparación con lo general, presta mucha atención a las palabras individuales, la sintaxis y el orden de las frases y las palabras, pero presupone una idea aproximada de un punto objetivo del proceso de conceptualización. De este modo, por ejemplo, podían encontrarse las siguientes denominaciones del concepto de ‘partícula’, especialmente importante en la lingüística misionera:

² <https://n9.cl/38pgb>

lat. *particula, particulae*; al. *Bestimmungswörter (particulae), eine Particul, Particuln, Wörterchen, Nebenwörtlein, Bestimmwörtlein, Beziehwörtlein, Zuwörter, Vorwörtlein, Fügwortlein, Zwischenwörtlein*; ingl. *particle, conjunction, preposition*; fr. *particules, interjection, interjectives, discursives, exclamatives, acclamatives, imprecatives, assertives, admonitives, imitatives, exhibitives, explétives, précurives*; esp. *particula, conjunciones; partículas, aquellas únicas y singulares partes*; sueco. *Partiklar*, finl. *partikkeli*. (Haßler & Neis, 2009, p. 1338)

Dada la enorme variedad de términos y denominaciones, no podemos limitarnos a buscar una palabra o un grupo de palabras en los textos. Existe una asimetría entre la vertiente conceptual y la denominativa que hace imposible establecer correspondencias directas. El enfoque onomasiológico se modifica en varios aspectos. En primer lugar, se tiene en cuenta el hecho de que los significados de los términos en las distintas lenguas solo coinciden parcialmente. En segundo lugar, la constitución de conceptos se concibe como un proceso dinámico que debe estudiarse en relación con la denominación. Por último, los conceptos se desarrollan en los textos, y, a menudo, su naturaleza solo se concibe a partir de la argumentación en la que aparecen.

5. EL PRINCIPIO ONOMASIOLÓGICO EN LA HISTORIOGRAFÍA DE LA LINGÜÍSTICA Y LAS GRAMÁTICAS ESPAÑOLAS

En español también se puede encontrar una gran variedad de términos relacionados con conceptos de la lingüística y la gramática. A modo de ejemplo, a continuación, se abordará la denominación de los tiempos, en particular los dos pretéritos perfectos, en la gramatización del español. Por último, nos referiremos a un manual de lengua de un autor español de la primera mitad del siglo XVII, que también ilustra el problema de la denominación.

La gramática de todas las lenguas románicas se basa en la tradición grecolatina, de la que no solo imitaron y adaptaron las primeras gramáticas de las lenguas vernáculas, sino que también tomaron prestada una y otra vez las gramáticas posteriores. Esto provocó problemas, sobre todo en la denominación de los tiempos, ya que la estructura de las formas verbales en las lenguas

romances no es congruente con la latina y los términos se reinterpretaron y utilizaron sin referencias precisas.

La gramática de Prisciano (hacia 500), que es extremadamente breve en su descripción de los tiempos, fue especialmente decisiva para las primeras gramáticas francesas (cf. Fournier, 2013, p. 14-15):

Sunt igitur tempora tria, *praesens*, *praeteritum* et *futurum*, sed *praeteritum* rursus dividitur in tria, in *praeteritum imperfectum*, *praeteritum perfectum*, *praeteritum plusquamperfectum*. nec mirum tam late patere *praeteritum* tempus, cum in notitiam nostram nihil sic naturaliter a longo saeculorum spatio potest venire, quomodo actus *praeteriti* temporis. in *praesenti* enim et futuro pleraque incerta nobis sunt angustissimaque *esteorum* cognitio nobis et dubia plerumque. (Priscian 1855-1859, *Institutiones grammaticae*, liber VIII. 405.8-14)

En este pasaje, Prisciano clasifica los tiempos basándose en la división cronológica del tiempo físico en presente (*praesens*), pasado (*praeteritum*) y futuro (*futurum*) y justifica la subdivisión ulterior del pasado con la certeza de nuestro conocimiento de los acontecimientos pasados. Por el contrario, atribuye un carácter incierto al tiempo presente y al futuro, pues nuestro conocimiento de ellos es muy limitado y, por lo general, dudoso. Prisciano hace otra distinción con respecto a la cualidad aspectual del tiempo pasado. Distingue entre el *praeteritum imperfectum*, que se utiliza para describir acciones que se han iniciado, pero no se han completado, y el *praeteritum perfectum*, que corresponde a acciones completadas, y el *praeteritum plus quam perfectum*, que corresponde a acciones que se han completado hace algún tiempo. Prisciano propone cinco tiempos diferentes, tomando como referencia, en el primer nivel de diferenciación, la posición del hablante en el tiempo físico (Gráfico 3). Para el pasado, abre subcategorías que se diferencian según el grado de realización (no realizado, realizado, realizado durante algún tiempo). La repetición del término *praeteritum* es notable en las designaciones, al igual que la aparición de *perfectum* en las tres designaciones del tiempo pasado (*imperfectum*, *perfectum*, *plus quam perfectum*):

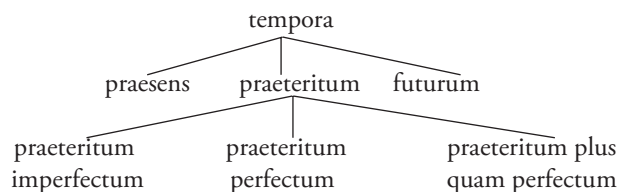


GRÁFICO 3: Denominaciones de los tiempos verbales en la gramática de Prisciano

Estas categorías se relacionan entre sí en el *genus proximum* y la *differentia specifica* y, además del momento del habla, remiten implícitamente a otro punto de referencia en el pasado. Con este planteamiento, Prisciano ya demuestra estar en la senda de las descripciones modernas de las relaciones temporales. Además de Prisciano, los textos *Ars minor* y *Ars maior* de Aelius Donatus (ca. 320-380) también fueron importantes referencias para las primeras gramáticas de las lenguas romances (cf. Schönberger, 2008, 2009). En su *Ars minor*, ya utilizaba los términos que también se encontraban en el Prisciano y los organizaba en un esquema de preguntas y respuestas que resultaba adecuado para la memorización y que posteriormente se aplicó con frecuencia a las lenguas vernáculas. Aquí, el término *tempus* se deriva de la cronología del tiempo físico y se distinguen cinco tiempos diferentes en la inflexión de los verbos:

tempora uerborum quot sunt? tria.
 quae? praesens, ut lego, praeteritum, ut legi, futurum, ut legam.
 quot sunt tempora in declinatione uerborum? quinque.
 quae? praesens, ut lego, praeteritum imperfectum, ut legebam, praeteritum perfectum, ut legi, praeteritum plusquamperfectum, ut legeram, futurum, ut legam.
 (Donatus, *Ars minor*, en Schönberger, 2008)

Los primeros gramáticos españoles se basaron principalmente en modelos latinos, entre los que destaca el *Ars minor* de Donat. Esta doctrina también influyó en Antonio de Nebrija y sus *Introductiones latinae* (1481; 2ª ed. 1485), que fueron la primera gramática humanística de modelo italiano en España. Estas *Introductiones latinae* tuvieron éxito y su autor introdujo cambios en ediciones posteriores. En la última edición antes de su muerte, por ejemplo,

aumentó a seis el número de formas verbales. Ya antes de su *Gramática castellana* (1492), Nebrija tradujo su gramática latina y señaló en una glosa marginal que el metalenguaje de la descripción del español determinado por el latín llevaba a nombrar categorías lingüísticas que no existían en español (Nebrija, 1996 [1488]; cf. Baldischwieler, 2004, p. 26):

E regione, id est 'ex opposito'; hoc est ut sermo hispaniensis sit positus contra latinum. *Hispanienses* vero dixit non *hispanas*, quia introductiones illae in hispanum sermonem non usque adeo migrare potuerunt ut non maximam latinitatis partem retinuerint. Nam *hispanum* est 'quod in Hispania' atque 'ex Hispania est,' *hispaniense* vero 'quod non penitus hispanum', sed aut 'in Hispania ab externis geritur', ut *bellum hispaniense*, quod ab Romanis in Hispania aut 'extra Hispaniam ab Hispanis', ut *collegium hispaniense*, quod est Bononiae ab immortalis memoriae viro Aegidio Albornozio S. R. C. cardinali hispano fundatum. (Nebrija 1996 [1488], ILC, prólogo, f. 5vo-6ro)

Para traducir los nombres de las formas verbales recurría principalmente a traducciones prestadas:

praeteritum imperfectum = pasado & no acabado (f. 14ro et passim)
praeteritum perfectum = pasado & acabado (f. 14ro et passim)
praeteritum plus quam perfectum = pasado & mas que acabado (f. 14ro et passim)

En otros casos, utilizó lexemas españoles y les dio el significado de palabras latinas:

futurum = *venidero* (f. 7vo et passim)
genus (verbi) = *genero* (f. 40ro et passim)
modus = *manera* (f. 7ro et passim)

Según la *Gramática castellana* de Nebrija, existen cinco tiempos de indicativo y cinco de subjuntivo. Para nombrarlos, utiliza préstamos del latín (*tiempo presente*, *tiempo futuro o venidero*) y traducciones (*tiempo pasado no acabado*, *tiempo pasado acabado*, *tiempo pasado más que acabado*). Nebrija pudo inspirarse en dos obras anteriores que contenían descripciones de las formas verbales: el *Compendium grammaticae* de Juan de Pastrana y un tratado de Andrés Gutiérrez de Cerezo (cf. Ridruejo, 1977, p. 56-58). Estos autores distinguían entre tres formas del tiempo perfecto en latín: el aoristo, el perfecto compuesto y el pretérito perfecto. Nebrija adaptó su sistema al

de estos autores y dedujo la función de las formas verbales a partir de sus nombres:

Presente tiempo se llama aquél en el cual alguna cosa se haze agora, como diziendo *io amo*.

Passado no acabado se llama en el cual alguna cosa se hazía, como diziendo *io amava*.

Passado acabado es aquél en el cual alguna cosa se hizo, como diziendo *io amé*.

Passado más que acabado es aquél en el cual alguna cosa se avía hecho, cuando algo se hizo, como *io te avía amado, cuando tú me amaste*.

Venidero se llama en el cual alguna cosa se a de hazer, como diziendo *io amaré*.

(Nebrija, 1980 [1492], 185, pp. 31-39)

Para las gramáticas de orientación latina, las formas verbales analíticas del español planteaban un problema, ya que había que integrarlas en el sistema. Nebrija decidió describir cinco formas verbales (*presente, pasado no acabado, pasado acabado, pasado más que acabado, venidero*), pero también tuvo en cuenta los tiempos que existen en latín y faltan en español. Describe su composición a partir de las formas del verbo *haber*, a las que se añade una forma verbal nominal. Nebrija (1492, pp. 83-84) añadió esta forma verbal nominal bajo el nombre *de nombre verbal infinito* como un nuevo tipo de palabra que expresa el sentido de la acción verbal. Por ejemplo, Nebrija describe un *passado acabado por rodeo*, que se presenta en dos formas:

Assi dize el passado acabado por rodeo en dos maneras, una por el presente del indicativo; y otra por el mesmo passado acabado; diziendo *io e amado*, y *ove amado*. El passado mas que acabado dize por rodeo del passado no acabado diziendo: *io avia amado*. (Nebrija, 1980 [1492], p. 79)

En las primeras gramáticas no se distinguía entre tiempo cronológico y tiempo gramatical, por lo que se utilizaba simplemente el término *pasado*. Desde Nebrija, sin embargo, se había establecido la distinción entre formas verbales terminadas (*acabado*) e incompletas (*no acabado*). Villalón, sin embargo, rechazó esta distinción y asumió solo tres tiempos. Nebrija había añadido dos pretéritos, pero estos solo eran necesarios para describir la elegancia de

la lengua latina. No había necesidad de aumentar el número de tiempos para el español:

[...] ay solas tres differēcias del tiempo por donde se varia el verbo por la lengua Castellana. Tiēpo presente, como yo amo. Y tiēto passado, como yo ame. Y tiempo futuro, que es el que esta por venir, como yo amare. Por solos estos tres tiēpos dizo toda su habla y cōtinua su conuersaçion el Castellano. Aunque Antonio de Nebrija en su Arte de Gramatica Latina añadió dos preteritos y dos futuros, perfecto e imperfecto y pluscuamperfecto. Pero esto fue porque ansi lo hallo en la elegancia latina: y ANSI lo puso el preceptiuamēte, para que ofrecido el verbo en algun auctor lo entendiesse el que lo leya en su verdadera sinificaçion y intincion. Pero nuestra lēgua Castellana no tiene neçesidad desta multiplicaçion de tiempos. (Villalón, 1558, p. 36)

En cuanto a las denominaciones de las formas verbales del pasado en las primeras gramáticas españolas, llama la atención que el pretérito perfecto compuesto no se tratara inicialmente en muchas gramáticas y, en el mejor de los casos, se mencionara simplemente bajo la conjugación del verbo auxiliar *haber*. Las gramáticas del siglo XVII coinciden en la designación del pretérito perfecto simple y del pretérito imperfecto (Tabla 1):

formas verbales	Nebrija 1492	Villalón 1558	Correas 1626	Jiménez Patón 1614	Villar 1651
<i>canté</i>	pasado acabado	passado	pasado cumplido i acabado pretérito perfecto	preterito perfeto	preterito perfeto
<i>cantaba</i>	pasado no acabado	-----	pasado no-cumplido ni acabado pretérito imperfecto	preterito imperfeto	preterito imperfeto
<i>be cantado</i>	passado acabado por rodeo	-----	-----	-----	-----

TABLA 1: Denominaciones de los tiempos verbales en gramáticas del siglo XVII

Correas (1626, p. 156) siguió un orden diferente al de Nebrija al describir las categorías del verbo. Empezó por los tiempos, luego trató las personas, los

números y los modos; supuso cinco tiempos y usó la doble característica como *cumplido i acabado*:

Presente, qe demuestra lo qe ahora se haze; *pasado no-cumplido ni acabado*, qe muestra lo qe se hazía i no se acabó; *pasado cumplido i acabado*, qe muestra lo qe se hizo i acabó; *pasado mas que cumplido*, qe muestra qe algo se hizo antes qe otra cosa, i antes qe otro hiziese; *venidero*, qe muestra lo qe se ha de hazer adelante. (Correas, 1626, p. 156)

También mencionó los términos latinos y justificó su uso diciendo que se conocían gracias al estudio de la lengua latina:

A estos cinco tiempos en la Gramática Latina llaman *Presente*, *pretérito imperfecto*, *pretérito perfecto*, *pretérito pluscuamperfecto*, *futuro imperfecto*, los cuales términos usaremos alguna vez como notorios á los qe estudian. (Correas, 1626, p. 156)

Al utilizar los términos españoles y latinos como equivalentes entre sí, Correas consideró que los sistemas de ambas lenguas eran similares y no complicó su descripción con los mismos términos.

Esta tradición de clasificación continuó en el siglo XVIII (Tabla 2). Con el tiempo, las formas verbales compuestas se integraron en la descripción del paradigma verbal. Martínez Gómez Gayoso distinguió solo seis tiempos y los relacionó con funciones temporales poco precisas. Denominó las formas verbales según la tradición de los gramáticos latinos:

El Tiempo es la vária inflexión del Verbo, según el tiempo que denota. Los Tiempos son seis: *Presente*, *Pretérito imperfecto*, *Pretérito perfecto*, *Pretérito plusquam perfecto*, *Futuro imperfecto*, y *Futuro perfecto*. (Martínez Gómez Gayoso, 1769, p. 124)

Además de los tiempos propiamente dichos, Gómez Gayoso se ocupó también de los verbos auxiliares que “ayudan à la conjugación de los demás Verbos” (Martínez Gómez Gayoso, 1769, p. 126). Integró las formas del perfecto compuesto (*he consumido*) y del *pretérito anterior* (*hube consumido*) en el paradigma verbal, pero no implementó ninguna diferencia funcional entre estas formas y el perfecto simple (*consumí*). Las tres formas aparecen bajo la denominación *pretérito perfecto* (por ejemplo, Martínez Gómez Gayoso, 1769, p. 179). En la gramática del Padre Benito de San Pedro, las formas

verbales compuestas se integran en el paradigma de conjugación. En esta gramática, el *pretérito* perfecto compuesto se denomina *pretérito indefinido*. San Pedro no justifica el uso del término *indefinido*, que se aparta de la tradición griega, pero probablemente lo utilizó para el perfecto compuesto porque esta forma verbal está relacionada con el presente y, por tanto, solo tiene un valor temporal relativo:

El Pretérito indefinido [denota] una cosa passada en un tiempo del que dura algo todavía, o que es passado poco a, v.g. *Yo e estado enfermo este año, oeste mes: E oido Missa esta mañana*. (San Pedro, 1769, vol. II, p. 56)

Al igual que en Francia, el concepto de ‘indefinido’ también se aplicó a diferentes formas verbales en el siglo XVIII. La referencia al aoristo griego se había perdido y los gramáticos categorizaron diferentes formas verbales como indefinidas basándose en diversos criterios. En el siglo XVIII, el pretérito perfecto compuesto dejó de tratarse simplemente como el presente de *haber* con un participio o un *nombre verbal infinito*. La Real Academia Española introdujo el término *pretérito perfecto próximo* para esta forma verbal, teniendo así en cuenta algo parecido a la regla de las 24 horas: los acontecimientos denotados por el tiempo pretérito perfecto compuesto deben estar próximos en el tiempo al momento del habla:

formas verbales	Martínez Gómez Gayoso 1769	San Pedro 1769	RAE 1771	RAE 1796
<i>canté</i>	pretérito perfecto	pretérito perfecto	pretérito perfecto remoto	pretérito perfecto remoto
<i>cantaba</i>	pretérito imperfecto	pretérito imperfecto	pretérito imperfecto	pretérito imperfecto
<i>be cantado</i>	-----	pretérito indefinido	pretérito perfecto próximo	pretérito perfecto próximo

TABLA 2: Denominaciones de los tiempos verbales en gramáticas del siglo XVIII

Ya en la primera edición de la Gramática de la Real Academia Española, se distinguieron las formas verbales *simples y propias* de las *compuestas e impropias*

(RAE, 1771, p. 74)³. Para justificar la denominación de *impropias*, los autores aludieron a la incapacidad de estas formas verbales para expresar un significado temporal con una sola palabra. Este criterio, que devaluaba la tendencia analítica del español, resultaba naturalmente de la comparación con el latín. No obstante, también se introdujo el término *compuesto* con base morfológica:

Llámanse *impropios*, porque no tienen la propiedad de expresar el tiempo con solo una palabra, y son formados para traducir y suplir por algun rodeo otros tiempos semejantes, propios de la lengua latina.

Llamanse también *compuestos*, porque se componen de dos, ó tres palabras. (RAE, 1771, p. 74)

La integración vacilante de las formas verbales compuestas en la descripción de la conjugación de los verbos españoles y las rarezas en la denominación de estas formas verbales parecen haber sido causadas por dos factores hasta el siglo XVIII. Por un lado, los autores de las gramáticas transpusieron la gramática latina y sus términos a la lengua española. A pesar de su alto grado de gramaticalización, las formas analíticas no encajaban en este sistema. Por otro lado, los autores solo tenían en cuenta los términos semasiológicamente, es decir, derivaban las funciones de las formas verbales del significado de los términos. Las características semánticas y el uso de estas formas apenas se tuvieron en cuenta y no se planteó la cuestión onomasiológica de su denominación adecuada.

Incluso en las gramáticas escritas para hablantes no nativos, se consideraba que los tiempos del español y del latín eran similares. En este contexto, es interesante la descripción de los dos tiempos perfectos del español en la introducción de Sanford (1611), quien los distingue de la siguiente manera: el tiempo aoristo solo indica la conclusión de la acción y no contiene un rasgo temporal, mientras que el tiempo perfecto compuesto expresa la conclusión reciente de un proceso y un rasgo temporal. De este modo, las condiciones

³ Sobre las categorías del verbo en las gramáticas de este período, véase Gómez Asencio (1981 y 2011).

existentes originalmente en griego se proyectan sobre el español, cuya realidad, sin embargo, distaba mucho de aquel:

They have two Preterperfect tenses; one faith *Bartholomeus Granius* which aunswereth to the Greek Aorist, because it signifieth a thing done and past, without determining the time when, as *yo ame*, I loued a while agoe, *yo andúve* I went, *yo vine* I came: The other signifieth a thing lately past with determination of the time, as *ya he comido oy venado*, I have eaton venison to day. (Sanford, 1611, p. 25)

La aplicación de los conocimientos procedentes del estudio onomasiológico de conceptos lingüísticos y gramaticales también puede ser relevante para la comprensión de textos individuales que han sido poco estudiados. Por ejemplo, *Das Neue Sprachbuch* (en español, *El nuevo libro de la lengua*) de Juan Ángel Zumarán (cf. Corvo Sánchez, 2007), publicado en Múnich en 1621, está destinado a apoyar el aprendizaje del alemán, francés, italiano y español. El siguiente análisis de este libro de texto de lengua se limitará a las áreas conceptuales del orden y enlace de los elementos oracionales, los tiempos y modos del verbo y las consideraciones contrastivas.

Los gramáticos del siglo XVII también contaban las propiedades del sustantivo entre los medios de ordenación y enlace, porque ayudan a reconocer la función del nexo oracional y a establecer relaciones en la oración. Con respecto a estas propiedades, la gramática de Zumarán es más progresista en cuanto al género, ya que escribe:

Die Franzosen und Welschen haben kein Genus neutrum, und do sie es schon haben/ thun sie es dem generi masculino zuaignen. (Zumarán, 1621, p. 14)

‘El francés y el galés no tienen género neutro, y si ya lo tienen, lo asignan al género masculino.’ (Zumarán, 1621, p. 14).

Aunque esto es lingüísticamente incorrecto, ya que los neutros también podrían convertirse en femeninos (cf. *gaudium* > pl. *gaudia* > fem. *joie*), al menos hace justicia a la lengua francesa actual y no se traslada del latín la existencia de tres géneros. Zumarán asume tres géneros para el italiano y el español, aunque el neutro se utilizaría en pocas ocasiones. Obviamente, asume la forma del artículo en - *lo* como neutro, que, en italiano, sin embargo, depende del sonido

inicial de los sustantivos masculinos. Al asignar el género y el caso que determina el papel de la cláusula al artículo, se reconoce su función.

Entretanto, sin embargo, se había iniciado un intenso debate sobre la disposición de los elementos oracionales como medio para establecer relaciones entre ellos, que comenzó en Francia, pero pronto se extendió a otros países (cf. Haßler, 2014). En el caso de Zumarán, esta discusión al menos no se refleja en enunciados metalingüísticos, aunque los patrones lingüísticos que transmitía en sus diálogos naturalmente también proporcionaban consejos prácticos sobre el orden de las palabras. Para él, el latín es la base para la enseñanza de las cuatro lenguas que describe en su libro (alemán, francés, italiano y español). Además, un profesor de idiomas también debería dominar la lengua del país en el que desea enseñar (Zumarán, 1621, p. 3). El hecho de que Zumarán no tenga problemas con las inversiones se demuestra también en sus diálogos de ejemplo, incluso en francés (Zumarán, 1621, p. 328) y, por supuesto, en alemán (Zumarán, 1621, p. 331).

Zumarán había formulado la idea de que la conjugación en lenguas extranjeras puede aprenderse a partir de la lengua materna y, como muchos otros, también empieza por los auxiliares (Zumarán, 1621, p. 193). Sin embargo, solo se debe aprender la conjugación una vez se haya comprendido la pronunciación y la declinación.

Zumarán comienza con el *verbum possessivum haber*, lo que resulta obvio dado el uso generalizado de *haber* para construir las formas verbales analíticas en español. A continuación, aparece el *verbum substantivum (ser)*, término que también se utilizaba para la cópula siguiendo la gramática de Port Royal. No se distingue entre la función como verbo pleno, cópula y auxiliar y tampoco se indica cuál de los dos verbos se utiliza para formar los tiempos compuestos en francés, italiano y alemán. Las clases de conjugación se presentan a continuación basándose en el latín.

En la descripción alemana del verbo (Zumarán, 1621, p. 64) como tipo de palabra, Zumarán traduce el término latino *pars orationis* literalmente

como *theil der Oration* así como *modus* con el alemán *manier* y el español *manera*. Distingue cinco modos, que nombra según el modelo latino, que también incluye variantes:

Indicativus, Demonstrativus
Imperativus, Mandativus
Optativus, Desiderativus
Conjunctivus, Subunctivus
Infinitivus

Zumarán basa inicialmente su descripción de los tiempos en la división cronológica en presente, pretérito y futuro (Zumarán, 1621, pp. 64-65), pero luego añade dos formas verbales compuestas, el perfecto y el pluscuamperfecto. Por último, distingue el tiempo *definitum o determinatum* del *indefinitum, absolutum o auristum (aoristo)* y asigna estos tiempos a lenguas extranjeras, ya que no se dan en alemán. En las descripciones de otras lenguas también hay absurdos terminológicos, como en el francés *présent passé* e *imparfait parfait* (Zumarán, 1621, pp. 143-144).

Al describir los paradigmas de conjugación del español, utiliza los términos morfológicamente basados *perfecto simple* y *compuesto*, pero esto no conduce a un resultado razonable para el alemán, ya que simplemente adopta el término *perfecto simple* en latín y lo aplica a una forma verbal compuesta (Zumarán, 1621, p. 215). Sin embargo, no se trata ni de un *perfecto simple* ni de un *perfecto compuesto*, sino del *pluscuamperfecto*, al que dos páginas más adelante se le asigna también la misma forma (Imagen 2).

vosotros amáuades, aquellos amáuan. Perfecto simple, Sing. Yo amé, amaste, amó. Plur. Amamos, amastes, amaron. Perfecto Compuesto. Sing.	Perfectum simplex. Sing. Ich hatte geliebt/ du hattest geliebt/ der hatte geliebt. Plur. Wir hatten geliebt/ ihr hattet geliebt/ sie hatten geliebt. Perfectum Compositum. Sing.	Yo he amado, has amado, ha amado. Plur. Hemos vel hauemos amado, haueis amado, han amado. Mas que perfecto. Sing. Yo auia amado, tu auias amado, el auia amado.	Ich habe geliebt/ du hast geliebt/ der hat geliebt. Plur. Wir haben geliebt/ ihr habt geliebt/ sie haben geliebt. Plusquamperfectum. Sing. Ich hatte/ du hattest/ der hatte geliebt.
---	--	--	---

IMAGEN 2: Perfecto “simple” en Zumarán (1621, pp. 215-217)

Sin embargo, no se trata de criticar los términos utilizados por Zumarán. Su observación de que las dos formas perfectas diferentes, la simple y la compuesta, solo se dan en otras lenguas distintas del alemán se obvió, evidentemente, cuando elaboró los paradigmas de conjugación. Se trata de un problema que se trasladó de la gramática española a la alemana sin necesidad.

En el siglo XVII, las observaciones contrastivas solían basarse en la suposición de un carácter diferenciado de cada lengua. El término proviene del francés, *génie de la langue*, y posteriormente se tradujo con más o menos éxito a otros idiomas (*genio della lingua*, *genio de la lengua*, *Genius der Sprache*, *genius of language*).

Por supuesto, Zumarán no utiliza el término *génie de la langue* ni una traducción del mismo, pero sin duda hay intentos de conceptualización, por ejemplo, cuando se explica que un alemán que sabe francés aprendería italiano más rápido que un francés nativo porque la lengua francesa es blanda, mientras que la italiana es varonil como la alemana (Zumarán, 1621, pp. 45-46). Estas atribuciones de características a las distintas lenguas se remontan al Renacimiento y se resumieron en el concepto de *génie de la langue*. En la obra de Zumarán, sin embargo, tales planteamientos son muy escasos. Por el contrario, describe detalladamente la diferente representación de sonidos similares mediante letras, como en caso de los españoles que quieren aprender italiano (Zumarán, 1621, pp. 173-174): “En lugar de la k, se siruen destas dos letras ch, como los Españoles de la qu”.

6. CONCLUSIÓN

El enfoque onomasiológico de la formación de términos metalingüísticos se centra en procesos cognitivos que incluyen distintos aspectos de la formación de conceptos y también errores objetivos. Por tanto, el estudio onomasiológico no puede reducirse a una lista de términos auténticos para un concepto fijo. En el proceso de conceptualización, el agrupamiento basado en características también puede determinarse en una fase preconceptual, que puede

combinarse lingüísticamente con términos inespecíficos o parafraseados. Sin embargo, la conceptualización propiamente dicha no suele completarse con la definición de un término, ya que la estructura de relaciones en la que este se inscribe solo puede consolidarse o constituirse de nuevo en una fase posterior; incluso es posible un cambio en las características constitutivas de la conceptualización.

En el desarrollo de los conceptos lingüísticos intervienen los siguientes procesos: construcción de un concepto a partir de la tradición latina, diferenciación del concepto, juego de términos contrarios y opuestos, comparación con otra ciencia, trabajo metalingüístico sobre un concepto y sus nombres, uso para formar opiniones contrarias. La mayoría de los conceptos que se han estudiado utilizan estos procedimientos, a veces de forma abreviada o incompleta.

Al leer un texto histórico sobre lingüística o gramática, un análisis onomasiológico de los nombres de términos individuales en el horizonte temporal correspondiente puede ser útil para comprender el fondo y el estado de la discusión. De este modo, se puede determinar si el autor en cuestión estaba a la vanguardia, rezagado o incluso aportaba sugerencias innovadoras. Observar el término elegido en el campo onomasiológico de términos para el concepto estudiado puede proporcionar información sobre las particularidades de la conceptualización por parte del autor.

REFERENCIAS BIBLIOGRÁFICAS

- Abad Nebot, F. (1986). *Diccionario de lingüística de la escuela española*. Gredos.
- Abraham, W. (1981). *Diccionario de terminología lingüística actual*. Gredos.
- Alcaraz Varó, E. (2004). *Diccionario de lingüística moderna*. Ariel.
- Baldischwieler, Th. (2004). *Antonio de Nebrija, Las introducciones latinas contrapuesto el romance al latín (1486)* [Tesis doctoral. Philosophische Fakultät der Heinrich-Heine-Universität]. Publikationsservice. <https://n9.cl/q68fm>
- Benito Lobo, J. A. (1982). *Lengua española: glosario de términos lingüísticos*. Servicio de Publicaciones del Ministerio de Educación y Ciencia (ediciones 1983, 1986, 1987, 1990, 1991, 1992, 1993).

- Blank, A. & Koch, P. (eds.) (2003). *Kognitive romanische Onomasiologie und Semasiologie*. Niemeyer. <https://doi.org/10.1515/9783110911626>
- Calleja, J. M. (1818). *Elementos de gramática castellana*. Don Pedro Antonio de Apraiz.
- Cardona, G. R. (1991). *Diccionario de lingüística*. Edición española a cargo de M. T. Cabello. Ariel.
- Cerdà Massó, R. (coord.) (1986). *Diccionario de lingüística*. Anaya.
- Correas, G. (1626). *Arte grande de la lengua castellana compuesto en 1626 por el Maestro Gonzalo Correas*. Publicalo por primera vez el Conde de la Viñaza. Madrid Tip. R. Fé 1903.
- Corvo Sánchez, M. J. (2007). *Los libros de lenguas de Juan Ángel de Zumaran*. Lang.
- Coseriu, E. (1967). L'arbitraire du signe. Zur Spätgeschichte eines aristotelischen Begriffes. *Archiv für das Studium der neueren Sprachen und Literaturen*, 204, 81-112.
- Crystal, D. (1992). *An Encyclopedic Dictionary of Language & Languages*. Blackwell.
- Crystal, D. (2000). *Diccionario de lingüística y fonética*. Traducción y adaptación de X. Villalba. Octaedro.
- Dubois, J. (1979). *Diccionario de lingüística*; versión... [del francés] de I. Ortega y A. Domínguez; dirección y adaptación de A. Yllera. Alianza Editorial, D.L.
- Dubois, J. et al. (1994). *Dictionnaire de linguistique et des sciences du langage*. Larousse.
- Fernández López, J. (2018). *Diccionario de Lingüística y ciencias afines. Definiciones y explicaciones según diversos planteamientos teóricos* <https://n9.cl/imhrm>
- Fernández-Domínguez, J. (2019). The Onomasiological Approach. En M. Aronoff (coord.), *Oxford Research Encyclopedia of Linguistics* (pp. 1-28). Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.579>
- Fournier, J.-M. (2013). *Histoire des théories du temps dans les grammaires françaises*. ENS Éditions. <https://doi.org/10.4000/books.enseditions.4724>
- Glück, H. (1993). *Metzler Lexikon Sprache*. Metzler. <https://doi.org/10.1007/978-3-476-03486-1>
- Gómez Asencio, J. J. (1981). *Gramática y categorías verbales en la tradición española, 1771-1847*. Universidad de Salamanca (Acta Salamanticensia).
- Gómez Asencio, J. J. (2011). *Los principios de las gramáticas académicas (1771-1962)*. Peter Lang. <https://doi.org/10.3726/978-3-0352-0097-3>
- Gómez Carrillo, E. & Sola, A. (1925). *Diccionario ideológico para facilitar el trabajo literario y enriquecer el estilo*. Renacimiento.
- Grzega, J. (2015). Word-formation in onomasiology. En P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (eds.), *Word-formation: An international handbook of the languages of Europe* (pp. 79-93). De Gruyter. <https://doi.org/10.3726/978-3-0352-0097-3>
- Hallet, W. (2010). Methoden kulturwissenschaftlicher Ansätze: *Close Reading* und *Wide Reading*. En V. Nünning, A. Nünning & I. Bauder-Begerow (eds.), *Methoden der literatur- und kulturwissenschaftlichen Textanalyse* (pp. 293-315). Metzler. https://doi.org/10.1007/978-3-476-00205-1_14

- Hallig, R. & Wartburg, W. (1952). *Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas*. Akademie-Verlag.
- Haßler, G. (2014). "Inversions of word order generate higher costs". Continuity and development of a topos since the rationalist language theories of the 17th century. En V. Kasevich, Y. Kleiner & P. Sériot (eds.), *History of Linguistics 2011. Selected papers from the 12th International Conference on the History of the Language Sciences (ICHOLS XII)* (pp. 3-11). John Benjamins Publishing Company. <https://doi.org/10.1075/sihols.123.01has>
- Haßler, G. (2016). *Temporalität, Aspektualität und Modalität in romanischen Sprachen*. De Gruyter. <https://doi.org/10.1515/9783110312997>
- Haßler, G. (2019). El tratamiento del orden de las palabras en el pensamiento lingüístico de los siglos XVIII y XIX en España. En M. Díaz, G. Vaamonde, A. Varela, M. C. Cabeza, J. M. García-Miguel & F. Ramallo (coords.), *Actas do XIII Congreso Internacional de Lingüística Xeral, Vigo, 13-15 de xuño de 2018* (pp. 447-452). Servicio de publicacións Universidade de Vigo.
- Haßler, G. & Neis, C. (2009). *Lexikon sprachtheoretischer Grundbegriffe des 17. und 18. Jahrhunderts*. De Gruyter. <https://doi.org/10.1515/9783110212624>
- Instituto Nacional de Bachillerato a Distancia (1984). *Glosario de términos lingüísticos*. Servicio de Publicaciones del Ministerio de Educación y Ciencia.
- Jiménez Patón, B. (1614). *Instituciones de la gramatica española*. Pedro de la Cuesta.
- Knobloch, C. (1996). Problemgeschichte und Begriffsgeschichte. En H. E. Brekle, E. Dobnig-Jülch & H. Weiss (eds.), *A Science in the Making. The 1994 Regensburg Symposia on European Linguistic Historiography. Europe* (pp. 259-273). Nodus Publikationen.
- Koselleck, R. (1979). *Vergangene Zeiten. Zur Semantik geschichtlicher Zeiten*. Suhrkamp.
- Lewandowski, Th. (1982). *Diccionario de lingüística*, traductores M. L. García-Denche Navarro, E. Bernárdez. Cátedra.
- Martínez Gómez Gayoso, B. (1769). *Gramática de la lengua castellana reducida a breves reglas, facil método para instruccion de la Juventud: nuevamente añadida y emendada*. Gabriel Ramirez.
- Mounin, G. (1982). *Diccionario de lingüística*, dirigido por Georges Mounin, traductor Ricardo Pochtar; adaptador Eugenio Martínez Celdrán. Labor.
- Nebrija, A. (1980 [1492]). *Gramática de la lengua castellana*. Estudio y edición de A. Quilis. Editora Nacional.
- Nebrija, A. (1996 [1488]). *Introducciones latinas contrapuesto el romance al latín* (c. 1488). Ed. por M. Á. Esparza Torres & V. Calvo. Nodus Publikationen.
- Noboa, A. M. (1839). *Nueva Gramática de la lengua castellana, según los principios de la filosofía gramatical*. Eusebio Aguado.
- Ogden, Ch. K. & Richards, I. A. (1923). *The Meaning of Meaning. A Study of the Influence of Language upon Thought and of the Science of Symbolism*. Kegan Paul.

- Priscian (1855-1859). *Prisciani grammatici Caesariensis Institutionum grammaticarum libri XVIII*. Ex recensione Martini Herzii. Teubner.
- Quadri, B. (1952). *Aufgaben und Methoden der onomasiologischen Forschung. Eine entwicklungsgeschichtliche Darstellung*. A. Francke AG. Verlag.
- RAE = Real Academia Española (1771). *Gramática de la lengua castellana compuesta por la Real Academia Española*. Don Joaquín Ibarra.
- RAE = Real Academia Española (1796). *Gramática de la lengua castellana compuesta por la Real Academia Española*. Cuarta edición corregida y aumentada. La viuda de Don Joaquín Ibarra.
- Ridruejo, Emilio (1977). Notas romances en gramáticas latino-españoles del siglo XV, *Revista de Filología Española*, 59, 47-80. <https://doi.org/10.1515/9783110212624>
- Salvá, V. (1852). *Gramática de la lengua castellana según ahora se habla*. Novena edición. Garnier; José María Andrade.
- San Pedro, B. (1769). *Arte del Romance castellano, dispuesta segun sus principios generales i el uso de los mejores autores*. Tomo II. Benito Monfort, Impresor del colegio Andresiano.
- Sanford, J. (1611). Προπύλαιν, or an entrance to the Spanish tongue. Th. Haueland.
- Schönberger, A. (2008). *Die Ars minor des Aelius Donatus: lateinischer Text und kommentierte deutsche Übersetzung einer antiken Elementargrammatik aus dem 4. Jahrhundert nach Christus*. Valentia.
- Schönberger, A. (2009). *Die Ars maior des Aelius Donatus: lateinischer Text und kommentierte deutsche Übersetzung einer antiken Lateingrammatik des 4. Jahrhunderts für den fortgeschrittenen Anfängerunterricht*. Valentia.
- Villalón, C. (1558). *Gramática castellana: Arte breue y compendiosa para saber hablar y escreuir en la lengua Castellana congrua y decentemente*. Guillermo Simon.
- Villar, J. (1651). *Arte de la lengua española reducida a reglas y preceptos de rigurosa gramática*. Francisco Verengel.
- Vossler, K. (1919). *Französische Philologie*. (Wissenschaftliche Forschungsberichte. Geisteswissenschaftliche Reihe, 1). Perthes.
- Zauner, A. (1902). *Die romanischen Namen der Körperteile: Eine onomasiologische Studie* [Tesis doctoral]. Universität Erlangen.
- Zumarán, J. Á. (1621). *Das Neue Sprachbuch. Liure & unstruction pour apprendre les langues*. Apud Viduam Bergianam.



APROXIMACIÓN A LOS SENTIDOS METAFÓRICOS DE *GRANDE* Y *PEQUEÑO*: PROPUESTA LEXICOGRÁFICA

APPROACHE TO THE METAPHORICAL SENSES OF *GRANDE* AND *PEQUEÑO*: LEXICOGRAPHIC PROPOSAL

Carlos Ynduráin Pardo de Santayana
Universidad del Atlántico Medio
carlos.yndurain@pdi.atlanticomedio.es

RESUMEN

El objetivo de este artículo es analizar los valores semánticos que suelen asociarse con los adjetivos *grande* y *pequeño* más allá de sus sentidos puramente dimensionales (vinculados, en principio, con el tamaño de las entidades del mundo físico). En el apartado dedicado a la propuesta lexicográfica, dichos valores semánticos se separarán y agruparán de modo que, una vez ordenados y definidos, se muestren como posibles acepciones de un diccionario. A lo largo del artículo se pondrá el foco en mostrar el carácter polisémico de las palabras que nos ocupan y se tratará de explicar cuáles son los procesos cognitivos que subyacen al desarrollo de los sentidos metafóricos o metonímicos con que estas se relacionan habitualmente. Las decisiones de carácter lexicográfico tomadas posteriormente se pondrán en relación con ese primer análisis semántico; con ello se busca dotar de coherencia la propuesta y justificar, en la medida de lo posible, los pasos dados a la hora de elaborar las entradas de los vocablos *grande* y *pequeño*.

Palabras clave: adjetivos dimensionales, polisemia, metáfora cognitiva, tamaño.

ABSTRACT

The aim of this article is to analyze the semantic values that are usually associated with the Spanish adjectives *grande* and *pequeño* beyond their purely dimensional senses (linked, in principle, with the size of the entities of the physical world). In the part dedicated to the lexicographic proposal, these semantic values will be separated and grouped in such a way that, once ordered and defined, they will be shown as possible meanings of a dictionary. Throughout this article, the focus will be on showing the polysemic character of the concerning words and on explaining the cognitive processes that underlie the development of the metaphorical or metonymical senses to which they are usually related. The lexicographical decisions made later on will be related to this first semantic analysis; the aim is to provide coherence to the proposal and to justify, as far as possible, the steps taken in the elaboration of the entries for the words *grande* and *pequeño*.

Keywords: dimensional adjectives, polysemy, cognitive metaphor, size.

Recibido: 01-08-2024
Aceptado: 30-10-2024

DOI: <https://doi.org/10.17561/rilex.8.1.9118>



1. INTRODUCCIÓN

Resulta habitual que las palabras cuenten con más de un sentido convencional asociado a ellas. En el caso de los adjetivos dimensionales (*alto*, *bajo*, *profundo*, *estrecho*, etcétera) este carácter polisémico se manifiesta de manera especialmente acusada: más allá de sus significados prototípicos, vinculados con cuestiones relativas a las dimensiones físicas de los objetos, presentan toda una red semántica que los relaciona con conceptos de muy diversa índole (UBICACIÓN, CANTIDAD, CALIDAD, ESPIRITUALIDAD...). A nadie le resultan extrañas asociaciones como *nubes altas*, *altos pensamientos*, *ventas bajas*, *bajas pasiones*, *mentalidad estrecha* o *sentimientos profundos*.

El caso concreto que nos ocupa, el de los adjetivos –también dimensionales– *grande* y *pequeño*, no es una excepción: con la metáfora como mecanismo de multiplicación semántica, este par de antónimos polares ha desarrollado diversos sentidos alejados de lo estrictamente dimensional. Así, en *grandes ventas*, *un problema pequeño* o *un gran hombre* los adjetivos no se vinculan con el tamaño de las entidades de las que predicen una propiedad: las ventas son eventos que carecen de dimensión en el mundo físico, un problema es una entidad abstracta que no cuenta con un volumen medible y un *gran hombre* puede ser bajo y delgado.

A lo largo de este artículo mostraremos los sentidos que consideramos apropiado distinguir dentro de la red polisémica surgida en torno a las palabras que nos ocupan. A pesar de que en algunos casos podemos estar ante un continuo semántico, trataremos de perfilar una serie de categorías que nos permitan separar de la forma más transparente posible todos los usos no dimensionales de estos adjetivos. Además, mostraremos cuáles son los mecanismos cognitivos que subyacen a la aparición de estos sentidos metafóricos a partir de un único significado nuclear de carácter puramente dimensional. Nuestro objetivo es, por tanto, llevar a cabo un estudio de semántica léxica que –desde un punto de vista sincrónico– arroje luz sobre el significado de un determinado par de antónimos.

En la segunda parte del trabajo nos apoyaremos en el análisis realizado previamente para llevar a cabo una propuesta de carácter lexicográfico: los sentidos diferenciados deberán subdividirse o agruparse para encontrar acomodo dentro de un posible diccionario monolingüe de uso general. Explicaremos los motivos que nos llevarán a proponer una serie de acepciones como las más adecuadas para dar cuenta del significado de los vocablos *grande* y *pequeño* en una obra de consulta. Además, ordenaremos y definiremos estas acepciones a partir de criterios lingüísticos y lexicográficos que trataremos de hacer transparentes.

Entendemos que el análisis semántico y el trabajo lexicográfico son dos procesos complementarios y que la combinación de ambos nos permite abordar de manera efectiva el estudio del significado de dos palabras que presentan un marcado carácter polisémico.

2. EXTENSIONES SEMÁNTICAS

2.1. SENTIDO METONÍMICO

Los adjetivos denominados *dimensionales* hacen referencia, en su sentido prototípico, a las magnitudes de un objeto en el espacio: “En cuanto a los adjetivos dimensionales, el término *dimensión* denota principalmente las tres dimensiones espaciales de los objetos físicos, a saber, la largura/altitud, la anchura y el volumen o profundidad” (Mulier, 2009, p. 12).

El sentido nuclear de *grande/pequeño* se diferencia del de otros adjetivos dimensionales (como *alto/bajo*, *ancho/estrecho* o *profundo/somero/superficial*) en que este no se relaciona con una dimensión concreta de un objeto, sino con los conceptos holísticos de ÁREA o VOLUMEN (o, de manera más general e informal, TAMAÑO).

En ocasiones, como ocurre en los ejemplos 1, 2 y 3, los adjetivos *grande* y *pequeño* acompañan a sustantivos sin atribuir de forma directa un tamaño a la entidad con que se relacionan: así, en el caso de *grandes terratenientes* nos encontramos con que son las tierras que poseen los terratenientes aquello

que, en última instancia, se considera que es grande. A través de un proceso metonímico evidente, que lleva a identificar lo poseído con su poseedor, el adjetivo se aplica de manera directa a los terratenientes, aunque, en sentido estricto, no se esté predicando nada acerca del tamaño de dichas entidades:

- (1) El poder político residía en los militares y los grandes terratenientes alemanes¹.

Lo mismo puede decirse que ocurre en *pequeños empresarios*: a pesar de que en este caso el adjetivo dimensional pueda emplearse también para hacer referencia a la importancia de las empresas poseídas (y no a su tamaño, en sentido estricto), su aplicación directa sobre el sustantivo que nombra a los poseedores de las empresas puede entenderse como un proceso metonímico análogo al que se produce en el caso de los *grandes terratenientes* visto más arriba:

- (2) El desempleo, ha aumentado la inflación y ha arruinado a muchos pequeños empresarios.

De manera similar, los *grandes propietarios* son propietarios que cuentan con grandes cantidades o grandes extensiones de aquello que poseen. En esta metonimia media la relación entre TAMAÑO y CANTIDAD que analizaremos más adelante:

- (3) “Hay que tener en cuenta”, dijo, “que los grandes propietarios de las aguas no ven con agrado que el Ayuntamiento de Telde no sea controlado por la derecha”.

En los casos que hemos recogido (y en otros similares) podría interpretarse que los adjetivos dimensionales se aplican a las personas para señalar el poder de estas: los *grandes terratenientes/propietarios* serían terratenientes/propietarios importantes y los *pequeños empresarios* serían empresarios poco importantes. Desde esta perspectiva no estaríamos ya ante usos metonímicos: el adjetivo se aplicaría de forma directa al poseedor para predicar una

¹Tomamos todos los ejemplos de *Corpus del español del siglo XXI* (CORPES XXI) y del *Corpus de Referencia del Español Actual* (CREA).

propiedad de este, y no de aquello que este posee. Nos encontraríamos, como veremos más adelante, ante uno de los usos metafóricos fundamentales de los adjetivos dimensionales que nos ocupan: aquel que se relaciona el TAMAÑO con el concepto de IMPORTANCIA.

Desde nuestro punto de vista entendemos que el uso de *grande* y *pequeño* con los sentidos puramente metonímicos señalados (sin que medie la idea de IMPORTANCIA DEL POSEEDOR) constituye una primera extensión semántica convencionalizada de los adjetivos que nos ocupan. En la propuesta lexicográfica recogeremos esta idea como subacepciones² metonímicas de las acepciones prototípicas de *grande* y *pequeño*.

2.2. SENTIDO TEMPORAL

El TIEMPO es un concepto complejo al que solemos aproximarnos mediante metáforas espaciales: el mundo físico sirve de base cognitiva para entender, a través de lo tangible y concreto, aquello que resulta más abstracto y escurridizo.

Como apuntan Lakoff y Johnson (1999, p. 139), resulta virtualmente imposible conceptualizar el tiempo sin recurrir a metáforas. Sinha (2014, p. 184) entiende que “it is difficult, if not impossible, to think of and talk about time without employing metaphors, and many of these have as their source domain space and spatial motion”. Yu (1998, p. 84), en la misma línea, señala que “it seems that time cannot be approached directly or literally, without getting onto the vehicle of a spatial metaphor”.

Algunos autores consideran que, más que la propia conceptualización subjetiva del tiempo en sí, lo que lleva a recurrir a metáforas espaciotemporales es una necesidad comunicativa: “We need metaphors to speak about time in the same way that we need metaphors to speak about other internal states such as emotions or thoughts” (Radden, 2011, p. 2). Esta línea estaría

²Entendemos que una *subacepción* es una acepción que constituye un matiz especial de otra y que se registra como dependiente de esta.

en sintonía con lo expresado por San Agustín en sus *Confesiones* (2015, p. 18): “¿Qué es, pues, el tiempo? Si nadie me lo pregunta lo sé, pero si trato de explicárselo a quien me lo pregunta no lo sé”. Gale (1968, p. 4) señala al respecto que “Augustine’s perplexity is due to the fact that we both have an experiential awareness of time and know the correct use of temporal language but are mysteriously reduced to silence when we try to verbalize this understanding”.

Las metáforas más frecuentes mediante las que concebir y hacer referencia al TIEMPO son las de origen espacial, pero no parece haber consenso en lo que respecta a su posible carácter universal. Fauconnier y Turner (2008, p. 4) consideran que “time as space is a deep metaphor for all human beings. It is common across cultures, psychologically real, productive, and profoundly entrenched in thought and language”. Yu (1998, p. 86), mientras, entiende que todavía resulta necesario confirmar dicha hipótesis: “It seems that the time as space metaphor has a universal status. But this status has not received adequate support in the form of systematically worked-out evidence across languages”. Sinha (2014), por su parte, rechaza la universalidad de estas metáforas tomando como ejemplo la lengua amondawa (2014, p. 195): considera que a partir de ella puede llegarse a conclusiones muy similares a las que Whorf extrae del hopi (aunque sin las implicaciones deterministas de este):

our idea of space has also the property of acting as a surrogate of non-spacial relationships like time, intensity, tendency, and as a void to be filled with imagined formless items, one of which may even be called ‘space’. Space as sensed by the Hopi would not be connected mentally with surrogates, but would be comparatively pure, unmixed with extraneous notions (Whorf, 1941, p. 93).

En el caso del español es frecuente hacer referencia a conceptos temporales a través de palabras relacionadas, en principio, con lo espacial. Los adjetivos que con más frecuencia se vinculan con el TIEMPO son *largo* y *corto*: “Los usos temporales del par *largo/corto* se basan en la idea de que el tiempo es una línea sobre la que se sitúan los eventos” (Ynduráin, 2019, p. 36).

Los adjetivos *grande* y *pequeño* se suelen emplear acompañando a sustantivos que se vinculan con el tiempo de una forma especial: como se muestra en los ejemplos 4, 5 y 6, estos, en muchos casos, hacen referencia a eventos que suponen una especie de ruptura de la línea temporal normal.

- (4) Podemos hacer un pequeño descanso y aprovechar este momento para comer a la sombra de los pinos.
- (5) Una pausa grande porque Antígona no puede imaginarse qué es lo que Hemón puede haber contado.
- (6) Ciertos hábitos culturales favorecen los desplazamientos de fin de semana o de pequeñas vacaciones (Semanas de la Nieve, o de Carnaval, etc.).

Además, *grande* y *pequeño* (como se muestra en los ejemplos 7 y 8) se emplean también junto a nombres de fragmentaciones del tiempo (*rato*, *momento*, *temporada*, etc.) para señalar la duración de estas:

- (7) Habían estado todos ellos hacía un rato grande dando una vuelta.
- (8) Si algún compañero crea compañía y forma empresa eso es algo esporádico y por una pequeña temporada.

Las combinaciones *una gran fiesta* o de *una fiesta grande*, sin embargo, no activan un sentido temporal del adjetivo. Entendemos que el uso temporal de *grande* y *pequeño* se da exclusivamente cuando resulta poco probable una interpretación vinculada con los conceptos de IMPORTANCIA o CALIDAD (abordados en el apartado 2.4). Casos como *grandes vacaciones* pueden resultar ambiguos: resultan aceptables interpretaciones tanto temporales como relativas a la calidad (que se abordarán en el apartado 2.4). De modo similar, *un gran descanso* puede interpretarse también (como se verá en el apartado 2.5) como un descanso que es intenso. La desambiguación semántica puede apoyarse, además de en cuestiones pragmáticas, en factores sintácticos: los sentidos relativos a la IMPORTANCIA, la CALIDAD o la INTENSIDAD se ven potenciados si los adjetivos aparecen como adyacentes antepuestos. El sentido temporal que hemos abordado en este apartado resultará más accesible, por lo tanto, si los adjetivos *grande* y *pequeño* se

sitúan tras el nombre al que acompañan (*unas vacaciones pequeñas*) o si aparecen como atributo en oraciones copulativas (*las vacaciones no fueron muy grandes*).

2.3. SENTIDO RELATIVO A LA EDAD

En español resulta frecuente expresar que un niño es *pequeño* si este, independientemente de cuál sea su tamaño, es de corta edad. El uso de *grande* para indicar que una persona es adulta (o que un niño ya tiene cierta edad) es también habitual. El proceso cognitivo que subyace a esta relación entre conceptos se basa, como es evidente, en nuestro conocimiento del desarrollo de las personas y de muchos de los seres vivos que nos rodean.

Los adjetivos *mayor* y *menor* han experimentado el mismo proceso de extensión semántica: su significado primario se relaciona con el tamaño de las entidades, pero ambos, como ocurre en los ejemplos 9 y 10, suelen emplearse para hacer referencia a la edad de los seres vivos.

- (9) La casa estaba cerrada y las personas grandes aún consagradas al rito de la siesta.
- (10) Kendler ha encontrado que las ratas y los niños pequeños operan mejor bajo el paradigma del cambio no reversible que bajo el reversible, mientras los niños mayores y los universitarios realizan mejor los cambios reversibles.

2.4. SENTIDOS RELATIVOS A LAS IDEAS DE PODER, IMPORTANCIA Y CALIDAD

El TAMAÑO es un concepto que resulta muy productivo como base metafórica: Bosque (1985, p. 72) señala que, al menos entre las lenguas por él estudiadas, las metáforas que parten de la idea de tamaño son (“junto con las de cantidad”) las más frecuentes.

El PODER y la IMPORTANCIA son algunos de los conceptos fundamentales con los que se relaciona el TAMAÑO. Esta misma traslación semántica la encontramos también relacionada con el concepto ALTURA:

Social hierarchic relations are often constituted, confirmed, communicated, and talked about with reference to symbolic representations of spatial dimensions such as size and height, even if physical size of the ranked individuals or groups is not a determinant of the hierarchy (Schubert, 2005, p. 1).

En el caso de la ALTURA, el esquema de imagen³ subyacente es que el PODER y el CONTROL se identifican con la idea de ARRIBA y la sumisión con la idea de ABAJO (Lakoff y Johnson, 1980): el eje vertical sirve de base para una jerarquización de las relaciones, tanto desde un punto de vista abstracto como desde una perspectiva tan concreta como pueda ser la representación gráfica del organigrama de una empresa. Esta metáfora cognitiva resulta (interlingüísticamente) muy productiva. La siguiente cita muestra ejemplos del inglés que, en muchos casos, resultan fácilmente extrapolables al español (y a otras lenguas)⁴:

When we talk about power, we often use metaphors about up and down. Examples abound: When someone has a high status, or is up in the hierarchy, he or she has control over and can oversee others who have lower status. One can look up to those who rose to the height of their power or look down on underlings. When a picture of a hierarchy is drawn, the most powerful person is usually at the top, and the subordinates are drawn below. All these metaphors are cases of the “control is up, lack of control is down” metaphor (Lakoff, 1987; Lakoff & Johnson, 1980). In short, power is metaphorically described as a vertical dimension in physical space (Schubert, 2005, p. 1).

Al margen de que el PODER se identifique con lo que queda ARRIBA y que esto pueda presentar cierta conexión con el concepto TAMAÑO (pues, frecuentemente, las entidades altas⁵ son también entidades grandes), consideramos que el volumen general de los objetos físicos (es decir, su TAMAÑO) puede

³ Un esquema de imagen (*image schema* o *image schemata*) es una estructura cognitiva, consistente en un patrón o esquema básico preconceptual, que surge de la interacción directa con el mundo físico. Los esquemas de imagen sirven de base para conceptualizar otros patrones menos básicos del mundo físico, así como para estructurar metafóricamente conceptos abstractos: “They are preconceptual schematic structures that emerge from our bodily experience and that are constantly operating in our perceptual interaction, bodily movement through space, and physical manipulation of objects” (Yu, 1998, p. 24). Johnson (1987, p. XIV) define los esquemas de imagen del siguiente modo como “a recurring, dynamic pattern of our perceptual interactions and motor programs that give coherence and structure our experience” y señala que estos juegan un rol crucial “in what we take as meaningful and in how we reason” (Johnson, 1987, p. XXXVII).

⁴ Como señala Bosque (1985, p. 66), “nuestros *alta posición* o *alto cargo* se pueden traducir casi literalmente a lenguas muy diferentes”.

⁵ Las extensiones metafóricas de los adjetivos *alto* y *bajo* en lo que respecta a los conceptos de PODER e IMPORTANCIA se recogen en Ynduráin (en prensa).

también, por sí mismo, servir como base experiencial a partir de la que conceptualizar las relaciones de PODER. Esta conexión entre TAMAÑO y PODER está, desde nuestra perspectiva, mediada por el concepto FUERZA FÍSICA: “Physical size typically correlates with physical strength” (Lakoff & Johnson, 1980, p. 15).

Frecuentemente, “people who have social control over other are conceived and spoken of as *powerful*. An effective leader is *strong*, regardless of how much she can bench-press” (Fiske, 2004, p. 96). Parece evidente que existe una base natural y tal vez evolutiva para este tipo de conexiones: en la naturaleza la norma es “the bigger the more dangerous” (Freedman, 1979, p. 29). Entre las personas, especialmente durante la infancia y la adolescencia, el tamaño también resulta determinante a la hora de establecer relaciones de poder. Schwartz, Tesser y Powell (1982) señalan que los roles de padres e hijos contribuyen también a reforzar la asociación entre TAMAÑO y PODER. En la edad adulta el tamaño también puede resultar determinante en contextos en los que hay un enfrentamiento físico (deportes, violencia...) entre dos o más oponentes. Además, los grupos *grandes* de individuos (más adelante veremos que MÁS GRANDE suele identificarse con MÁS CANTIDAD) tienden a ser más poderosos que los grupos con pocos miembros.

En nuestras sociedades existe una relación objetiva entre TAMAÑO y PODER: esto sucede porque, curiosamente, existe cierta tendencia a dotar de poder a las personas altas (Judge & Cable, 2004) y porque las mujeres y los jóvenes suelen tener un acceso menor a las posiciones dominantes. Resulta especialmente llamativo el hecho de que, intuitivamente, la gente tienda a atribuir a las personas importantes una altura mayor de la que realmente tienen (Higham & Carment, 1992).

Además, el hecho de que exista también una relación (muchas veces real, pero también simbólica) entre TAMAÑO y CALIDAD (*lo grande es mejor*) lleva a que las cosas grandes acaben por rodear a las personas de alto estatus:

Magnitude in the conformation of authority rankings is also exhibited when people of high rank are accorded greater social space than people of lower rank. They have larger abodes, wider ritual zones, and more extensive domains. Executives have offices that consist of architecturally reified personal spaces, while lower functionaries may have cubicles and subordinate workers have no individual space of their own. Higher ranking executives are often entitled to bigger desks, bigger cars, and so forth. Moreover, their “personal” space is greater: They have more social volume (Fiske, 2004, p. 100).

A lo largo de este apartado hemos visto que la relación entre TAMAÑO y PODER tiene una base experiencial sólida y que parece estar profundamente arraigada en nuestro subconsciente. Esto, evidentemente, tiene su reflejo en el lenguaje: “Language links power and size in many metaphors” (Schubert, Waldzus & Giessner, 2009, p. 2). En el caso concreto que nos ocupa, el del español actual, resulta fácil encontrar conexiones como las que recogemos en los ejemplos 11 y 12:

- (11) El descabezamiento temporal expone la vulnerabilidad de los grandes ejecutivos.
- (12) Que yo sepa, las autoridades y los grandes políticos de la Villa y Corte han permanecido callados todo este tiempo.

El concepto de IMPORTANCIA discurre, en muchas ocasiones, paralelo al de PODER: una persona poderosa es siempre una persona importante. Esto seguramente haya llevado a que también el concepto IMPORTANCIA tienda a identificarse directamente con el TAMAÑO, como parece ocurrir en el ejemplo 13. Como vemos, algo puede ser considerado *importante* sin que deba entenderse necesariamente que, además, es *poderoso*. Esta distinción se hace especialmente evidente en el ejemplo 14, pues *grandes encuentros* podría sustituirse por *encuentros importantes*, pero no por *encuentros poderosos*:

- (13) Stalin no vaciló en manejar todos los recursos del Estado para amordazar a grandes y pequeños artistas cuyo único delito era su talento creador o su lucidez.
- (14) Juan Viñas permanece al pie del cañón de las transmisiones deportivas, brillando de forma sobresaliente en los grandes encuentros de hockey sobre patines.

Este vínculo de *grande* (y *pequeño*) con la IMPORTANCIA puede entenderse también a partir de la idea de que hay entidades que (independientemente de su tamaño físico) destacan, sobresalen y resultan eminentes⁶ dentro de un contexto determinado. Otras, mientras, pueden pasar casi desapercibidas y ser ignoradas. La conexión con el mundo físico y lo sensorial resulta evidente: de manera general, los objetos de gran volumen son inevitablemente detectados por nuestros sentidos (vista y tacto, fundamentalmente), pero los pequeños resultan menos perceptibles. Los primeros resultan relevantes, los segundos tienden a ser ignorados.

Grady, Oakley y Coulson (1999, pp. 134-174) consideran la relación entre TAMAÑO e IMPORTANCIA una metáfora primaria presente en todas las lenguas y culturas. Hermanson (2006), por su parte, aduce una muestra de esta relación metafórica a través de los siguientes casos de la lengua zulú:

Size [...] may be used metaphorically, with the complementary conceptual metaphors Big is Important and Small is Insignificant being invoked so that the nouns *inkunzi* (bull), *inkosi* (chief) refer to something big or somebody important; *intaba* (mountain) refers to something big, tall or high (Hermanson, 2006, p. 40).

Existe también una conexión metafórica entre el TAMAÑO y la CALIDAD de las entidades: se describe como *grande* aquello que se considera bueno. Este vínculo resulta fácil de entender si se tiene en cuenta el carácter positivo que se les atribuye habitualmente a los conceptos de PODER e IMPORTANCIA. Lakoff y Johnson (1980, p. 22) consideran que la identificación *bigger is better* es de origen cultural, aunque resulta coherente con el sistema metafórico primario (y universal) en el que se inserta: se relaciona armónicamente con *arriba es más* y *arriba es mejor*.

⁶ *Destacado, sobresaliente y eminente* son, de hecho, adjetivos que se relacionan tanto con las propiedades físicas y tangibles de los objetos como con aquellas relativas a la importancia de las entidades.

En el español actual no se da una identificación de *pequeño* con *malo*; sin embargo, el uso de *grande* con sentido positivo resulta muy frecuente, como vemos en los ejemplos 15, 16 y 17:

- (15) El curso pasado usted comentó que empezaba a darse cuenta del gran jugador que era en el Atlético. ¿En la selección es usted tan gran futbolista?
- (16) Orantes, pese a la derrota, cuajó una gran actuación.
- (17) De una buena novela acostumbra a salir una mala película y de una novela mediocre una gran película.

En estos usos de *grande* y *pequeño* consideramos que se predica de las entidades que estas poseen de forma marcada las cualidades positivas con que se asocian típicamente. En *Está siendo un gran año para la música*, mientras, no se destacan las cualidades típicas de los años: no se está ante un año de especial calidad, sino que se señala que un año está siendo especialmente beneficioso o positivo para (alguien o) algo. Esto ocurre también en el ejemplo 18:

- (18) Así que 2015 está siendo un gran año para la División de Ciencias Planetarias.

Bosque (1985, p. 72) apunta que en su estudio comparativo no ha encontrado “ningún idioma en el que la palabra *grande* no signifique también figuradamente ‘importante’, ‘notable’, ‘destacado’, ‘insigne’ u otro valor parecido”. Aunque todas estas ideas tiendan a mezclarse en nuestro sistema conceptual, creemos que, a lo largo de este apartado, hemos distinguido de forma clara cuatro sentidos de *grande* (y dos de *pequeño*) presentes de manera habitual en el español actual. En la propuesta lexicográfica abordaremos la cuestión de si estos sentidos (distintos, pero, como hemos visto, interrelacionados) deben agruparse o separarse en las distintas acepciones de un diccionario.

En lo sintáctico, debemos destacar que cuando se busca activar los sentidos que recogemos en este apartado el adjetivo suele anteponerse al nombre: resulta más probable que interpretemos como excelentes unas vacaciones que se describen como *unas grandes vacaciones* que aquellas descritas como *vacaciones grandes*. En el segundo caso aparece más claramente la

posibilidad de activar un sentido temporal del adjetivo. Lo mismo ocurre entre *un gran niño* y *un niño grande*: con el adjetivo pospuesto son más probables interpretaciones relativas a la dimensión o a la edad del niño. La aparición del adjetivo como atributo (*Las vacaciones fueron grandes*, *El niño es grande...*) también hace menos probables las interpretaciones relativas a los conceptos de PODER, IMPORTANCIA O CUALIDAD.

2.5. SENTIDOS RELATIVOS A LAS IDEAS DE CANTIDAD, FRECUENCIA E INTENSIDAD

Expresiones como *pequeñas dosis*, *gran tensión* o *grandes ventas* son una clara muestra de cómo los adjetivos que nos ocupan pueden relacionarse con nombres que hacen referencia a realidades que, en sentido estricto, no presentan un tamaño físico y que, por lo tanto, no son susceptibles de ser calificadas dimensionalmente.

Lo que subyace en estos casos es la asociación de *grande* con MUCHO y de *pequeño* con POCO: la base experiencial de este vínculo se halla en que el volumen que ocupan los objetos acumulados es mayor cuando estos son muchos que cuando son pocos. Además, la masa de una sustancia suele hacerse también más voluminosa cuando la cantidad de esta aumenta⁷.

En español resulta frecuente el vínculo de *grande* y *pequeño* con nombres que, como ocurre en los ejemplos 18 y 19, hacen referencia a propiedades: *intensidad*, *resolución*, *mortalidad*, *cantidad*, etc. En estos casos los adjetivos indican el grado⁸ (mayor o menor) en que se dan dichas propiedades: en *gran intensidad* entendemos que hay *mucha intensidad*, es decir, que la

⁷ Un hecho similar subyace también a la interpretación cuantitativa de los adjetivos *alto* y *bajo* (Ynduráin, 2020): “If you add more of a substance or of physical objects to a container or pile, the level goes up” (Lakoff & Johnson, 1980, p. 16).

⁸ Hacemos referencia al concepto de grado por ser precisamente su graduabilidad una de las características básicas de los adjetivos que nos ocupan: “Adjectives are gradable if they denote a property that can be present in an object to a greater or lesser degree” (Tribushina, 2008, p. 78). Aunque habitualmente los grados se vinculan con el concepto ESCALA y, por lo tanto, se proyectan sobre el eje vertical, el hecho de que podamos hablar de *un grado mayor o menor* muestra que, al menos en español, una identificación con el TAMAÑO resulta también posible.

intensidad es una propiedad que está muy presente en aquello a lo que se le atribuye.

- (19) La radiación solar es ahora de pequeña intensidad, careciendo de energía suficiente para provocar la rotura de estas moléculas.
- (20) Mi opinión es que la gran mortalidad actual de los gatos de Copenhague se debe precisamente a la saturación de electricidad que hay en la atmósfera.

Otros tipos de entidades son también susceptibles de describirse mediante los adjetivos *grande* y *pequeño* si se busca señalar que alguna de sus propiedades típicas se da de forma más o menos intensa. Así, en *un sonido grande* podemos entender que su volumen es la propiedad que presenta especial intensidad⁹. Algo muy similar ocurre en los ejemplos 21 y 22:

- (21) Cierra con gran estruendo la puerta metálica de la celda.
- (22) Marco finge que duerme, escucha cada pequeño sonido a través del pladur.

Además, en español “los adjetivos de tamaño expresan cuantificación de grado con sustantivos que se usan también como adjetivos, como en *un embustero colosal* o en *eres un gran tonto*” (NGLEM, 2010, p. 924). En estos casos los adjetivos (habitualmente antepuestos) ponen el foco en las propiedades típicas de las entidades: en el ejemplo 23 estamos ante un *gran tonto*, es decir, una persona que presenta las propiedades típicas que lo convierten en tonto de manera especialmente acentuada.

- (23) Y has perdido, ¡gran tonto!, la ocasión de rogarme, por favor, que no me marchase.

Tal y como ocurre con el uso de los diminutivos, la atribución de un tamaño pequeño a las entidades puede tener también como fin la expresión de cierto afecto hacia estas (Ivanov, 1979, p. 158). Habitualmente, tal y como ocurre en el ejemplo 24, dicha muestra de afecto contrasta con el carácter negativo del sustantivo mediante el que se alude al referente:

⁹ Muchos de los casos se sitúan a medio camino entre una sinestesia creativa y el uso convencional relativo a la INTENSIDAD que describimos.

- (24) Eres un pequeño cabrón —suspiró ella, tendida en el sofá—. Un pequeño cabrón.

Aunque no es tan habitual, los adjetivos *pequeño* y, sobre todo, *grande* pueden relacionarse también con sustantivos eventivos vinculando el concepto de TAMAÑO con el de FRECUENCIA/CANTIDAD. Mediante estos adjetivos se hace referencia a la mayor o menor repetición de un evento: *grandes ventas* (ejemplo 25) sería equivalente a *muchas ventas*.

- (25) Alcanza grandes ventas un ensayo que define la sinceridad como el arte de no intentar explicarse.

Además, es frecuente también la asociación de *grande* y *pequeño* con el sustantivo *número*. En muchos de estos casos, entendemos que *número* actúa como sinónimo de *cantidad* (ejemplo 26): *un número pequeño de pacientes* sería equivalente a *una pequeña cantidad de pacientes* o, sencillamente, a *pocos pacientes*. Sin embargo, como muestra el ejemplo 27, también resulta posible la identificación metafórica directa entre los conceptos de TAMAÑO y de VALOR NUMÉRICO:

- (26) Montar servicios de Medicina Nuclear en centros con un número pequeño de pacientes no es rentable.
- (27) De alguna manera, no sólo somos capaces de percibir claramente que un número pequeño sea par, sino en general si es compuesto o no.

En los usos de *grande* y *pequeño* que hemos visto en este apartado los adjetivos suelen aparecer como adjuntos antepuestos, aunque es habitual la posposición y la posición predicativa en los casos que relacionábamos con el valor de los números (*un número pequeño*, *los números son grandes*) y con la intensidad que presentan las propiedades de las entidades (*estruendo grande*, *el estruendo fue grande*).

Cuando el adjetivo (antepuesto) se relaciona directamente con nombres de propiedades (*gran fuerza*, *gran resolución*) apenas hay lugar para ambigüedades entre interpretaciones relativas a la INTENSIDAD y a la IMPORTANCIA. Si el adjetivo acompaña a *ventas* (como ocurre en el ejemplo 28), sin embargo,

si debemos recurrir al contexto para determinar si se trata de muchas/pocas ventas o de ventas de mucha/escasa importancia:

- (28) ¿Por qué no aceptas mi ayuda y empiezas con pequeñas ventas en La Habana, Trinidad y Santiago?

2.6. OTROS SENTIDOS

Podemos distinguir un uso de carácter irónico bastante asentado en el español peninsular actual: aquel en el que se dice de alguien que es *grande* cuando resulta “gracioso por absurdo o extravagante [...]: *Es grande este hombre, ahora se le ocurre que empecemos por el final*” (Becerra Hiraldo, 2000, pp. 392-393).

Además, *grande* puede emplearse también como sustantivo: en estos usos el TAMAÑO se vincula de nuevo con la IMPORTANCIA o la CALIDAD. Así, *un grande* se entiende que es una persona de alta posición social. Nuestra propuesta lexicográfica se ciñe a los usos de *grande* y *pequeño* como adjetivos calificativos, por lo que dejaremos este caso fuera de los siguientes apartados.

3. EXPLICACIÓN DE LA PROPUESTA LEXICOGRÁFICA

En este apartado tratamos de mostrar cuáles han sido las cuestiones lingüísticas que subyacen a las decisiones lexicográficas (micro y macroestructurales) adoptadas en la propuesta lexicográfica. Nuestro objetivo es, por lo tanto, hacer explícito cómo se ha tratado de conjugar la transparencia requerida por un diccionario con la complejidad de algunos de los conceptos manejados en los análisis semánticos de la primera parte (apartado 2) del trabajo.

3.1. SEPARACIÓN Y AGRUPAMIENTO DE LAS ACEPCIONES

Más allá de la lexicografía, determinar cuáles son los diferentes significados que se pueden relacionar con una palabra es una cuestión todavía no resuelta en la lingüística actual: “The trouble with word sense disambiguation is word senses” (Kilgarriff, 2006, p. 29). De hecho, puede afirmarse que “there is very little agreement about what word senses are” (Atkins & Rundell, 2008, p. 269).

La lexicografía, como consecuencia, tampoco dispone de unos criterios objetivos y estandarizados mediante los que separar o agrupar las acepciones: “After centuries of practical lexicography, there is still hardly any consensus on how to divide the semantics space of a lexical item” (Van der Meer, 2006, p. 604).

Parece indudable que los sentidos que hemos identificado en el apartado 2 de este trabajo deben recogerse en un diccionario al margen del significado prototípico de *grande* y *pequeño*. Sin embargo, las extensiones metafóricas a las que nos enfrentamos constituyen muchas veces un continuo semántico que la elaboración de una propuesta lexicográfica obliga a romper a través de un listado de acepciones discretas y mutuamente excluyentes: “Dictionaries greatly exaggerate the measure of discreteness of meanings, and are inclined to set clear-cut borders where a closer examination... reveals only a vague intermediate area of overlapping meanings” (Apresjan, 1973, p. 9).

Por otro lado, la creación de una acepción supone agrupar dentro de una misma categoría sentidos que podrían mantenerse aislados en forma de múltiples (tal vez infinitas) microacepciones; establecer cuáles de los posibles significados asociados con una forma léxica deben aparecer listados en una entrada de diccionario supone una generalización en la que se desecha aquello que resulta excesivamente coyuntural y se opta por lo más estandarizado: “dictionaries generalize (that is their job), and from the infinite number of individual situations in which a word appears, lexicographers derive a finite set of Lexical Units which collectively explain how that word contributes to the meaning of all the individual events” (Atkins & Rundell, 2008, p. 272).

Mediante el uso de las metáforas relativas al TAMAÑO los hablantes tratan, en muchos casos, de evocar una idea general que no acaba de identificarse de forma precisa con un sentido concreto y claramente delimitado. El concepto de TAMAÑO presenta, además, numerosas implicaciones susceptibles de ser evocadas (por separado o de forma conjunta) en cada contexto específico de uso. Puede decirse, por lo tanto, que nuestra labor a la hora de

redactar las entradas de *grande* y *pequeño* ha consistido (simultáneamente) en agrupar y aislar los sentidos del espectro semántico surgido en torno a estas palabras. En los apartados 3.3 y 3.4 trataremos de hacer explícitos los factores lingüísticos (semánticos y pragmáticos) que motivan cada una de las decisiones lexicográficas adoptadas.

3.2. LAS DEFINICIONES

Las acepciones de nuestra propuesta lexicográfica presentan definiciones que intentan conjugar la necesaria claridad lexicográfica con las explicaciones teóricas recogidas en la primera mitad del trabajo.

Las definiciones que se emplean responden a aquello que Seco (2003, p. 35) denomina *definiciones impropias*, ya que estas “no son definiciones en metalengua del contenido, sino en metalengua del signo”. Las definiciones, en estos casos, aportan información sobre el uso de las palabras y son, según el criterio de Seco (2003, p. 34), más explicaciones que auténticas definiciones¹⁰.

En las definiciones recurrimos a *aquello* cuando consideramos apropiado no restringir en exceso el tipo de entidades (objetos humanos, objetos no humanos, eventos, propiedades, etc.) con que se puede relacionar el adjetivo: esta referencia ontológicamente abierta aporta flexibilidad a las definiciones sin limitar su valor explicativo. Con el mismo objetivo, y para evitar repetir la palabra *aquello*, también hemos recurrido a subordinadas sustantivas introducidas por *lo que*. En algunas definiciones sí hemos considerado apropiado hacer referencias explícitas a tipos de entidades susceptibles de ser calificadas como *grandes* o *pequeñas* (con un sentido determinado): “seres vivos”, “propiedades”.

En ocasiones empleamos referencias a “lo normal o adecuado”. La referencia a “lo normal” muestra que los adjetivos que expresan propiedades

¹⁰ Seco señala que esta clase de definiciones ‘irregulares’ son frecuentes “en un amplio sector de la lexicografía, dentro del cual figuran todos los diccionarios españoles, encabezados por el de la Academia” (Seco, 2003, p. 34).

graduables suelen emplearse cuando estas se dan en un grado relevantemente mayor o menor que el que se considera propio del prototipo de aquella categoría que actúe como clase de comparación¹¹: “Los adjetivos graduables requieren algún punto de referencia, aunque no esté explicitado. Si decimos de alguien que es *delgado*, lo estamos comparando de algún modo con alguna norma vigente en la comunidad o en un determinado grupo” (Di Tullio, 2010 [2005], p. 193). Con “lo adecuado” recogemos la posibilidad de que una determinada referencia actúe como norma en un contexto dado: Ebeling y Gelman (1994, p. 1179) señalan, acerca del uso de los adjetivos *big* y *small*, que, en una de sus posibles interpretaciones, “the object is judged with regard to its intended use”. Así, un enunciado como *The bat is big* puede tener, entre otras, una interpretación de carácter funcional: “e.g. too big for a tiny doll”. En las definiciones optamos por evitar referencias explícitas a las ideas y conceptos señalados: mediante “lo normal o adecuado” hacemos llegar esta información apoyándonos de forma natural en el conocimiento intuitivo que poseen los hablantes respecto al funcionamiento de los adjetivos que presentan un valor graduable.

Los ejemplos empleados para acompañar las definiciones de la propuesta lexicográfica han sido adaptados (siempre que ha resultado posible) a partir de casos registrados en el *Corpus de Referencia del Español Actual* y en *Corpes*. Con ellos se ha buscado facilitar la comprensión de las definiciones

¹¹ Para saber qué magnitudes se deben asociar con un adjetivo resulta imprescindible poner en relación la entidad descrita por este con una determinada categoría de cosas. La categoría concreta respecto a la que se debe relativizar el valor de un adjetivo relativo es lo que tradicionalmente se ha denominado *comparison class*: “A comparison class is a subset of the universe of discourse which is picked out relative to a context of use” (Klein, 1980, p. 13). Consideramos que el prototipo de una categoría es la entidad que actúa como referencia de esta y permite establecer si sus elementos son, por ejemplo, *grandes*, *pequeños* o *medianos*. No cualquier elemento distinto del prototipo debe ser necesariamente considerado susceptible de describirse mediante un adjetivo dimensional: en torno a este prototipo existe una zona neutra que recoge todo elemento dotado de la propiedad que se esté evocando en un grado irrelevantemente distinto al grado en que esta propiedad se presente en ese prototipo. Solo se emplearán los adjetivos si se está ante una diferencia que se entienda como relevante (Ynduráin, 2020, pp. 201-205).

y situar dentro de unos límites precisos algunos usos cuya ubicación pudiera resultar poco intuitiva.

3.3. EL ARTÍCULO *GRANDE*

En la entrada de *grande* recogemos primero el sentido dimensional prototípico mediante una definición que subsume también el sentido temporal al que hacíamos referencia en el apartado 2.2. La concepción del tiempo a partir de metáforas de origen espacial es un recurso cognitivo profundamente arraigado en los hablantes: entendemos que estos pueden aceptar con total naturalidad que las ideas de DIMENSIÓN FÍSICA y de DURACIÓN TEMPORAL se recojan de manera conjunta.

La primera acepción cuenta con una subacepción en la que se recoge el sentido al que hacíamos referencia en el apartado 2.1. Introducimos la definición con la fórmula “por extensión” para, mediante un lenguaje asequible, mostrar la idea de que se está ante un uso metonímico del adjetivo. Aunque esta metonimia consiste habitualmente en una referencia al poseedor para señalar el tamaño de aquello que este posee, dejamos abiertas otras posibilidades mediante el uso de “lo que se relaciona con...”.

En la segunda acepción se recoge el uso vinculado con la edad al que hacíamos referencia en el apartado 2.3. Situamos esta acepción en segundo lugar debido a que esta extensión de significado mantiene todavía una conexión estrecha y transparente con lo estrictamente dimensional.

En la tercera acepción abandonamos la referencia a “lo normal o adecuado” debido a que puede resultar algo confusa en el contexto de la IMPORTANCIA de las entidades. Además, a través del verbo *destacar*, se mantiene, creemos, cierta conexión con la idea de PROMINENCIA FÍSICA que sirve (al menos en parte) como base del uso metafórico que nos ocupa. Consideramos, por último, que con “relevancia” recogemos de un modo general y amplio las ideas de PODER e IMPORTANCIA analizadas en el apartado 2.4.

La cuarta acepción recoge un uso al que también hacíamos referencia en el apartado 2.4: el relativo al concepto de CALIDAD. Aunque este uso presente

un vínculo estrecho con los usos relativos a la IMPORTANCIA, creemos que es pertinente recogerlo en una acepción independiente. Esta acepción cuenta, además, con una subacepción en la que se contempla la posibilidad de que el adjetivo no se relacione exactamente con las cualidades típicas de las entidades, sino que lo haga con la presencia de cierto valor de carácter moral. Cuando el elemento adjetivado es, por ejemplo, *escritor*, parece evidente que con *gran escritor* se predica que alguien escribe bien; en *gran persona*, sin embargo, las “cualidades típicas” resultan algo más escurridizas como recurso explicativo y se hace necesario activar una idea de GRANDEZA vinculada con lo MORAL¹². La sutileza de la distinción nos lleva a incluir el segundo caso (*gran persona*) como una suerte de especialización del primero (*gran escritor*).

En la quinta acepción se recoge un caso que contemplábamos también en el apartado 2.4: aquel en que se indica que algo es bueno debido a lo “beneficioso” que resulta para alguien o algo. En la definición obviamos la referencia directa al hecho de que sea necesaria una entidad que reciba unos efectos positivos: creemos que está implícito en la palabra *beneficioso* y que, además, se muestra de forma clara en los ejemplos.

En la sexta acepción recogemos el uso irónico al que hacíamos referencia en el apartado 2.6. No hemos considerado apropiado señalar que se aplica exclusivamente a personas y dejamos abierta la posibilidad de que *grande* pueda emplearse con el mismo sentido para calificar también otro tipo de entidades. Los ejemplos recogen un uso vinculado a las personas que hemos tomado de Becerra Hiraldo (2000, pp. 392-393) y un uso en el que se califica una acción.

La séptima acepción recoge el caso en que *grande* se relaciona con el grado o intensidad con que se manifiesta una propiedad. En la definición evitamos la palabra *intensidad*, ya que resultaría redundante a la hora de explicar la asociación *gran intensidad*.

¹² Becerra Hiraldo considera que en la asociación sintagmática *Grande de España* el concepto de GRANDEZA se relaciona con las ideas de “bondad y excelencia moral” (2000, p. 393).

La subacepción 7b también está relacionada con el concepto de INTENSIDAD, pero el adjetivo, en este caso, no se aplica directamente a propiedades, sino a otras clases de entidades (“aquello”, en la definición). En este caso aquello a lo que hace referencia el nombre presenta una propiedad relevante en un grado elevado: así, un *gran estúpido* presenta un elevado grado de ESTUPIDEZ y un *gran sonido* presenta un grado alto de VOLUMEN. La propiedad en cuestión debe ser inferida pragmáticamente, pero este es un dato que no consideramos pertinente incluir en la definición: “alguna de sus propiedades típicas” hace llegar al hablante la idea de que las propiedades que resulte adecuado activar en cada caso concreto deberán ser seleccionadas a partir del contexto comunicativo.

En la acepción número 8 recogemos un uso de *grande* relativo a la CANTIDAD. En los ejemplos aparecen las asociaciones *gran número* y *grandes ventas*. La primera podría haberse incluido dentro de la acepción 7, pues *número*, en este contexto, es un sinónimo de *cantidad* (y el concepto CANTIDAD puede entenderse como una propiedad). El caso relativo a las ventas podría haberse incluido en 7b, considerando el VOLUMEN DE VENTAS también como una propiedad típica de las ventas que puede darse en un mayor o menor grado. Hemos considerado más oportuno, sin embargo, que los casos relacionados con la ABUNDANCIA aparezcan en una acepción independiente, debido a que su relación con las propiedades resulta menos evidente.

La última acepción se relaciona con el VALOR de los números, aunque hemos optado por no cerrar la posibilidad de que *grande* se emplee con un sentido similar al vincularse con el nombre de otro tipo de entidades.

3.4. EL ARTÍCULO *PEQUEÑO*

La primera acepción de *pequeño* se opone a la primera acepción de *grande*. También las primeras subacepciones se oponen. Cuando ha resultado posible, hemos tratado de redactar definiciones (casi) idénticas para ambos adjetivos. En la segunda acepción oponemos “primeras etapas de su desarrollo” a “edad adulta” (*grande* 2). La tercera acepción es aquella relativa a la RELEVANCIA: se opone a la tercera acepción de *grande*. Las acepciones 5, 6 y 7 (relacionadas

con la INTENSIDAD, la ABUNDANCIA y el VALOR) se oponen, respectivamente, a las acepciones 7, 8 y 9 de *grande*.

No encontramos para *pequeño* sentidos análogos a los de las acepciones cuarta y quinta de *grande*: no son posibles asociaciones en las que el adjetivo tenga por objeto señalar la escasa calidad de una entidad o indicar que algo resulta perjudicial. Tampoco encontramos para *pequeño* el sentido moral que recogíamos en *grande* 4b.

La cuarta acepción de *pequeño* es aquella que presenta ciertas analogías con el uso del diminutivo: antepuesto a un sustantivo se emplea para mostrar cierto afecto por la entidad a la que este hace referencia. No hemos considerado oportuno recoger en la definición el contraste que suele producirse entre la expresión de afecto que supone el uso del adjetivo y el carácter negativo del sustantivo con el que este se relaciona. El ejemplo que acompaña a la definición sí muestra dicho contraste.

4. PROPUESTA LEXICOGRÁFICA

Recogemos a continuación nuestra propuesta lexicográfica. En ella aparecen las acepciones de los adjetivos *grande* y *pequeño* que hemos considerado pertinente distinguir a partir de las cuestiones lingüísticas y pragmáticas abordadas en el apartado 2.

***Grande* (adjetivo):**

1. Se aplica a aquello que presenta un tamaño o una duración superior a lo normal o adecuado. *Prefieren una casa grande en la que cada uno pueda tener su espacio. Se quedaron hablando un rato grande.*
- 1b. Por extensión, se aplica también a lo que se relaciona con aquello que presenta un tamaño o duración superior a lo normal o adecuado. *La agricultura se convirtió en una opción interesante para los grandes terratenientes.*
2. Se aplica a los seres vivos que han alcanzado la edad adulta. *Los niños jugaban mientras la gente grande dormía la siesta.*

3. Se aplica aquello que destaca por su relevancia. *El pequeño festival ha acabado por convertirse en uno de los grandes eventos del verano. Esta ermita es uno de los más grandes representantes del estilo.*
4. Se aplica a aquello que presenta de forma destacada las cualidades típicas con que se asocia. *Si continúa practicando llegará a convertirse en un gran guitarrista.*
 - 4b. Se aplica a aquello que es especialmente bueno en el sentido moral. *Es una gran persona y todo el mundo le adora.*
5. Se aplica a aquello que resulta especialmente beneficioso. *Fue un gran año para la industria discográfica.*
6. Se aplica a aquello que resulta gracioso por absurdo o extravagante. *Es grande este hombre: ahora se le ocurre que empecemos por el final. ¡Qué grande tu intervención de ayer!*
7. Se aplica a propiedades cuando estas se dan en un grado superior a lo normal o adecuado. *Se produjo un terremoto de gran intensidad. Los nuevos misiles destacaban por su gran alcance.*
 - 7b. Por extensión, se aplica a aquello que presenta alguna de sus propiedades típicas en un grado superior a lo normal o adecuado. *Un gran estruendo nos despertó. Uno puede quedar indefenso ante un gran idiota.*
8. Se aplica a aquello que es más abundante de lo normal o adecuado. *El sector despegó impulsado por las grandes ventas de vehículos en Asia. Las autoridades intervinieron ante el gran número de afectados.*
9. Se aplica a aquello que presenta un valor elevado. *Los cálculos son más difíciles con números grandes.*

Pequeño (adjetivo):

1. Se aplica a aquello que presenta un tamaño o una duración inferior a lo normal o adecuado. *Prefieren una casa pequeña ahora que sus hijos se han ido. Hicimos una pequeña pausa para descansar.*

- 1b. Por extensión, se aplica también a lo que se relaciona con aquello que presenta un tamaño o duración inferior a lo normal o adecuado. *Los pequeños terratenientes se unieron frente al gobierno.*
2. Se aplica a los seres vivos que están en las primeras etapas de su desarrollo. *Son programas destinados a los más pequeños de la casa.*
3. Se aplica aquello que tiene escasa relevancia. *Los grupos de música pequeños ya no encuentran hueco en los festivales de verano. El escenario se llenó con los pequeños políticos de la zona.*
4. Se emplea para mostrar cierto afecto por alguien o algo. *Eres un pequeño cabrón —suspiró ella, tendida en el sofá.*
5. Se aplica a propiedades cuando estas se dan en un grado inferior a lo normal o adecuado. *Para el Parlamento, fue una importante victoria, aunque de pequeño alcance. Su misión es apoyar en operaciones de pequeña intensidad.*
- 5b. Por extensión, se aplica a aquello que presenta alguna de sus propiedades típicas en un grado inferior a lo normal o adecuado. *Respondió con un pequeño grito de guerra. ¿Mi pequeña bofetada te ha hecho daño?*
6. Se aplica a aquello que es menos abundante de lo normal o adecuado. *La condena final fueron las pequeñas ventas de su negocio.* (Se opone a grande 8)
7. Se aplica a aquello que presenta un valor bajo. *Los cálculos resultaban más sencillos con números pequeños.*

5. CONCLUSIÓN

Nuestro sistema conceptual tiene en su base elementos primarios de naturaleza espacial: “The conceptual system is founded on spatial information” (Mandler, 2010, p. 22). El contacto con el mundo físico supone, pues, el punto de partida para el desarrollo de estructuras más complejas: “Spatio-physical experience provides much of the fundamental semantic (or

conceptual) structure from which other concepts are constructed” (Tyler & Evans, 2003, p. 24).

El mecanismo que permite emplear los conceptos físicos, concretos y tangibles para asimilar aquello que resulta abstracto y escurridizo es la metáfora cognitiva: “The contemporary theory of metaphor claims that abstract concepts are at least in part understood and expressed metaphorically in spatial terms and that abstract reason is achieved by using certain mechanisms for the perception of spatial relations” (Yu, 1998, p. 3).

Este proceso de naturaleza mental tiene su reflejo en el lenguaje: resulta evidente que “concrete meanings serve as structural templates to denote more abstract meanings” (Heine, 1997, p. 36). Dentro del tipo de palabras que nos ocupa, este fenómeno se muestra de manera especialmente transparente: el hecho de que adjetivos dimensionales como *grande* y *pequeño* puedan emplearse con diversos sentidos no es un simple accidente; no es un fenómeno de homonimia resultante de la convergencia azarosa de étimos no relacionados, sino un claro proceso metafórico en el que lo concreto, tangible y primario ha actuado como soporte para el desarrollo de significados de carácter abstracto.

Nuestro análisis trata de hacer comprensibles las bases experienciales en que se apoya la red polisémica surgida en torno a *grande* y *pequeño*: el mundo físico observable como punto de partida para un complejo entramado en el que participan conceptos que no siempre presentan una vinculación real (o, al menos, directa e intuitiva) con lo dimensional.

Tras abordar de forma teórica los valores semánticos vinculados con el par de antónimos en que se centra nuestro estudio, hemos pasado a distinguir una serie de acepciones con las que trasladar lo puramente lingüístico a la práctica lexicográfica. Ibarretxe-Antuñano (2010) considera que la linealidad en la disposición de las acepciones en los diccionarios tradicionales es una herencia del estructuralismo en la que no se tiene en cuenta “su estructura interna ni las posibles relaciones entre las mismas” (Ibarretxe-Antuñano, 2010, p. 197). Esta autora aboga por una nueva disposición espacial acorde

con las bases teóricas de la lingüística cognitiva, en la que las distintas acepciones se distribuyan en torno a un significado central prototípico del que se deriven, tejiendo una estructura como de tela de araña (una *red polisémica radial*), los otros significados que se haya convenido diferenciar.

Otros autores, como Geeraerts (2007, p. 1169), ahondan en esta necesidad de dotar a los diccionarios de cierta transparencia a la hora de recoger los mecanismos que subyacen a las redes polisémicas que constituyen las distintas acepciones y subacepciones:

Swanpoel (1992, 1998) and Van der Meer (2000), for instance, argue for devoting more explicit attention to the motivational link between core senses and figurative subsenses. Such motivational links could specifically involve conceptual metaphors in the Lakoffian sense (Van der Meer, Swanepoel), or even image schema (Swanepoel) (Geeraerts, 2007, p. 1169).

Aunque en nuestra propuesta hemos empleado una estructura lineal tradicional, en los análisis previos que sirven de base para las decisiones lexicográficas adoptadas se muestra de modo claro la naturaleza radial mediante la que concebimos la polisemia y los mecanismos (cognitivos y lingüísticos) que subyacen a cada ampliación de significado. Entendemos, por lo tanto, que con este trabajo sentamos las bases para futuras aproximaciones a los adjetivos *grande* y *pequeño* desde perspectivas lexicográficas de carácter eminentemente cognitivista.

REFERENCIAS BIBLIOGRÁFICAS

- Apresjan, J. D. (1973). Regular Polysemy. *Linguistics*, 142, 5-39. <https://doi.org/10.1515/ling.1974.12.142.5>
- Atkins, B. T. S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press. <https://doi.org/10.1093/oso/9780199277704.001.0001>
- Becerra Hiraldo, J. M. (2000). Adjetivos de dimensión: sentido figurado para personas. En G. Wotjak (ed.), *En torno al sustantivo y adjetivo en el español actual: aspectos cognitivos, semánticos, (morfo)sintácticos y lexicogenéticos* (pp. 291-400). Iberoamericana. <https://doi.org/10.31819/9783865278425-034>
- Bosque, I. (1985). Usos figurados de los adjetivos que denotan dimensiones físicas. En M. Alvar López (hom.), *Philologica hispaniensia: in honorem Manuel Alvar* (Vol. 2) (pp. 63-80). Gredos.

- De Hipona, A. (2015). *Confesiones*. Verbum.
- Di Tullio, A. (2010 [2005]). *Manual de gramática del español*. Waldhuter.
- Ebeling, K. S. & Gelman, S. A. (1994). Children's use of context in interpreting "big" and "little". *Child Development*, 65(4), 1178–1192. <https://doi.org/10.2307/1131313>
- Fauconnier, G. & Turner, M. (2008). Rethinking Metaphor. En R. Gibbs (ed.), *Cambridge Handbook of Metaphor and Thought* (pp. 53-66). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.005>
- Fiske, A. P. (2004). Four Modes of Constituting Relationships: Consubstantial assimilation; space, magnitude, time, and force; concrete procedures; abstract symbolism. En N. Haslam (ed.), *Relational models theory. A contemporary overview* (pp. 61-146). Lawrence Erlbaum Associates.
- Freedman, D. G. (1979). *Human sociobiology*. Free Press.
- Gale, R. M. (1968). *The language of time*. Routledge and Kegan Paul.
- Geeraerts, D. (2007). Lexicography. En D. Geeraerts & H. Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 1160-1174). Oxford University Press.
- Grady, J., Oakley, T. & Coulson, S. (1999). Blending and metaphor. En R. W. Gibbs Jr. & G. J. Stee (eds.), *Metaphor in Cognitive Linguistics: Selected papers from the 5th International Cognitive Linguistics Conference* (pp. 101-124). John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.175.07gra>
- Heine, B. (1997). *Cognitive foundations of grammar*. Oxford University Press. <https://doi.org/10.1093/oso/9780195102512.001.0001>
- Hermanson, E. A. (2006). *Metaphor in Zulu*. Sun press. <https://doi.org/10.18820/9781920109288>
- Higham, P. A. & Carment, D. W. (1992). The rise and fall of politicians: The judged heights of Broadbent, Mulroney and Turner before and after the 1988 Canadian federal election. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 24(3), 404-409. <https://doi.org/10.1037/h0078723>
- Ibarretxe-Antuñano, I. (2010). Lexicografía y Lingüística Cognitiva. *Revista Española de lingüística aplicada*, 23, 195-214.
- Ivanov, V. V. (1979). La semiótica de las oposiciones mitológicas de varios pueblos. En Y. M. Lotman & N. Méndez (aut.), *Semiótica de la Cultura* (pp. 159-172). Cátedra.
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226177847.001.0001>
- Judge, T. A. & Cable, D. M. (2004). The Effect of Physical Height on Workplace Success and Income: Preliminary Test of a Theoretical Model. *Journal of Applied Psychology*, 89(3), 428-441. <https://doi.org/10.1037/0021-9010.89.3.428>
- Kilgariff, A. (2006). Word senses. En E. Agirreand & P. Edmonds (eds.), *Word Sense Disambiguation: Algorithms and Applications* (pp. 29-46). Springer. https://doi.org/10.1007/978-1-4020-4809-8_2

- Klein, E. (1980). A Semantics for Positive and Comparative Adjectives. *Linguistics and Philosophy*, 4, 1-45. <https://doi.org/10.1007/BF00351812>
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. The University of Chicago Press.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books.
- Mandler, J. M. (2010). The spatial foundations of the conceptual system. *Language and Cognition*, 2(1), 21-44. <https://doi.org/10.1515/langcog.2010.002>
- Mulier, U. (2009). *Análisis semántico – sintáctico de cuatro adjetivos de dimensión: alto, bajo, largo y corto* [Tesis de maestría. Universiteit Gent. Faculteit Letteren en Wijsbegeerte]. Universiteitsbibliotheek Gent. <https://lib.ugent.be/catalog/rug01:001414042>
- NGLEM = Real Academia Española & Asociación de Academias de la Lengua Española (2010). *Nueva gramática de la Lengua Española. Manual*. Espasa.
- Radden, G. (2011). Spatial time in the West and the East. En M. Brdar, M. Omazic, V. Pavicic Takac, T. Gradecak-Erdeljic & G. Buljan (eds.), *Space and Time in Language* (pp. 1-40). Peter Lang.
- Real Academia Española. *Corpus de referencia del español actual* [en línea] [consultado en julio de 2024]. <http://corpus.rae.es/creanet.html>
- Schubert, T. W. (2005). Your Highness: Vertical Positions as Perceptual Symbols of Power. *Journal of Personality and Social Psychology*, 89(1), 1-21. <https://doi.org/10.1037/0022-3514.89.1.1>
- Schubert, T. W., Waldzus, S. & Giessner, S. (2009). Control over the association of power and size. *Social Cognition*, 27, 1-19. <https://doi.org/10.1521/soco.2009.27.1.1>
- Schwartz, B., Tesser, A. & Powell, E. (1982). Dominance cues in nonverbal behavior. *Social Psychology Quarterly*, 45(2), 114-120. <https://doi.org/10.2307/3033934>
- Seco, M. (2003). *Estudios de lexicografía española*. Gredos.
- Sinha, Ch. (2014). Is space-time metaphorical mapping universal?: Time for a cultural turn. En L. Filipović & M. Pütz (eds.), *Multilingual Cognition and Language Use Processing and typological perspectives* (pp. 183-202). John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.44.07sin>
- Swanpoel, P. (1992). Linguistic motivation and its lexicographical application. *South African Journal of Linguistics*, 10, 49-60. <https://doi.org/10.1080/10118063.1992.9723878>
- Swanpoel, P. (1998). Back to basics: Prepositions, schema theory, and the explanatory function of the dictionary. En T. Fontenelle, P. Hiligsmann, A. Michiels, A. Moulin, & S. Theissen (eds.), *Euralex'98 Proceedings* (pp. 655-666). Université de Liège. <https://n9.cl/87lmj>
- Tribushinina, E. (2008). *Cognitive reference points. Semantics beyond the prototypes in adjectives of space and colour* [Tesis doctoral. Leiden University. Landelijke Onderzoekschool Taalwetenschap]. Scholarly Publications. <https://hdl.handle.net/1887/13224>

- Tyler, A. & Evans, V. (2003). *The Semantics of English Prepositions. Spatial Scenes, Embodied Meaning and Cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486517>
- Van der Meer, G. (2000). Core, subsense, and the New Oxford Dictionary of English. En U. Heid, S. Everet, E. Lehmann & C. Rohrer (eds.), *Euralex 2000 Proceedings* (pp. 419-431). Institut für Maschinelle Sprachverarbeitung. <https://n9.cl/jy328>
- Van der Meer, G. (2006). It's about Time: On Coherence and Simplicity in Dictionary Entries. *English Studies*, 87, 602-616. <https://doi.org/10.1080/00138380600768361>
- Whorf, B. (1941). The relation of habitual thought and behavior to language. En L. Spier *et al.* (eds.), *Language, culture, and personality. Essays in memory of Edward Sapir* (pp. 75-93). Sapir Memorial Publication Fund.
- Ynduráin Pardo de Santayana, C. (2019). *Largo y corto: de lo espacial a lo temporal*. *RILEX. Revista sobre investigaciones léxicas*, 2/1, 25-39. <https://doi.org/10.17561/rilex.v2.n1.2>
- Ynduráin Pardo de Santayana, C. (2020). *Alto y bajo: de lo dimensional a lo cuantitativo*. Propuesta lexicográfica. *Futhark, Humanities and Social Sciences Review*, 15, 191-212. <https://doi.org/10.12795/futhark.2020.i15.12>
- Ynduráin Pardo de Santayana, C. (En prensa). *Alto, bajo, profundo y superficial: de lo dimensional a lo cualitativo*. Propuesta lexicográfica. *Lexis. Revista de Lingüística y Literatura*.
- Yu, N. (1998). *The Contemporary Theory of Metaphor. A perspective from Chinese*. John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.1>



ANÁLISIS DE LAS FRECUENCIAS LÉXICAS NOMINALES, ADJETIVALES Y ADVERBIALES PRESENTES EN ESCRITOS DE ESTUDIANTES UNIVERSITARIOS DE LA CLÍNICA DE ESCRITURA DE LA FACULTAD DE PSICOLOGÍA DE LA UASLP¹

ANALYSIS OF THE NOMINAL, ADJECTIVAL AND ADVERBIAL LEXICAL FREQUENCIES PRESENT IN WRITINGS OF UNIVERSITY STUDENTS OF THE WRITING CLINIC OF THE FACULTY OF PSYCHOLOGY OF THE UASLP

Marco Antonio Pérez Durán
Universidad Autónoma de San Luis Potosí
marco.duran@uaslp.mx

Ma. Guadalupe Rojas Corona
Universidad Autónoma de San Luis Potosí
rojasgpe@outlook.es

RESUMEN

Este estudio léxico-matemático analiza la frecuencia léxica de vocablos nominales, adjetivales y adverbiales de 68 producciones de estudiantes universitarios que asistieron a la Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí, México, antes de pandemia por COVID-19. El objetivo es realizar un estudio exploratorio-descriptivo para obtener un panorama general de las frecuencias léxicas de dichos vocablos. El método de análisis es de corte cuantitativo descriptivo de tipo exploratorio contrastivo con un enfoque empírico no transversal. Para ello, se utilizaron el programa libre de cómputo *AntConc*, las fórmulas matemáticas Porcentaje de Vocablos (PV) y Fórmula de Palabras Nacionales (PN) (López Morales, 2011). Los datos obtenidos se contrastaron con el corpus especializado *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). Los resultados indican que de los 100 vocablos obtenidos en la muestra solo 10 aparecen entre las primeras 50 posiciones de dicho corpus. Esto revela la necesidad por desarrollar intervenciones enfocadas en incrementar el vocabulario disciplinar.

Palabras clave: escritura académica, frecuencia léxica, léxico académico, léxico general, vocablo.

ABSTRACT

This lexical-statistical study analyzes the lexical frequency of nominal, adjectival and adverbial words of 68 productions of college students who attended the Writing Clinic of the Faculty of Psychology of Autonomous University of San Luis Potosi, Mexico, before the COVID-19 pandemic. The objective is to carry out an exploratory-descriptive study to obtain a general overview of the lexical frequencies of these words. The analysis method is quantitative, descriptive, exploratory-contrastive with a non-cross-sectional empirical approach. To do this, the free computing program *AntConc*, the mathematical formulas Percentage of Words (PV) and Formula of Notional Words (PN) (Lopez Morales, 2011) were used. The data obtained were contrasted with the specialized corpus *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). The results indicate that the 100 words obtained in the sample, only 10 appear among the first 50 positions of this corpus. The data reveals the need to increase the disciplinary vocabulary.

Keywords: academic writing, academic lexicon, general lexicon, lexical frequency, word.

¹ La realización del presente estudio no hubiera sido posible sin la información de los textos de escritura que nos facilitó la Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí.

Recibido: 12-04-2024
Aceptado: 23-07-2024



1. INTRODUCCIÓN

La escritura es una habilidad compleja que contribuye al fortalecimiento de la comunicación en el ámbito académico. Esta habilidad se conforma de otras competencias básicas, por ejemplo, el conocimiento del vocabulario, la ortografía y la organización discursiva (Montersen, Smith-Lock & Nickels, 2009, como se ha citado en Gómez Vera, Sotomayor, Bedwell, Domínguez & Jéldrez, 2014; Rojas Corona, Silva-Maceda & Romero-Contreras, 2024). Dichas habilidades son antecesoras del desarrollo de la escritura, lo que supone que su dominio impacta en la producción escrita y, por lo tanto, en el desarrollo académico de los estudiantes dentro de la disciplina que han decidido estudiar (Rojas Corona, Silva-Maceda & Romero-Contreras, 2024). Conocer cuál es el vocabulario que los estudiantes emplean puede ayudar a los docentes a desarrollar estrategias para dominar el vocabulario académico y disciplinar. El objetivo de este trabajo es hacer un estudio exploratorio de las frecuencias léxicas de vocablos nominales, adjetivales y adverbiales extraídos de redacciones universitarias de estudiantes que asistieron a la Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí (UASLP), México, antes de la pandemia COVID-19. La premisa de la que parte este estudio es que los estudiantes que asistieron a este centro ya evidenciaban uso del vocabulario académico propio de su disciplina, pero no un dominio satisfactorio.

La literacidad académica se define como una práctica de interacción social oral o escrita que se ubica y se produce dentro de las actividades académicas. A través de ella, se observa la capacidad del estudiantado de poder interactuar con la información académica a partir de la manipulación, uso y estructuración del discurso académico (Vanegas, 2020). En estos estudios, se define como texto académico a toda producción escrita que se realiza en el ámbito universitario (Martín-Sánchez *et al.*, 2015). Dentro de la escritura académica, el dominio del léxico se percibe como unidad formal, semántica y pragmática que cumple con una función comunicativa

específica. Haché de Yunén (1991) y Sánchez-Avendaño (2005) han afirmado que, en cuanto al vocabulario, los estudiantes manifiestan resultados poco favorables de uso y dominio, los cuales se observan cuando tratan de leer o escribir (transcribir) lo que han leído, por lo que fallan en la transformación del conocimiento. Esto significa una desventaja para el alumnado, ya que es por medio de la escritura que pueden validar los conocimientos de su disciplina.

Para mejorar los procesos de escritura, en muchas universidades en Latinoamérica, se han abierto centros cuyo énfasis está en ayudar a los estudiantes a mejorar sus habilidades básicas y necesarias de escritura, por ejemplo, el dominio del vocabulario académico y el vocabulario de su disciplina (Molina-Natera, 2016; Starkie & Sidotti, 2017). En México, la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí cuenta con un centro de escritura denominado Clínica de Escritura, el cual se fundó en 2017 y tomó como referencia el esquema de trabajo aplicado en la Facultad de Educación de la Universidad de Diego Portales en Chile. Este es un centro estratégico enfocado en la mejora y atención personalizada a estudiantes universitarios que requieran de apoyo educativo en una habilidad tan indispensable y necesaria como la escritura. Este apoyo se logra mediante la autorevisión y corrección de su texto, así como la práctica oral de los recursos del lenguaje y de discurso antes de la escritura.

En este trabajo, se estudia el léxico de las redacciones de un grupo de estudiantes a través de la frecuencia léxica (Fl desde este momento). Según Bargetto Fernández y Riffo Ocares (2019), la Fl se refiere a una palabra que por su uso se vuelve frecuente frente a otro conjunto de palabras que también aparecen en el texto de uso menor o mínimo. A través de la frecuencia léxica, se pueden conocer tres aspectos relevantes sobre el dominio del léxico por parte del estudiante: a) la comprensión sobre la relación interna de las palabras frecuentes en el corpus estudiado y su determinación de las

causas léxicas asociativas que pudieran explicar mejor el comportamiento léxico del hablante y de la lengua en general; b) la idea de que las palabras que alcanzan una frecuencia alta son unidades léxicas iniciales del aprendizaje, por lo que son ítems que frecuentemente aparecerán en el léxico de los informantes, en los textos académicos y en la comunicación discursiva sin importar su posición; y c) la idea de que el léxico es pieza angular en la transmisión y recepción del conocimiento, por lo que con su estudio se pueden explicar mejor las variaciones léxicas de los textos y discursos académicos.

En el siguiente apartado se presenta la base teórica sobre la cual se sustenta este trabajo de corte léxicométrico enfocado en el estudio del vocabulario académico y vocabulario disciplinar.

1.1. MARCO TEÓRICO

En general, la escritura es 1) una habilidad clave para el desarrollo del aprendizaje y la comunicación dentro de las diferentes disciplinas, 2) el vehículo por el cual se accede al conocimiento, 3) una estructura mediante la cual se organiza el pensamiento, 4) un sistema secundario del pensamiento y 5) un sistema mediante el cual se crea, se transmite y se conserva el conocimiento. Además, la escritura se construye con los recursos lingüísticos adecuados a nivel académico, es decir, dominio de la ortografía acentual, literal y puntual, de los aspectos morfológicos y sintácticos, del léxico académico y disciplinar, conocimiento de los géneros discursivos, entre otros (Snow & Ucelli, 2009; Ucelli, Dobbs & Scott, 2013).

Enseñar a leer y escribir en la universidad significa ayudar institucionalmente a los estudiantes a ingresar y participar en las prácticas discursivas propias de las profesiones para las que se preparan y en las que se van a especializar (Narvaja de Arnoux, 2011; Carlino, 2013). Uno de los aspectos relevantes en la producción del texto escrito es el léxico, el cual se define como el conjunto de lexías obtenidas de producciones textuales (Wagner, 2000) que se usan de forma recurrente en los géneros textuales dentro del

ámbito académico (Parodi, 2008; Castillo Fadić & Sologuren, 2020). Este último sirve como mecanismo para el fortalecimiento de la comprensión lectora y la escritura académica (Romero-Contreras, Silva-Maceda & Snow, 2021). Conocer el conjunto de lexías que comúnmente aparecen en determinados géneros textuales puede ayudar a los docentes a decidir sobre el léxico que se debe incorporar en la planeación de sus clases.

Por su parte, la escritura académica es un tipo de escritura que se enseña a partir de los dos últimos grados del nivel educativo básico, y los grados de la educación media, pero que se utiliza más en el ámbito universitario. De acuerdo con Vanegas (2020) esta escritura permite construir el conocimiento, conocer y aprender sobre la disciplina a través de prácticas letradas con características y formas de escribir del área de conocimiento. Esto se debe a que todas las tareas y trabajos que forman parte de la vida académica universitaria están contruidos con base en la escritura del tipo académico. Se reconocen varias características de este tipo de escritura 1) ser clara, entendible y puntual, 2) contar con un propósito académico, 3) tener un tono impersonal, 4) tener argumentos sólidos que validen el propósito del trabajo y, lo principal, 5) generar o renovar el conocimiento en torno al área de estudio. Un texto académico se construye a partir de un léxico del tipo general y del tipo académico o disciplinar. El primero se entiende como el vocabulario que se utiliza para cualquier situación comunicativa, mientras que el segundo se centra más en el ámbito especializado. Para la construcción de un escrito académico se requiere de ambos en el sentido de que el primero ayuda a soportar al segundo.

Aunque en un contexto general, los términos léxico y vocabulario pueden tomarse como sinónimos, para los estudios de corte lexicológico como es el caso de este trabajo, esto no sucede así. El léxico es un término más abstracto que se refiere a la totalidad de palabras que el locutor (escritor en este caso) sabe en un momento determinado, de ahí su valor global y general de la lengua, mientras que el término vocabulario alude a los vocablos propios

de la lengua que se han usado en un acto de habla particular, por tanto, es un concepto más material (Madrigal & Vargas, 2016). Dicho de otro modo, el vocabulario es un hecho concreto que sirve para construir un texto, su unidad de análisis es el *type* o palabra diferente, mientras que el léxico es algo universal que pertenece a la lengua, su unidad de estudio es la palabra repetida o *token*.

Los *types* y *tokens* son unidades que se pueden encontrar en un corpus, o bien que pueden conformar uno. Un corpus se define como el conjunto general de las muestras orales o escritas que se han recogido en un contexto y momento determinado y que pueden servir para una investigación. Para estudiar los corpus académicos se utilizan pruebas de riqueza léxica en las que los *types* y los *tokens* son las unidades de análisis. Estas pruebas son una fuente de información valiosa para los especialistas porque se analiza el desarrollo cuantitativo del léxico de cualquier corpus, en este caso, de los estudiantes universitarios. Tomando en cuenta los resultados de estas pruebas se puede obtener la riqueza léxica que es el grado de madurez y de dominio léxico general de un estudiante medidos en su producción escrita por el porcentaje de vocablos de contenido nocional (Álvarez Martín, 2019) y por el intervalo de palabras y vocablos de contenido nocional (Cuba Vega, 2007).

En el caso del léxico académico, el corpus más conocido es la lista de vocabulario académico elaborada y evaluada por Coxhead en el año 2000 (para el contexto español esta lista fue traducida y adaptada por Juan C. Ripoll). Dicho corpus se obtuvo de muestras que pertenecían a diferentes disciplinas universitarias, entre 15 a 28 disciplinas diferentes. Como resultado la autora extrajo las 2.000 palabras más frecuentes del inglés que se repetían por lo menos 100 veces en el corpus. Sin embargo, esta lista de vocabulario académico no presenta la clasificación de vocabulario por disciplina que es otro tipo de estudio que se puede realizar.

El vocabulario disciplinar pertenece al ámbito de la especialización académica. En este tipo de léxico, se pueden encontrar palabras cuyo significado cambie en relación con la disciplina que haga uso de ellas. Un corpus de esta naturaleza es *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* que está compuesto por 143 millones de palabras. Este se obtuvo de 20.000 textos sobre investigaciones de psicología, por lo que pertenece a un solo ámbito de estudio. Este último se ha utilizado para esta investigación como corpus de contraste debido al tipo de población que permitió obtener la muestra del corpus de este estudio, escritos realizados por estudiantes de la Facultad de Psicología.

2. METODOLOGÍA

La presente investigación se enmarca en los estudios léxico-matemáticos de tipo cuantitativo descriptivo de tipo exploratorio contrastivo, con un enfoque empírico no transversal por la naturaleza del estudio. En este se busca cuantificar y describir las voces nominales, adjetivales y adverbiales que se obtuvieron de textos escritos por estudiantes de una universidad estatal y pública de México.

2.1. MUESTRA

La muestra está compuesta por 68 escritos realizados por estudiantes que asistieron a tomar sesiones de asesoría en escritura académica en el centro de prácticas Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí. La recolección de estos escritos se realizó en dos momentos. El primero se refiere a los textos que se recogieron antes de que los estudiantes tomaran sus sesiones de asesoría, 34 textos pre-asesoría. Una vez que los estudiantes terminaron dichas sesiones, ellos elaboraron otro escrito denominado textos post-asesoría, 34 en total. La muestra se obtuvo durante enero y febrero de 2020, antes del inicio de la pandemia de COVID-19 y el cierre de escuelas en México. Con los datos obtenidos de

la muestra pre y post-intervención se conformó el corpus de esta investigación. Para su estudio, se realizó el análisis del corpus por relación: *types* y *tokens*, lematización de las unidades léxicas y N-grama de los *types*.

2.2. LA CLÍNICA DE ESCRITURA

La Clínica de Escritura es un centro de prácticas que se encuentra ubicado dentro de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí. Este centro cuenta con asesores, alumnos de la licenciatura en Psicopedagogía, con énfasis en lengua, que realizan sus prácticas profesionales. Estos asesores son capacitados en procesos de escritura académica y recursos del lenguaje académico para brindar apoyo en el desarrollo de las habilidades de escritura académica mediante sesiones individuales de 50 minutos. Para poder obtener este tipo de asesoría, el alumno interesado la solicita mediante la página de Facebook². Una vez hecho esto, se le agenda una fecha de asistencia. Durante la primera cita los estudiantes deben realizar un escrito *in situ* a partir de la pregunta de elicitación: *¿Qué has aprendido sobre tu carrera?* De esta forma se busca que el alumno escriba un texto real. Este sirve de base para la detección de los errores de escritura. Por lo regular, la extensión de este debe ser mayor a una cuartilla (entre 200 y 250 palabras en promedio).

Una vez obtenido el texto, se somete a una evaluación ciega por tres asesores con el propósito de focalizar y jerarquizar de forma objetiva los problemas de escritura y planear las asesorías de acuerdo con los problemas más relevantes. Al terminar las sesiones, se solicita al estudiante redactar, nuevamente, un texto, para ello se toma de base la misma pregunta del inicio. A este paso se le denomina evaluación post-asesoría.

2.3. HERRAMIENTAS DE ANÁLISIS

Para el análisis de los datos, se utilizó el programa *AntConc*³, software de acceso libre. Con este programa, se identificaron los *types* y *tokens* de la muestra. La

² https://www.facebook.com/ClinicaEscrituraUASLP/?locale=es_LA

³ <https://www.laurenceanthony.net/software/antconc/>.

relación entre *types* y *tokens* se denomina *ratio* y se representa con las siglas TTR. La relevancia de establecer la relación, *ratio* radica en que esta es necesaria para determinar la riqueza léxica de la muestra. Para encontrar la riqueza léxica, se utiliza la fórmula: $TTR = \text{Número de } types * 100 / \text{número de } tokens$ (Capsada Blanch & Torruella Casañas, 2017). Como se sabe, el TTR presenta la desventaja de que el cálculo depende de la longitud del texto, ya que existe una relación inversa entre la cantidad de *tokens* y *ratio*. Además de esta fórmula, se utilizó la de Honoré (1979), la cual se puede relacionar con la Ley de Zipf (1949) que indica que con cada aparición de una palabra esta desarrollará un patrón de frecuencia exponencial en un contexto de uso en relación con la palabra que le antecede (Rodríguez Velásquez, Prieto Bohórquez, Ortiz Salamanca, Bautista Charry, Bernal & Avilán Vargas, 2006). A continuación, se presenta la ejemplificación de la fórmula de Honoré (Imagen 1).

$R = \frac{100 \log N}{1 - (V_1 / V)}$	donde	R = riqueza léxica N = n° palabras en el texto V = n° de formas en el texto V ₁ = n° de palabras usadas una sola vez
--	-------	--

IMAGEN 1: Ejemplificación de fórmula de Honoré

El análisis que se presenta a través de esta fórmula parte de la idea de que, en la palabra con más apariciones, se tiene una frecuencia igual al doble de la frecuencia de la segunda palabra y así sucesivamente. Los resultados obtenidos se organizan en tablas de rasgos y frecuencias. Una vez aplicada la fórmula se obtienen los lemas, es decir, la unidad autónoma que constituye el léxico de un idioma.

Otro análisis que se obtuvo mediante el software *AntConc* fue la organización de la información en N-gramas o *clústers* de los *types*. Un N-grama es un conjunto de *n* elementos consecutivos que se relacionan a partir de las relaciones sintácticas entre las unidades léxicas. Estos N-gramas pueden incluir palabras, números, símbolos. En este trabajo, se utilizaron los N-gramas para comprender las relaciones asociativas estructurales entre las palabras del

corpus. Para ello se tomó un ítem que sirve de núcleo con la cual se asocia el resto de las respuestas. A esto se le denominó familia léxica. El concepto de familia léxica es relevante en estos estudios porque funciona como la precursora de la organización de la información.

Además de lo anterior, se aplicaron las fórmulas Porcentaje de Vocablos (pv) y de Palabras Nocionales (pn), propuestas por López Morales (2011). Para encontrar el resultado de la fórmula Porcentaje de Vocablos (pv), se requiere que se divida el total de vocablos (v) entre el total de las unidades léxicas comprendidas en el texto (n). Después el resultado de la división se multiplica por 100, su representación es $pv = v \cdot 100 / n$. Esta fórmula se usa para encontrar el porcentaje de vocablos, es decir, el indicador grueso para la riqueza léxica. Por último, se hizo el contraste entre los vocablos del corpus de estudiantes universitarios con el de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (Xodabande & Xodabande, 2020). Para este análisis, se tomaron las diez unidades léxicas más frecuentes del corpus de estudiantes y se buscaron por su posición en el corpus de corte psicológico. La posición de estas es una característica relevante en el conocimiento del vocablo, ya que permite conocer la activación del léxico de los informantes de forma clara. Por último, el corpus se redujo a dos subclasificaciones: 1) léxico general, y 2) léxico especializado.

2.4. PROCEDIMIENTO DE ANÁLISIS DEL CORPUS

Una vez obtenidas las muestras de escritura pre y post-intervención, se conformó el corpus de este estudio y se procedió a su análisis. Para ello, se llevaron a cabo los siguientes pasos:

- 1) Se procedió a la anonimización de los textos según el reglamento de ética de la Facultad de Psicología.⁴
- 2) Posteriormente, se segmentó cada una de las redacciones en unidades léxicas y se creó una sola lista de palabras.

⁴Este reglamento se puede consultar en <https://www.psicologia.uaslp.mx/Paginas/Servicios/3737#gsc.tab=0>.

- 3) Mediante el proceso de lematización, se separaron las palabras funcionales (sin contenido semántico) de aquellas que tuvieran contenido semántico. Una vez conformado este conjunto de palabras, se quitó las voces verbales, ya que esta submuestra forma parte de otro estudio, por lo que solo se conservaron voces adverbiales, adjetivales y nominales.
- 4) Posterior a la selección de las voces nominales, adjetivales y adverbiales de los escritos donados por los estudiantes que asistieron a la Clínica de Escritura, estas se organizaron por familias léxicas, lo cual facilitó el conteo final y puso en evidencia todas las relaciones entre vocablos. Cabe mencionar que no se consideraron para el análisis las preposiciones, onomatopeyas, artículos y conjunciones.
- 5) Después, se hizo la corrección tanto de forma como de contenido de cada unidad léxica que compone el corpus. Se encontraron y se corrigieron errores de escritura y de acentuación comunes en las unidades léxicas que pertenecen a la escritura hecha a lápiz y papel. La depuración de la información tuvo una duración de 200 días por el tipo de procedimiento manual que se implementó.
- 6) Posterior a la depuración de la información, se organizó el corpus en Excel y en formato txt. En este último, se pudieron quitar acentos del corpus para mejorar el análisis de la información en el software *Ant-Conc*. Después de procesar la información en este software, se utilizó Excel para ingresar la fórmula de Honoré (1979) y poder generar los resultados que se exponen. Todo el procedimiento de análisis tuvo una duración de 280 días de trabajo.
- 7) Se compararon los datos obtenidos con el corpus especializado de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). Para ello, se tomaron las diez unidades léxicas más frecuentes que aparecieron en ambos corpus.
- 8) La información se organizó en tablas para un mejor manejo de los datos y para poder hacer más evidente el análisis.

3. RESULTADOS

Para identificar los resultados generales, se presenta la siguiente tabla con los datos obtenidos del corpus distribuido en total de palabras (tp), vocablos (v), *types tokens ratio* (ttr), valor de riqueza léxica (vrl) y palabras de contenido nocional (iat), que aparecieron durante el análisis. Esto se puede observar en la Tabla 1.

Total	tp	v	ttr	vrl	iat
Muestra general	17,167	2,306	13.43	6.32	7.44
Preasesoría	16,890	2,290	13.55	6.28	7.49
Postasesoría	17,050	2,297	13.47	6.23	7.49

TABLA 1: Resultados generales del análisis del corpus. *Elaboración propia*

Como se aprecia en la Tabla 1, los resultados muestran que el número total de palabras del corpus asciende a 17,167. Al analizarlas por pre-asesoría y post-asesoría se obtuvo lo siguiente: el total de términos por pre-asesoría es de 16,890 y post-asesoría es de 17,050. Al comparar los resultados se encontró una diferencia a favor del corpus general de 277 unidades léxicas, en relación con los resultados pre-asesoría y de 117, en relación con post-asesoría.

Sobre el número de vocablos, el corpus ascendió a 2,306. Al contrastar este resultado, el número total de términos pre-asesoría es de 2,290 y el de post-asesoría es de 2,297. En relación con el análisis ttr, parámetro primordial del análisis de la variación léxica, la muestra general es de 13.43 del corpus. Vinculado al análisis del ttr, se encuentra el intervalo de aparición de contenido nocional (iat) que refleja, como se ha mencionado, las palabras nocionales en el texto. El corpus general presenta un iat de 7.44. En contraste, la muestra pre y post-asesoría tiene un iat de 7.49. Esto significa que el corpus analizado presenta un alto nivel de palabras nocionales.

Además de los indicadores anteriores, el promedio de palabras asciende a 14.24, la longitud del corpus es de 8.61 términos y la media es de 10. Sobre el análisis de lemas o vocablos lematizados del corpus, hay 1,153 vocablos lematizados y 2,324 vocablos totales. Al comparar estos dos resultados, hay una reducción de 1,171 vocablos (50%) de la totalidad. A su vez, el corpus lematizado se dividió en 682 nominales (58.2%), 386 adjetivos (32.9%) y 85 adverbios (7.2%).

La información de los vocablos más frecuentes por distribución de las diez respuestas se organiza usando las siguientes clasificaciones: lugar de aparición (n), vocablo (v), frecuencia (f), frecuencia acumulada (Fi) y la frecuencia absoluta (Hi). Todo esto se expone en la Tabla 2.

N	Vocablo	Frecuencia	Fi	Hi
1	<i>Psicología</i>	343	343	0.03695
2	<i>No</i>	342	685	0.02807
3	<i>Niño</i>	278	963	0.02232
4	<i>Mucha</i>	260	1223	0.02091
5	<i>Importancia</i>	208	1431	0.02091
6	<i>Carrera</i>	197	1628	0.01864
7	<i>Su</i>	196	1824	0.01506
8	<i>Materia</i>	193	2017	0.01441
9	<i>Todo</i>	179	2196	0.01355
10	<i>Mi</i>	178	2374	0.01355

TABLA 2: Orden de la frecuencia por distribución de los diez vocablos más frecuentes. Elaboración propia

En las primeras diez respuestas, hay cinco nominales, cuatro adjetivos y un solo adverbio. En relación con los nominales, el vocablo *psicología* acumuló el mayor número de frecuencias con 343, le siguió el adjetivo *mucha* con 260 y el adverbio *no* con 342 apariciones. Los mismos resultados se aprecian tanto en la frecuencia acumulada como en la absoluta. La categoría que menos apariciones presentó en las 100 palabras de las listas fue la adverbial, aunque el adverbio negativo no haya aparecido en segundo lugar de la lista de vocablos, de 2,306. La categoría que presentó mayor número de vocablos es la nominal y después la adjetival. En la Tabla 3, se presenta la distribución de las diez respuestas según sea su categoría gramatical.

	Nominales	Adjetivales	Adverbios
1	<i>Psicología</i>	<i>Mucha</i>	<i>No</i>
2	<i>Niño</i>	<i>Su</i>	<i>Muy</i>
3	<i>Importancia</i>	<i>Todo</i>	<i>Sí</i>
4	<i>Carrera</i>	<i>Mi</i>	<i>Siempre</i>
5	<i>Materia</i>	<i>Este</i>	<i>Bien</i>
6	<i>Infante</i>	<i>Conocida</i>	<i>Ahora</i>
7	<i>Personas</i>	<i>Esta</i>	<i>Aquí</i>
8	<i>Diferencia</i>	<i>Sus</i>	<i>Actualmente</i>
9	<i>Patología</i>	<i>Otros</i>	<i>Abí</i>
10	<i>Aprendizaje</i>	<i>Cada</i>	<i>Principalmente</i>

TABLA 3: Distribución de las diez respuestas según su categoría gramatical.
Elaboración propia

En el conjunto de vocablos nominales, aparecieron términos como: *psicología*, *carrera*, *materia*, *patología* y *aprendizaje*. Estas respuestas aparecieron

como esperadas, ya que estos nominales están vinculados al área de psicología como *patología* o *psicología*, o bien pertenecen estas unidades al vocabulario académico: *materia*, *carrera* o *aprendizaje*. En cuanto a la aparición de los adjetivos, estos se clasifican por la ubicación en relación con el nominal, la mayoría de ellos se usan antepuestos al nominal: adj+nominal. Solo dos adjetivos presentan un uso menos restrictivo, el adjetivo *mucha* por lo regular acompaña a un adverbio o nominal y el vocablo *conocido(a)* es una palabra deverbal. Sobre los adverbios, estos pueden ser de tipo modal (*muy*, *siempre*, *bien*, *principalmente*); o de negación (*no*), afirmación (*sí*), temporal (*ahora*, *actualmente*) y lugar (*aquí*, *ahí*).

Como se mencionó en la metodología, se sometió el corpus de vocablos a un nuevo procesamiento estadístico para obtener los vocablos lema de sus familias léxicas. Así pues, de los 2,306 se redujo a 1,152 vocablos lematizados (50 %) de 199 familias léxicas. Para el análisis, se presentan, en la Tabla 4, los seis vocablos lema que aparecen en la familia léxica con mayor producción.

	Vocablo lematizado	Vocablos	Familia léxica
1	Psicología	21	Psicopedagogía 81, psicopatología 75, psicólogo 61, psicóloga 26, psicopedagogo 24, psicológico 17, psíquica 9, psicosis 8, psicótica 6, psíquico 6, psicométrica 6, psicopedagógico 5, psicótico 4, psicoterapia 3, psicomotora 2, psicóticas 2, psicopatológico 2, psicoanalítica 2, psicoterapia 2, psicopatología 1, psicobiólogo 1.
2	Mucho	6	Mucho 128, mucha 68, muchísimo 8, muchísimas 2, muchísima 1, muchísimo 1.
3	Patología	6	Patología 74, patológico 37, patológicamente 7, patológica 4, patólogo 3, patologías 1.
4	Profesional	6	Profesional 51, profesor 21, profesión 8, profesora 6, profesionalidad 2, profesionalmente 2.
5	Conocimiento	4	Conocimiento 119, conocida 3, conocido 3, conociendo 1.
6	Todo	4	Todo 89, todos 44 toda 13, todas 32.

TABLA 4: Familia léxica de los seis vocablos lemas con mayor producción de vocablos.
Elaboración propia

La información que se desprende del cuadro anterior permite observar las relaciones entre el vocablo lema y el vocablo. Nótese que las familias léxicas más productivas del corpus están vinculadas al léxico de tipo nominal: *psicología*, *patología* y *conocimiento*, adjetival: *profesional* y *todo* y adverbial: *mucho*. De estas familias léxicas, las más productivas fueron: *psicología* con 21 vocablos, *mucho*, *patología* y *profesional* con seis y *conocimiento* y *todo* con solo dos unidades.

Sobre el análisis de los N-gramas o clúster, el número total asciende a 9,072 de un total de 17,166. El clúster nominal-nominal fue el que más apareció en el corpus con 5,789 ocasiones, seguido del adjetivo-nombre con 2,779, adjetivo-adjetivo o adverbio con 378 y adverbio-adverbio con 126. En la Tabla 5, se ubican los clústeres que más apariciones tuvieron en el corpus.

P	Relación gramatical	Frecuencia	Ejemplos
1	N-n	0.097	<i>niños adolescentes</i> <i>infancia adolescencia</i> <i>enseñanza aprendizaje</i> <i>comportamiento humano</i> <i>énfasis lengua</i>
2	Adj-n N-adj	0.086	<i>estas etapas</i> <i>esta carrera</i> <i>mi carrera</i> <i>objeto amoroso</i>
3	Adv-adj Adj-adv	0.042	<i>no sólo</i>

TABLA 5: Clústeres con más apariciones en el corpus de esta investigación.
Elaboración propia

En el análisis, se divide el corpus de estudiantes en dos grupos: léxico general y léxico académico. En este último, se engloban unidades léxicas denominadas especializadas. Para este análisis, se tomaron de nuevo el número de vocablos lematizados de 1,152. Se observó que el vocabulario general asciende a 807 (70%) vocablos y el académico a 345 (29.7%). La organización de esta información se puede apreciar en la Tabla 6.

Vocabulario	Ejemplos
General	<i>No, niño, mucha, importancia, su, mi(s), persona, diferencia, esto, otro, alguna, muy, cada, formación, desarrollo, observable, normal, manera, algo, nueva.</i>
Académico	<i>Psicólogo(a), carrera, materia, aprendizaje, lengua, humano, profesión, semestre, teoría, maestro(a), estudio, comportamiento, conducta, contexto, adolescente, clases, área, tema, educación, variable, factor, metodología, enseñanza, clínica, ideales, literaria, alumno(a), historia, curso, característica, licenciatura, escolar, proceso, concepto, experiencia, universidad, base, académica, autor(a), redacción, prueba, función, escrita, herramienta, ámbito, comunicación, enfermedad, identidad, institución, tarea, especializado, estructura, lectura, paciencia, frecuencia, texto, docente, duración, mental, atención, crítica, dinámica, ejemplo, interés, escuela, proyecto, arqueología, sistema, cultura, plan, salud, entorno, perspectiva, técnica, asignatura, artificial, conclusión, duda, facultad, análisis, búsqueda, consecuencia, doctor, economía, información, mente, suma, jurídica, pensamiento, sujeto, biología, nivel, percepción, personalidad, relevancia, castración, genética, interacción, manual.</i>

TABLA 6: Representación de las familias léxica de este corpus de investigación.
Elaboración propia

Por último, para el análisis contrastivo, se utilizó el corpus de vocabulario especializado de psicología de *Academic Vocabylary in Psychology Articles: A Corpus-Based Study* (2020) que contiene 1,526 términos y es parecido al corpus de estudiantes universitarios. En este corpus, de los 100 vocablos contrastados, 10 vocablos de los estudiantes aparecen en el listado especializado. Esto se puede observar en la Tabla 7.

P	Corpus estudiantes	Corpus especializado (<i>Academic Vocabylary in Psychology Art</i>)
1	<i>Psicólogo</i> (1)	<i>Psychological</i> (75)
2	<i>Conducta</i> (48)	<i>Conducted</i> (61)
3	<i>Contexto</i> (50)	<i>Context</i> (55)

(Continúa)

P	Corpus estudiantes	Corpus especializado (<i>Academic Vocabulary in Psychology Art</i>)
4	<i>Variables</i> (65)	<i>Variables</i> (65)
5	<i>Factor(es)</i> (69)	<i>Factors</i> (29)
6	<i>Metodología</i> (70)	<i>Methods</i> (62)
7	<i>Tiempo</i> (71)	<i>Temporal</i> (78)
8	<i>Clínica</i> (76)	<i>Clinical</i> (67)
9	<i>Características</i> (87)	<i>Features</i> (95)
10	<i>Proceso</i> (96)	<i>Process</i> (54)

TABLA 7: Análisis contrastivo del corpus universitario y el corpus de vocabulario especializado de psicología de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). Elaboración propia

4. CONCLUSIÓN

Sobre los resultados que se han encontrado a lo largo de esta investigación, se ha observado que independientemente de la extensión de los escritos que conformaron el corpus, los vocablos lema guardan un parecido con la lista de palabras del corpus *Academic Vocabulary in Psychology Articles: A Corpus-Based Study*, aunque no en el mismo orden y frecuencia. Llama la atención que los vocablos con más aparición por frecuencia, frecuencia acumulada y absoluta sean unidades pertenecientes a categorías como adverbios y adjetivos, unidades que cumplen la función de determinantes. Después aparecen con alto grado de frecuencia términos de categoría nominal, pero de vocabulario general como *niño*, *importancia*, *carrera* y *materia* y solo después aparecen vocablos que pertenecen al vocabulario

académico y muy pocos al disciplinar como *conocimiento*, *psicología*, *patología*, entre otras. Se observó una disminución del 10% de la totalidad del corpus de este trabajo en relación con otras investigaciones realizadas sobre el mismo tema (Read, 2000; Vine Jara, Fuentes Riffo, Neira Martínez & Poza Molina, 2021).

La escritura académica es una habilidad compleja que se solicita al estudiantado universitario, su estructura está basada en un dominio de léxico de corte académico y de corte disciplinar. El conocimiento de este es básico para que los estudiantes puedan comprender lo que leen y para escribir sobre un tema. Cuando el docente fortalece el dominio del vocabulario académico y el vocabulario disciplinar ayuda al estudiante en el proceso de adquisición y comprensión de las familias de palabras y del desarrollo de una escritura académica.

Como se ha comentado en este trabajo, la escritura académica, al ser un proceso complejo, se debe fortalecer desde el ámbito de la práctica de las habilidades que la componen, entre estas, la adquisición del vocabulario. La corrección del sistema y la revisión de las normas se harán en un momento posterior, es decir, en una revisión de estilo. Sobre el estudio que se llevó a cabo en este trabajo, la importancia que tiene el léxico es más que evidente porque este es la unidad de análisis y comunicación que se utiliza para todo.

Así, los resultados de esta investigación constituyen un aporte valioso, puesto que, no se han encontrado trabajos de léxico universitario donde se hayan obtenido y analizado tres categorías gramaticales (nominales, adjetivales y adverbiales) después de la pandemia por COVID-19 y mucho menos investigaciones que hayan sido obtenidas de textos como los que se presentaron durante el análisis de este artículo. Si bien es cierto que los resultados han aparecido después del COVID-19, todavía hay mucha información que sigue fortaleciendo la descripción del léxico dentro de la escritura académica

universitaria, nivel educativo en el que la cantidad de estudios, por lo menos en México, todavía no son suficientes.

El conocimiento de este tipo de información es significativo, ya que la presencia alta de los nominales es característica de la escritura en general. Este resultado es relevante para efectos didácticos porque se observan que, de las seis palabras más frecuentes, dos pertenecen al ámbito disciplinar (*psicología* y *patología*), dos al léxico académico (*profesional*, y *conocimiento*) y dos más al léxico general (*mucho* y *todo*). En este caso, la tendencia es que el léxico general sea más que el académico como se aprecian en los resultados. Sobre el léxico académico y el comparativo que se hace con el especializado de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study*, se observó como de las 10 palabras, *psicología* en el corpus de los estudiantes ocupa la primera posición, mientras que en otro corpus aparece en el lugar 75. Esto indica que existen otras palabras que están por debajo de esta posición en el corpus especializado. Sobre el resto de las unidades, la mayoría ocupa posiciones por encima del lugar 50.

Los resultados que se han visto en este artículo son un preámbulo para los interesados en la didáctica de la escritura disciplinar, o bien, en la producción de materiales administrativos de esta área. Además, los vocablos académicos identificados podrían relacionarse con los textos disciplinares por semestre, lo que ayudaría a la integración y dominio de estas respuestas como parte del repertorio léxico y como parte de los géneros académicos propios de las disciplinas.

En este sentido, esta investigación puede ser útil para distintas áreas: lingüística, lingüística aplicada, psicología, psicología social, psicología métrica, entre otras. Respecto a esto último, esta investigación cobra todavía más relevancia, pues al tratarse de un estudio del tipo cuantitativo descriptivo con enfoque exploratorio contrastivo realizado con base en textos reales producidos por estudiantes que asistieron a la Clínica de Escritura genera pistas en torno a que los estudiantes presentan vocabulario de su disciplina y que solo faltaría enseñarles a utilizarlo adecuadamente.

REFERENCIAS BIBLIOGRÁFICAS

- Álvarez Martín, L. (2019). Índices de riqueza léxica en estudiantes de 11 y 16 años de La Habana. *Universidad de La Habana*, 287, 267-286. <https://doi.org/10.4067/S0718-93032019000100341>
- Bargetto Fernández, M. Á. & Riffo Ocares, B. (2019). El reconocimiento de palabras y el acceso léxico: revisión de modelos y pruebas experimentales. *Boletín de filología*, 54(1), 341-361.
- Capsada Blanch, R. & Torruella Casañas, J. (2017). Métodos para medir la riqueza léxica de los textos. Revisión y propuesta. *Verba: Anuario galego de filoloxia*, 44, 347-408. <https://doi.org/10.15304/verba.44.3155>
- Carlino, P. (2013). Alfabetización académica diez años después. *Revista mexicana de investigación educativa*, 18(57), 355-381.
- Castillo Fadić, M. N. & Sologuren, E. (2020). Léxico frecuente, riqueza léxica y estereotipos sobre la lectura de profesores en formación. *Logos*, 30(1), 69-85. <https://doi.org/10.15443/RL3006>
- Coxhead, A. (2000). A new academic word list, *TESOL quarterly*, 34(2), 213-238.
- Cuba Vega, L. (2007). Riqueza léxica en escolares de primaria de Ciudad de La Habana. En M. Domínguez Hernández (ed.), *La lengua en Cuba. Estudios* (pp. 283- 308). Universidad de Santiago de Compostela.
- Gómez Vera, G., Sotomayor, C., Bedwell, P., Domínguez, A. M. & Jéldrez, E. (2014). Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria. *Documento de Trabajo*, 16, 1-19.
- Haché de Yunén, A.M. (1991). Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna. En H. López Morales (ed.), *La enseñanza del español como lengua materna* (pp. 47-60). Universidad de Puerto Rico.
- Honoré, A. (1979). Some Measures of Richness of Vocabulary. *ALLC Bulletin*, 7(2), 172-177.
- López Morales, H. (2011). Los índices de riqueza léxica y la enseñanza de lenguas. En F. J. de Santiago-Guervós, H. Bongaerts, J. J. Sánchez Iglesias & M. Seseña Gómez (coords.), *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE* (pp. 15-28). Asociación para la Enseñanza del Español como Lengua Extranjera-ASELE.
- Madrigal, M. & Vargas, E. (2016). Índice de riqueza léxica en redacciones escritas por estudiantes universitarios. *Kañina*, 40, 139-147. <https://doi.org/10.15517/rk.v40i3.29260>
- Martín Sánchez, T., Pascual Escagedo, C., & Puigdevall Balafuy, N. (2015). Nuevas metodologías para el aprendizaje de la conversación en ELE a través de las TIC.
- Molina-Natera, V. (2016). Los centros de escritura en Latinoamérica: consideraciones para su diseño e implementación. En G. Bañales, M. Castelló & N. Vega (eds.), *Enseñar a leer y escribir en la educación superior. Propuestas educativas basadas en la investigación* (pp. 339-362). Instituto de Investigación y Desarrollo Educativo.

- Mortensen, L., Smith-Lock, K., & Nickels, L. (2009). Text structure and patterns of cohesion in narrative texts written by adults with a history of language impairment. *Reading and Writing*, 22, 735-752.
- Narvaja de Arnoux, E. (2011). La escritura de tesis: apoyos institucionales y propuestas pedagógicas. En A. Vázquez, M. del C. Novo, I. Jakob & L. Pelliza (comps.), *Lectura, escritura y aprendizaje disciplinar* (pp. 25- 40). Subsede Universidad Nacional de Río Cuarto de la Cátedra unesco de Lectura y Escritura, Facultad de Ciencias Humanas, Centro de Investigación y Enseñanza de la Lectura y la Escritura.
- Parodi, G. (2008). Géneros académicos y géneros profesionales. En G. Parodi (ed.), *Accesos discursivos para saber y hacer* (pp. 75-96). Ediciones Universitarias de Valparaíso.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Rodríguez Velásquez, J., Prieto Bohórquez, S., Ortíz Salamanca, L., Bautista Charry, A., Bernal, P., & Avilán Vargas, N. (2006). Diagnóstico Matemático de la monitoria fetal aplicando la ley de Zipf-Mandelbrot. *Revista de la Facultad de Medicina*, 54(2), 96-107.
- Rojas Corona, M. G., Silva-Maceda, G. & Romero-Contreras, S. (2024). Relación del lenguaje académico y la comprensión lectora en la escritura académica en educación superior en pandemia. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 14(28). <https://doi.org/10.23913/ride.v14i28.1813>
- Romero-Contreras, S., Silva-Maceda, G. & Snow, C. E. (2021). Vocabulario académico y habilidades de lenguaje académico: predictores de la comprensión lectora de estudiante de primaria y secundaria en México. *Revista Pensamiento Educativo*, 58(2), 1-16. <https://doi.org/10.7764/PEL.58.2.2021.4>
- Sánchez-Avedaño, C. (2005). Los problemas de redacción de los estudiantes costarricenses: una propuesta de revisión desde la lingüística del texto. *Revista de Filología y Lingüística*, 31(1), 267-295. <https://doi.org/10.15517/rfl.v31i1.4419>
- Snow, C. E. & Uccelli, P. (2009). The challenge of academic language. En N. Torrance & D. R. Olson (eds.), *The Cambridge Handbook of literacy* (pp. 112-133). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609664.008>
- Starkie, E. I. G. & Sidotti, P. T. (2017). La evolución histórica de los Centros de Escritura Académica The Historic Evolution of Centres of Academic Writing. *Revista de educación*, 378, 9-29.
- Uccelli, P., Dobbs, C. L. & Scott, J. (2013). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication*, 30(1), 36-62. <https://doi.org/10.1177/0741088312469013>
- Vanegas, D. A. M. (2020). La escritura académica y la universidad: Una aproximación desde el discurso ideológico. *Revista Educación Superior y Sociedad (ESS)*, 32(1), 115-140.

- Vine Jara, A., Fuentes Riffo, K., Neira Martínez, A. & Poza Molina, C. (2021). Riqueza léxica en textos escritos por aprendientes de ELE. una contribución a las nociones específicas del plan curricular del instituto cervantes. *Nueva Revista del Pacífico*, 75, 201-226. <https://doi.org/10.4067/S0719-51762021000200201>
- Wagner, C. (2000). La diversidad léxica en la comunicación. *Revista Documentos Lingüísticos y Literarios UACH*, 23, 67-73.
- Xodabande, I. & Xodabande, N. (2020). Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study. *MEXTESOL Journal*, 44(3), 1-21.



LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR* EN ESPAÑA E HISPANOAMÉRICA

DIATOPIC VARIATION IN THE SPANISH-AMERICAN LEXICON: THE CASE OF *AGREGAR* AND *AÑADIR*

Elodie Weber
Universidad de La Sorbona
el.web@free.fr

RESUMEN

Ciertas parejas léxicas (*agregar/añadir*, *elegir/escoger*, *trozo/pedazo*, etc.) nunca vienen mencionadas en los estudios de dialectología tradicional o de geografía lingüística dedicados a los geosinónimos con diferente frecuencia de uso. Los términos que componen estas parejas se distinguen por ligeros matices denotativos relacionados con la existencia de varias acepciones. El examen, mediante el corpus electrónico *CORPESXXI*, de las frecuencias de uso de *agregar* y *añadir* muestra que estos verbos están efectivamente sometidos a la variación diatópica, la cual concierne dos acepciones en particular. Convendría realizar el mismo análisis con las demás parejas mencionadas arriba para ver si forman un conjunto coherente desde el punto de vista de la variación diatópica, e identificar posibles causas.

Palabras claves: variación diatópica, léxico hispanoamericano, geosinónimos, *agregar*, *añadir*.

ABSTRACT

Certain lexical pairs (*agregar/añadir*, *elegir/escoger*, *trozo/pedazo*, etc.) are never mentioned in the studies of traditional dialectology or linguistic geography studies devoted to geosynonyms with different frequencies of use. The terms that make up these pairs have the particularity of offering very slight denotative nuances linked to the existence of several meanings. The examination, using the *CORPESXXI* electronic corpus, of the frequencies of use of *agregar* and *añadir* shows that these verbs are indeed subject to diatopic variations, and that these variations concerns two acceptations of the verb in particular. The same analysis should be carried out with the other pairs mentioned above to see if they form a coherent set from the point of view of diatopic variation, and to identify possible causes.

Keywords: diatopic variation, Spanish-American lexicon, geosynonyms, *agregar*, *añadir*.

Recibido: 24-04-2024
Aceptado: 21-07-2024

DOI: <https://doi.org/10.17561/rilex.8.1.8882>



1. INTRODUCCIÓN

Entre los diferentes tipos de variación lingüística figura la variación diatópica, también llamada geolingüística, o variación de la lengua en función de la geografía. Diferentes rasgos están sujetos a este tipo de variación, entre los cuales el léxico. La variación diatópica del léxico se observa no solamente entre los distintos países de habla española que configuran el espacio americano, sino también entre estos mismos países y España (variación hispanoamericana).

Entre el léxico peninsular e hispanoamericano existen diferencias tanto gráficas como fonéticas, morfosintácticas y semánticas, siendo estas últimas el objeto del presente estudio. Las diferencias semánticas pueden ser examinadas bajo el punto de vista de la sinonimia, parcial o total, la cual permite identificar tres tipos de relaciones léxicas (Chacón García, 2016, pp. 184-185):

- oposiciones diatópicas simples o geosinonimia: unidades léxicas distintas pero misma denotación (sinonimia total). Por ejemplo, *ordenador* (España) y *computadora* (México)
- oposiciones connotativas: coexistencia de dos o varias voces con valores connotativos específicos (sinonimia parcial), en función del sexo, grupo de edad, registro o marco histórico. Por ejemplo, *sujetador* y *sostén* que coexisten en España, teniendo el segundo la connotación de palabra más anticuada; y “sostén”, único vocablo que se emplea en Chile.
- oposiciones denotativas: “voces idénticas con significados distintos en los diversos subsistemas del español” (Chacón García, 2016, p. 187). Por ejemplo, la palabra *tinto*, que en España refiere al vino tinto, tiene un significado único en América, el cual solo existe en Colombia: *café solo*.

Los estudios especializados que se han publicado acerca de los *geosinónimos* en el marco de la dialectología tradicional se limitan a enumerar, según un punto de vista onomasiológico, las distintas denominaciones que existen

en el mundo hispanohablante para referirse a un mismo concepto¹. Dichos estudios carecen de exhaustividad y, sobre todo, pasan por alto los problemas de coexistencia de varias voces en una misma variedad dialectal y, por consiguiente, de frecuencia.

Partiendo de una definición de americanismo que Buesa Oliver y Enguita Utrilla sintetizan como “todo elemento léxico que se difiere de su empleo en la norma peninsular, bien porque no sea conocido en España y su uso esté, por lo tanto, restringido al territorio americano [...], bien porque se emplee en América con acepciones diferentes o adicionales a las que tiene en la Península [...]” (1992, p. 24), se pueden distinguir dos tipos de americanismos, los que podrían calificarse de *estrictos* (elementos presentes en el español de América y ausentes en el español peninsular, *papa* frente a *patata* en España) y los que podrían recibir el calificativo de *relativos* (elementos presentes tanto en el español de América como en español peninsular, pero con distinto significado, uso contextual o con diferente connotación, frecuencia, etc.). En su análisis sobre las diferencias entre el léxico peninsular y el léxico americano, Haensch repasa los diferentes americanismos de cada categoría, destacando, en la de los americanismos relativos, un tipo de *geosinónimos* que va a ser el objeto del presente estudio: “misma unidad léxica formal, mismo significado, pero diferente frecuencia de uso” (2002, p. 45)². Desafortunadamente, Haensch se limita a mencionar una serie de ejemplos sin ningún comentario: *patrocinar* (más frecuente en España) frente a *auspiciar* (más frecuente en

¹ Cabe mencionar en especial el estudio de Moreno de Alba (1996, pp. 487-501) sobre las denominaciones de una gran cantidad de conceptos en Madrid y en las capitales americanas. Se destaca también el estudio de Molero Pintado (2003), que presenta el resultado de su investigación en un cuadro que registra un gran número de conceptos clasificados según los distintos campos de la vida cotidiana (vivienda, cocina, alimentación etc.) y menciona las denominaciones que le corresponden en 6 países de habla hispana, España, Argentina, Chile, México, Uruguay y Venezuela.

² A esta categoría también alude indirectamente Company Company (2007, p. 31) al describir el tercer tipo de mexicanismo existente en la variante dialectal mexicana como “construcciones compartidas en forma y significado, con el español peninsular, pero que muestran en México una mucho mayor frecuencia de empleo y generalización”.

América), *pequeño*, *-a* (España) frente a *chico*, *-a*, *tardar* (España) frente a *demorarse* (América) *llevar retraso* (un tren, un avión) (España) frente a *estar demorado* (América), etc. (Haensch, 2002, p. 45)³. Mas generalmente, la información brindada por Haensch deja cuestiones pendientes: ¿ocurre en toda América o solo en algunas zonas la mayor frecuencia de un término? ¿Hay países que ofrecen una proporción de uso de ambos términos de la pareja similar a la de España?

La proporción de uso de las distintas voces a través del mundo hispanohablante fue uno de los objetivos del proyecto internacional de geolingüística sobre variación léxica hispanoamericana, *Varilex* (Hueda & Takagaki, 1993), que propuso crear una red internacional de investigación para conseguir un panorama informatizado del léxico urbano vigente en los distintos territorios hispanohablantes de la actualidad. Partiendo de un concepto determinado, por ejemplo “prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados”, se sometió al informante una lista de términos equivalentes, entre los cuales tuvo que escoger el que él mismo usaba (*americana*, *capa*, *chaleco*, *gabán*, *leva*, *saco*, *saco de terno*, *saco de traje*, *terno*, *traje*, *traje de chaqueta*, *vestón*). Se consiguió de este modo una gran cantidad de datos cuya compilación permitió elaborar cuadros y mapas de distribución de uso de las formas en cuestión. *Varilex* resulta muy útil para conseguir una visión panorámica de la variación léxica, pero presenta los defectos

³De hecho, la legitimidad de algunas parejas en la lista se puede cuestionar. La pareja *estrecho*, *-a* (España) frente a *angosto*, *-a* (América) en el apartado 12, ya que también viene evocada en el apartado 11 dedicado a las oposiciones connotativas: en España, *angosto*, *-a* “solo se refiere a determinados lugares (calle, barranco) y es propio del código escrito” (Haensch, 2002, p. 45). En cuanto a la palabra *bochinche*, más frecuente, según Haensch, en América, que *follón*, *lío* o *cacao*, figura en el *Diccionario de americanismos* (2010) con acepciones distintas en América, según la zona donde se usa. Por fin, la oposición entre *lindo* (América) y *bello/hermoso/majo/mono* (España) también se puede cuestionar, primero porque *hermoso* y *bello* también se usan en América, segundo porque existen diferencias denotativas entre los adjetivos que, según Haensch, son de empleo más frecuente en España.

generales de la geografía lingüística: los mapas y atlas, que se limitan a un momento temporal concreto, y seleccionan una red de puntos de un territorio determinado, fueron criticados por su aparente superficialidad, en comparación con los glosarios y monografías elaboradas por los dialectólogos. Además, el proyecto *Varilex* equipara, sin comentarios ni análisis adicionales, parejas de distinta índole: verdaderos geosinónimos (*delantal/mandil*, *ordenador/computador*, *hierro/fierro*), términos con leve matiz denotativo (*bonito*, *hermoso*, *espléndido*), términos con diferencias connotativas (*matar/asesinar/chingar*, *sostén/sujetador/brassiere*). Ahora bien, llama la atención la ausencia, en todos los estudios sobre variación léxica hispanoamericana, de una serie de parejas, *responder/contestar*, *agregar/añadir*, *escoger/elegir*, *alcanzar/lograr/conseguir*, *pedazo/trozo* que, a nuestro parecer, corresponden a los geosinónimos con diferente frecuencia de uso destacados por Haensch. La idea de estas parejas surgió de nuestro estudio comparativo de dos traducciones al español (peninsular y de Colombia) de la obra francesa *L'Etranger* de A. Camus (Weber, 2020), cuyas conclusiones confirmaron las observaciones que hicimos en España y en varios países latinoamericanos: contrariamente a lo que se podría pensar, creemos que hay una diferencia en la frecuencia de uso de ambos términos a ambos lados del océano. ¿Por qué no aparecen estas parejas en los estudios sobre variación diatópica del español? ¿A qué categoría pertenecen? ¿En qué medida su análisis puede ayudar a completar la comprensión de la variación léxica hispanoamericana?

Para conocer el tipo de relación semántica que vincula los términos de estas parejas y su papel en la variación léxica hispanoamericana, uno no puede limitarse a destacar proporciones, sino que resulta imprescindible examinar detenidamente el detalle de dichas proporciones, lo cual requiere un análisis minucioso. Es lo que se va a realizar ahora con la pareja *agregar/añadir*.

2. MÉTODO

El análisis ha de empezar por el examen de los diccionarios de referencia del español, tanto el *Diccionario de la lengua española* (*DLE*) y el *Diccionario de uso del español* (*DUE*), como el *Diccionario de los americanismos* (*DicAm*) que registra las voces o acepciones exclusivas de Hispanoamérica. El *DLE* y el *DUE* revelan que el punto común a *añadir/agregar*, *responder/contestar*, *escoger/elegir*, *alcanzar/lograr/conseguir* y *pedazo/trozo* es la pluralidad de acepciones y campos de aplicación. Para *añadir* y *agregar*, será necesario pues, una vez determinados sus acepciones y campos de aplicación, examinar su empleo efectivo en las distintas zonas del mundo hispanohablante. El objetivo es doble: primero comprobar si se observa cierta regularidad en la repartición de las proporciones (una de las voces más frecuente en España y la otra en América), segundo comprobar si las proporciones dependen del tipo de acepción.

Ya que no estamos en capacidad de realizar una encuesta del alcance de *Varilex*, y tenemos como objetivo la observación del uso real de ambos verbos en el mundo hispanohablante, no tuvimos otra opción que recurrir a un corpus lingüístico. Para que se puedan generalizar los resultados obtenidos gracias a él, el corpus debe ser representativo de una lengua, razón por la cual se escogió para el presente estudio el *Corpus del español del siglo XXI* (*CORPESXXI*). Por una parte, este corpus cubre el período más reciente (2001-2020) y por otra parte, ofrece un equilibrio España/América del 30%-70%, más en adecuación con la proporción de hablantes que el equilibrio 50%-50% del *Corpus de referencia del español actual* (*CREA*)⁴. En *CREA*, numerosos países latinoamericanos se ven poco representados (los

⁴ En este sentido, *CORPESXXI* cumple con los requisitos expuestos por Molina Salinas y Sierra Martínez (2015, p. 309): “Un corpus debería ser una selección de datos lingüísticos reales que constituya una buena aproximación a la realidad de la lengua estudiada, agrupados bajo ciertos criterios que garanticen que todas o algunas variedades lingüísticas estén representadas (dependiendo de la finalidad del corpus) [...]”

de América central principalmente), lo que impide llevar a cabo la investigación en todas las zonas por igual.

Uno de los principales problemas, destacados por Molina Salinas y Sierra Martínez, es que en este tipo de corpus, “la temática o el estilo de ciertos autores podrían condicionar la aparición de algunas formas lingüísticas o giros y la ausencia de otros” (2015, p. 314). Los verbos *agregar* y *añadir*, en particular, aparecen con mayor frecuencia en libros de recetas culinarias, pero no se sabe si están representados en igual proporción en todas las zonas. Sin embargo, pensamos que estos corpus son el medio más adecuado para conseguir sino resultados exactos, por lo menos tendencias, lo cual será el objetivo del presente estudio.

3. EXAMEN DE LAS ACEPCIONES DE *AÑADIR* Y *AGREGAR*

En el *DicAm*, solo figura el verbo *agregar* que, en su empleo pronominal, tiene tres acepciones específicas en República Dominicana y Puerto Rico: “Acurrucarse alguien cómodamente en la cama o en un sofá”, “Acostarse una persona muy pegada a otra”, “Juntarse una persona con otra para convivir sin contrato legal”. Como era de esperar, el *DicAm* no señala el fenómeno de los geosinónimos con diferente frecuencia de uso y, por consiguiente, no alude a la mayor frecuencia del verbo *agregar* en Hispanoamérica.

El examen del *DLE* y del *DUE* revela en cambio que los términos que configuran las parejas mencionadas en el apartado precedente, o bien poseen varias acepciones, o bien poseen una sola, con diferentes campos de aplicación. *Agregar* y *añadir* son dos verbos transitivos⁵, cuyas acepciones se sintetizan en la Tabla 1:

⁵Para el concepto de transitividad, nos referimos a la definición de Bosque y Demonte (1999, p. 1521), inspirada en las teorías de Tesnière: “Verbo transitivo, entonces, es aquel que tiene la capacidad de aparecer con un complemento directo, el cual está regido por el verbo y el cual completa el significado del verbo”.

	RAE	María Moliner
<i>añadir</i>	<ol style="list-style-type: none"> 1. Agregar, incorporar algo a otra cosa 2. Aumentar, acrecentar, ampliar 	<ol style="list-style-type: none"> 1. Juntar una cosa a otra, de modo que forme con ella un conjunto o quede formando parte de ella: “añadir” agua al vino. “añadir” una posdata a la carta’. 1^{era} subacepción: ponerle a una cosa algo más de su misma materia para hacerla más grande: “añadir” un vestido. “añadir” puntos en una labor de media’. 2^{da} subacepción: Poner en un escrito o en un recitado algo que no está en el texto original
<i>agregar</i>	<ol style="list-style-type: none"> 1. Unir o juntar personas o cosas a otras 2. “añadir” algo a lo ya dicho o escrito 3. Destinar a alguien a un cuerpo u oficina, o asociarlo a otro empleado, pero sin darle plaza efectiva 	<ol style="list-style-type: none"> 1. Unir una cosa a otra con la que queda formando un todo: ‘agregar un municipio a otro. Agregar los huevos a la masa’ 2. Refiriéndose a empleados, destinar a alguien accidentalmente a cierto puesto o servicio. 3. “añadir” algo a lo ya expresado

TABLA 1: Aceptaciones de *añadir* y *agregar* en el DLE y en el DUE

Lo primero que llama la atención es el aspecto circular de las definiciones, usual en los diccionarios: en varias de sus acepciones, un verbo de la pareja viene definido por el otro; tal es el caso de la primera acepción de *añadir*, de la segunda acepción de *agregar* en el *DLE* y de la tercera acepción de *agregar* en el *DUE*. Tal circularidad sugiere la conmutabilidad de los dos verbos en ciertos contextos, especialmente cuando *agregar* se refiere a un acto de enunciación (*añadir* algo a lo ya dicho o escrito/ya expresado), y supuestamente cuando *añadir* significa ‘incorporar algo a una cosa’, tal como lo sugiere el *DLE*.

Del examen de las definiciones, se destacan puntos comunes y diferencias en las potencialidades referenciales de ambos verbos.

Ambos verbos pueden ir seguidos de complementos con referentes inanimados. En este caso, expresan la unión de dos seres inanimados A y B, con dos resultados posibles que el *DUE* subraya de manera explícita:

- la formación de un todo, ya sea por incorporación de B a A (“añadir agua al vino”/“agregar los huevos a la masa”), ya sea por asociación de B a A (“agregar un municipio a otro”)
- la formación de una entidad en la que B queda formando una parte de A (“añadir una postdata a la carta”/“poner en un escrito o en un recitado algo que no está en el texto original”)

El *DLE* evoca el proceso de incorporación para *añadir* (“incorporar algo a otra cosa”) y el de unión para *agregar* (“Unir o juntar personas o cosas a otras”) sin aclarar el tipo de resultado conseguido, lo cual deja abiertas las dos posibilidades.

Las diferencias sugeridas por el examen de ambos diccionarios son las siguientes:

- tan solo *añadir* podría, según el *DUE*, referirse a un proceso de ampliación: “añadir puntos en una labor”.
- tan solo *agregar* podría emplearse en el metalenguaje: “agregar una frase, una palabra”/“agregar que...”
- tan solo *agregar* podría recibir complementos con referentes animados, ya sea para expresar la asociación de seres animados en general, ya sea para expresar una asociación de tipo laboral.

De hecho, el examen de las acepciones sugiere una proximidad semántica que deja entrever una posibilidad de conmutación de ambos verbos en numerosos contextos, debido al carácter bastante genérico de la acepción que comparten.

Como era de esperar, una búsqueda en los corpus *CORPESXXI*, *CREA*, y una búsqueda en *Google* revelan una realidad de los usos muy distinta a la que sugiere el análisis teórico de los diccionarios.

En su primera acepción, ambos verbos reciben complementos que se refieren a seres inanimados. Tanto *agregar* como *añadir* pueden remitir a un proceso de unión de un elemento B y de un elemento A que llegan a formar un todo, por incorporación, pero también por simple mezcla:

- (1) Ponga la crema de nuevo en la olla, *añada* la leche, mezcle y sirva con la papa o el pan tostado, según su preferencia (CORPESXXI. Consultado el 16/10/2023).
- (2) La consistencia final debe ser de una miel no muy espesa, puede *agregar* más agua si se requiere (CORPESXXI. Consultado el 16/10/2023).
- (3) Tape y cocine a fuego moderado durante una hora. Sirva después de *agregar* los pimientos morrones y el perejil picado (CORPESXXI. Consultado el 16/10/2023).
- (4) Pasar por una estancia, sazonar y *añadir* el estragón picado. Mantener a 40°C (CORPESXXI. Consultado el 16/10/2023).

Tanto *añadir* como *agregar* pueden expresar la unión de dos seres inanimados sin que ambos lleguen a formar un todo:

- (5) Los usuarios de flickr pueden *añadir* tags a las fotografías que visitan como favoritos y dejar comentarios (CORPESXXI. Consultado el 16/10/2023).
- (6) No debe faltar en este espacio de la casa un sofá, al menos de dos puestos. El mismo debe estar acompañado por una silla individual, tipo poltrona. Y nada mejor que *agregar* cojines, por si aparecen invitados (CORPESXXI. Consultado el 16/10/2023).

Los diccionarios no mencionan la posibilidad, para ambos verbos, de evocar la unión de dos elementos A y B inanimados abstractos⁶. Ahora bien, los corpus brindan numerosos ejemplos de este tipo:

⁶Para estar en capacidad de identificar los complementos con referentes abstractos, nos referimos a la definición bastante sencilla proporcionada por la *Nueva gramática de la lengua española* (2010, p. 210): los sustantivos abstractos “designan cuanto no es material, es decir, acciones, procesos y cualidades que atribuimos a los seres pensándolos como entidades independientes de ellos”. Esta definición, basada en una operación extractiva, coexiste con otras dos en el campo de la lingüística. La definición de tipo ontológico describe al sustantivo abstracto como “el objeto inmaterial cuyo referente escapa a los cinco sentidos” [“l’objet matériel dont le référent échappe aux cinq sens”] (Grevisse, 2001 [1936], p. 452), contrariamente al nombre concreto que “designa un objeto que puede ser percibido por los cinco sentidos o considerado como pudiendo ser percibido por los cinco sentidos” [“désigne un être ou un objet tombant sous nos sens ou considérés comme pouvant tomber sous nos sens”] (Grevisse, 2001 [1936], p. 452). La última definición se basa en un criterio derivacional, es decir puramente lingüístico (Anokhina, 2009, p. 247): todo nombre abstracto se deriva de un adjetivo o de un verbo.

- (7) Por otra parte, al secular hábito de los lugareños de matar rapaces, hay que *añadir* el desmesurado afán coleccionista de los naturalistas de la época [...] (*CORPESXXI*. Consultado el 23/02/2023)
- (8) Para que un sistema de comunicaciones por fibra óptica sea competitivo, no basta con *añadir* las ventajas inherentes al uso de un sistema más rápido de transmisión, sino que [...] (*CORPESXXI*. Consultado el 23/02/2023)
- (9) Para considerar una situación más cercana a la realidad, habrá que *agregar* al beneficio individual –beneficio interno– el beneficio social –beneficio externo– (*CORPESXXI*. Consultado el 23/02/2023)
- (10) Los pasos son: establecer un reto, crear una expectativa positiva, poner a prueba la voluntad, fracasar, *agregar* aún más voluntad y algún lejano día, lograr la meta trazada. (*CORPESXXI*. Consultado el 23/02/2023)

El proceso de ampliación, que supuestamente solo *añadir* puede expresar, también lo expresa *agregar*, como en los ejemplos siguiente:

- (11) Es muy sencillo *añadir* una manga para cambiar el aspecto de un vestido, una camiseta o un top. (Pinterest. s. f. *DIY: añadir mangas a un vestido o a un top*. <https://n9.cl/7nww6>. Consultado el 18/01/2024).
- (12) Cómo *agregar* tela al tejido crochet (Pinterest. s. f. *Cómo agregar tela al tejido crochet*. <https://n9.cl/vblyk2>. Consultado el 18/01/2024).
- (13) Polémica por propuesta de *añadir* una estrofa sobre la paz al himno nacional. (*El País*. 2016, 18 de julio. <https://n9.cl/vh8na>. Consultado el 23/10/2023).
- (14) Sobre todo eso me baso para apoyar la iniciativa presidencial de reformar el himno nacional y *agregarle* una estrofa (*CORPESXXI*. Consultado el 16/10/ 2023).

La unión de personas, y no de cosas, también se puede expresar con ambos verbos:

- (15) La fórmula es bien simple: juntar algunos actores jóvenes y con pinta de mostrencos, *añadirles* un par de chicas monas [...] (*CORPESXXI*. Consultado el 23/02/2023).
- (16) [...] habría estado más que dispuesto a *añadir* a Ruth a su lista de amantes femeninas si ella hubiera querido (*CORPESXXI*. Consultado el 23/02/2023).
- (17) Después pensé que tenía más sentido –y era menos sospechoso– *agregar* a Karen antes que a John ya que había tenido más contacto con ella (*CORPESXXI*. Consultado el 18/10/2023).
- (18) En esta lista puede *agregarse* Gilberto Velázquez, defensor que está con un esguince del tobillo, luego de que venía jugando en la Reserva para ponerse a punto (*CORPESXXI*. Consultado el 18/10/2023).

También en el campo laboral se pueden emplear ambos verbos para referirse a la agregación de personas a ciertos puestos o servicios:

- (19) Desde que Valicourt declinó *agregar* a investigadores afines a Coppens al comité de apoyo al proyecto y siguió hablando en público [...]. (*CORPESXXI*. Consultado el 18/10/2023).
- (20) [...] Laboy dijo que ha hecho acercamientos con legisladores para enmendar la Ley de la Policía para *añadir* a los secretarios de Corrección entre los funcionarios con derecho a escolta cuando finalicen sus términos. (*CORPESXXI*. Consultado el 18/10/2023).

Por fin, contrariamente a lo que sugieren los diccionarios, no solamente *agregar*, sino también *añadir* se pueden emplear en el campo del metalenguaje, con complementos de varios tipos:

- sustantivos que pertenecen al léxico metalingüístico como *palabra*, *frase*, *argumento*, *detalle*, etc.:
- (21) Sin *añadir* una palabra más nos empujó hacia la salida. (*CORPESXXI*. Consultado el 23/02/2023).

- (22) Sin *agregar* una palabra se dio vuelta, caminó hasta la mesa del pelirrojo, y le tiró un puñetazo a la nariz. (CREA. Consultado el 19/10/2023).
- palabras o frase autonómicas⁷, en estilo directo⁸:
 - (23) ¿Y se puede nadar aquí son problema? - averiguó Francisco, que tenía la costumbre de *agregar* un “y” antes de sus preguntas (CORPESXXI. Consultado el 23/02/2023).
 - (24) [...] él estuvo a punto de *añadir*: “Aquí mismo te besé el otro día, cuando abandonaba tu fiesta [...] (CORPESXXI. Consultado el 23/02/2023).
 - (25) “Es lamentable que [...]” dijo Chávez, tras *agregar*: “En todo caso Venezuela, con todo el dolor, actuará en consecuencia [...] (CORPESXXI. Consultado el 23/02/2023).
 - aserciones metalingüísticas en estilo indirecto (proposición subordinada completiva introducida por la conjunción “que”):
 - (26) Se puede *añadir* a lo ya expuesto que en todas las poblaciones bacterianas han surgido cepas resistentes [...] (CORPESXXI. Consultado el 23/02/2023).

⁷ “Se habla de autonimia cuando un signo se refiere a sí mismo como signo, y no al objeto, al mundo” [“On parle d'autonymie quand un signe renvoie à lui-même en tant que signe et non à l'objet, au monde”] (Dubois, Giacomo, Guespin, Marcellesi, Marcellesi & Mével, 2002 [1994], p. 60).

⁸ Como lo explica Rey-Debove (1974, p.113), el discurso en estilo directo es autonímico, contrariamente al discurso en estilo indirecto: “Si se habla en estilo indirecto, se habla del sentido de las palabras de otra persona. Resulta claro, pues que las palabras usadas en este estilo no tienen su habitual denotación, denotan lo que es habitualmente su sentido (...) En estilo directo una proposición denota inmediatamente una proposición; en estilo indirecto, denota un pensamiento.” [Si l'on parle au style indirect, on parle du sens des paroles d'un autre. Il est donc clair que, dans ce style, les mots n'ont pas leur dénotation habituelle, ils dénotent ce qui est habituellement leur sens [...] Au style direct une proposition dénote directement une proposition; au style indirect, elle dénote une pensée.].

- (27) A lo anterior, cabe *agregar* que algunos han sido obsequiados por países con los que tenemos relaciones diplomáticas [...] (*CORPESXXI*. Consultado el 23/02/2023).

Teniendo en cuenta esa posibilidad efectiva de conmutación en todos los contextos y basándonos en la observación del discurso, nuestra hipótesis es que la variación de las proporciones de empleo en España y en América podría relacionarse con el tipo de acepción en juego, acepciones que dependen del tipo de complemento empleado. Clasificamos las acepciones de la manera siguiente:

- En la primera acepción (“Unir o juntar personas o cosas a otras”), los verbos reciben complementos que se refieren a seres inanimados. Estos pueden ser concretos (de aquí en adelante [IN concreto]), abstractos (de aquí en adelante [IN abstracto]), o animados (de aquí en adelante [AN]).
- en la segunda acepción (“añadir algo a lo ya expresado”), ambos verbos reciben complementos que se refieren a un contenido metalingüístico (de aquí en adelante [metalingüístico]).
- en la tercera acepción (“Destinar a alguien a un cuerpo u oficina, o asociarlo a otro empleado, pero sin darle plaza efectiva”), el complemento del verbo siempre se refiere a un ser animado.

Realizamos pues la investigación en *CORPESXXI* clasificando los resultados obtenidos en función del tipo de acepción y, tratándose de la segunda acepción, también en función del valor referencial del complemento del verbo, ya que ambos verbos pueden recibir complementos que se refieren a seres inanimados concretos, a seres inanimados abstractos, o a seres animados.

El cálculo y la clasificación de las ocurrencias según los criterios mencionados supuso el examen minucioso del conjunto de ocurrencias obtenidas por el test de cada uno de los verbos. Aunque fue ardua, la tarea se pudo llevar a cabo sin dificultad en la mayoría de los casos. Tratándose de España, que tiene una representación muy importante en *CORPESXXI* en comparación

con las demás zonas, el número de resultados arrojados resultó considerable para el verbo *añadir* (6779 ocurrencias), con lo cual se limitó el análisis a las primeras mil ocurrencias. Se realizó después el trabajo en las siguientes mil ocurrencias, consiguiendo resultados muy parecidos a los obtenidos en las primeras mil ocurrencias. Se decidió pues aplicar una regla de tres al total de casos obtenidos para cada acepción para conseguir las proporciones correspondientes a las 6779 ocurrencias.

4. RESULTADOS

4.1. RESULTADOS GLOBALES

La búsqueda en *CORPESXXI* de la forma infinitiva de ambos verbos para la zona peninsular revela una preferencia aplastante por el verbo *añadir*, como lo indica la Tabla 2:

	<i>añadir</i> España	<i>agregar</i> España
Número de ocurrencias	6779	700
Proporciones	90,6%	9,3%

TABLA 2: *Proporciones de uso de añadir y agregar en España*

América ofrece unas proporciones globales opuestas a las del español peninsular, aunque menos contrastadas, ya que el verbo menos usado es empleado, sin embargo, en casi un tercio de los casos (Tabla 3):

	<i>añadir</i> América	<i>agregar</i> América
Número de ocurrencias	2984	6580
Proporciones	31,2%	68,8%

TABLA 3: *Proporciones de uso de añadir y agregar en América*

El detalle de los resultados, en función de cada zona, y de cada acepción aparece en las 20 tablas que figuran en los anexos.

4.2. INTERPRETACIÓN DE LOS RESULTADOS

El examen de los 20 cuadros presentados en los anexos revela una serie de puntos comunes en las diferentes zonas. Se observa primero que sea cual sea la zona considerada, la primera acepción resulta siempre la más representada al contrario de la tercera, siempre muy poco frecuente (entre un 0% y un 2%). Este resultado se puede explicar por el carácter más genérico de la primera acepción, constituyendo las otras dos aplicaciones a campos particulares (lenguaje y mundo laboral) de la primera acepción; ese carácter genérico permite que el verbo reciba complementos de todos tipos (entidades inanimadas, concretas o abstractas, entidades animadas), mientras que solo un tipo de objeto puede complementar el verbo en la segunda y en la tercera acepción (contenido metalingüístico y entidad animada respectivamente).

También hay que subrayar que en la primera acepción, los enunciados más numerosos son los que incluyen complementos que se refieren a seres inanimados concretos. A esta regla escapan algunos países, en particular Cuba donde ambos verbos se construyen preferentemente con complementos [IN abstracto]. Nicaragua y Bolivia parecen mostrar la misma tendencia, pero el número de casos obtenidos por el test es lo insuficientemente importante para que se pueda sacar la misma conclusión. En otros países, la tendencia se observa únicamente para uno de los dos verbos: para *agregar* en El Salvador y en Guatemala; para *añadir* en Panamá y Uruguay. Sin embargo, el número de casos es otra vez demasiado reducido en la mayoría de los casos para que se pueda afirmar con certeza. Al fin y al cabo, puede que estas proporciones reflejen cierta realidad de la lengua (uso más frecuente de ambos verbos con complementos [IN concreto] en general, excepto en algunas zonas), o que se expliquen por el tipo de documentos registrados en *CORPESXXI*; es posible que el número reducido de documentos disponibles en algunas zonas y, por consiguiente, el tipo de documentos disponibles no hayan favorecido

la aparición de tal o cual empleo. A lo que se tiene que agregar la gran cantidad de libros de recetas culinarias en todas las zonas, los cuales han de incluir más formulaciones del tipo “añadir/agregar + entidad concreta” que los demás documentos.

Examinemos ahora las diferencias. En una mayoría de países latinoamericanos, las proporciones globales resultan opuestas a las de España: casi todas las zonas de América muestran una clara preferencia por el verbo *agregar* sea cual sea el tipo de acepción. Dos países, sin embargo, se asemejan a España en cuanto a proporciones globales: Cuba (*añadir*: 60,8%; *agregar*: 39,2%) y Puerto Rico (*añadir*: 73,8%; *agregar*: 26,2%).

La principal diferencia entre el español peninsular y el español americano concierne la segunda acepción: mientras que en España se escoge preferentemente *añadir* cuando el complemento del verbo se refiere a un contenido metalingüístico, en la misma situación se escoge preferentemente *agregar* en América. Esto se puede apreciar en la Tabla 4 que solo recoge el número total de casos correspondientes a la segunda acepción y calcula las proporciones de empleo de cada verbo:

	<i>añadir</i>	<i>agregar</i>	Total de casos Acepción 2
España	293 83,5%	58 16,5%	351
Argentina	43 14,6%	251 85,4%	294
Bolivia	22 61,1%	14 38,9%	36
Chile	21 11,6%	160 88,4%	181
Colombia	55 31,6%	119 68,4%	174
Costa Rica	12 30%	28 70%	40
Cuba	69 58%	50 42%	119
Ecuador	18 51,4%	17 48,6%	35

	<i>añadir</i>	<i>agregar</i>	Total de casos Acepción 2
El Salvador	5 17,8%	23 82,2%	28
Guatemala	16 32%	34 68%	50
Honduras	9 22%	32 78%	41
México	142 38%	231 62%	373
Nicaragua	3 13,6%	19 86,4%	22
Panamá	14 41,2%	20 58,8%	34
Paraguay	17 26,1%	48 73,9%	65
Perú	60 43,5%	78 56,5%	138
Puerto Rico	22 53,6%	19 46,4%	41
República Dominicana	9 22%	32 78%	41
Uruguay	10 12,2%	72 87,8%	82
Venezuela	45 35,4%	82 64,6%	127

TABLA 4: *Proporciones de uso de añadir y agregar para la segunda acepción*

En este cuadro, Cuba y, en menor medida, Bolivia, se destacan con una tendencia opuesta, similar a la de España. En Ecuador y Puerto Rico, la diferencia muy reducida y el número de casos también limitado impiden que se saque la misma conclusión.

Otra diferencia concierne, en la primera acepción, el verbo que se emplea en función del tipo de complemento, [IN concreto] o [IN abstracto]. De los cuadros anteriores, se deduce la Tabla 5, que recoge las proporciones de empleo de *agregar* y *añadir* con complementos [IN concreto] y [IN abstracto]:

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Complementos [IN concreto]		Complementos [IN abstracto]	
	<i>añadir</i>	<i>agregar</i>	<i>añadir</i>	<i>agregar</i>
España	65,7%	34,3%	78,4%	21,6%
Argentina	20%	80%	10,7%	89,3%
Bolivia	60%	40%	63%	37%
Chile	29,3%	70,7%	15,5%	84,5%
Colombia	46,4%	53,6%	36,2	63,8%
Costa Rica	23%	77%	27,4%	72,6%
Cuba	65,5%	34,5%	61,1%	38,9%
Ecuador	34%	66%	57,7%	42,3%
El Salvador	37,5%	62,5%	13,9%	86,1%
Guatemala	32,8%	67,2%	26,3%	73,7%
Honduras	38%	62%	18,7%	81,3%
México	39,2%	60,8%	36%	64%
Nicaragua	30,2%	69,8%	35,9%	64,1%
Panamá	20%	80%	61,5%	38,5%
Paraguay	22%	78%	15,8%	84,2%
Perú	28,6%	71,4%	43,2%	56,8%
Puerto Rico	77%	23%	81,4%	18,6%
Rep. Dom.	48,4%	51,6%	37,7%	62,3%
Uruguay	6%	94%	11,5%	88,5%
Venezuela	49%	51%	33,3%	66,7%

TABLA 5: *Proporciones de uso de añadir y agregar con complementos inanimados concretos y abstractos*

Ya señalamos que en una mayoría de zonas, la proporción de enunciados que incluyen complementos [IN concreto] es superior a la de los enunciados que incluyen complementos [IN abstracto], sea cual sea el tipo de verbo empleado. En España, donde se emplea más el verbo *añadir* con cualquier tipo de complemento, se comprueba que *agregar* se emplea aún menos cuando el complemento es de tipo [IN abstracto]. En cambio, en una mayoría de países latinoamericanos, donde la proporción global es inversa (*agregar* más frecuente que *añadir*), se observa que la diferencia es más importante aún en el caso de complementos de tipo [IN abstracto], empleándose preferentemente *agregar*. A esta regla hay que hacer una excepción con Bolivia

y Cuba, donde se emplea más el verbo *añadir*, tanto con complementos [IN concreto] como con complementos [IN abstracto]. La misma tendencia parece darse en Puerto Rico que muestra proporciones muy contrastadas, a pesar de un número total de casos bastante bajo. En Ecuador, esa tendencia parece observarse únicamente con los complementos de tipo [IN abstracto] pero el número reducido de casos y las proporciones poco contrastadas (*añadir*: 57,7%; *agregar*: 42,3%) impiden que se saquen conclusiones definitivas.

5. CONCLUSIONES

El análisis comparativo del empleo de *agregar* y *añadir* en España y en América revela frecuencias de uso globales opuestas en ambas zonas: mientras que en España se emplea preferiblemente *añadir*, una mayoría de países latinoamericanos opta preferiblemente por *agregar*. Se trata pues de una pareja léxica que los estudios de dialectología y de geolingüística tendrían que examinar más detenidamente: coincidimos en la necesidad de trabajos que precisen los valores semánticos según su utilización dialectal.

El examen del detalle de las frecuencias revela matices en la variación España/América según las acepciones y, eventualmente, el tipo de complemento del verbo. La diferencia España/América resulta más visible en la segunda acepción, que se caracteriza por la presencia de un complemento [contenido metalingüístico]: España opta categóricamente por *añadir* cuando América prefiere *agregar*, aunque en esta zona las proporciones de empleo de cada verbo sean menos divergentes que en España en ciertos casos. La otra conclusión concierne la primera acepción, donde se nota que la diferencia España/América es más importante aun cuando el verbo va acompañado por un complemento de tipo abstracto: un complemento de este tipo favorece todavía más el empleo de *añadir* en España, de *agregar* en América.

Por tanto, habría que comprobar si esta pareja verbal es representativa de una categoría léxica no mencionada en los estudios existentes, lo cual permitiría completar y matizar la descripción del apartado de Haensch

(2002, p. 45) dedicado a los *geosinónimos* con diferente frecuencia de uso. Entre las parejas con diferente frecuencia de uso, quizás convenga añadir la subcategoría de *geosinónimos* parciales con varias acepciones, entre las cuales una o algunas favorecen una variación de frecuencia más importante. Sería pues relevante llevar a cabo el mismo tipo de análisis con las demás parejas mencionadas arriba (*responder/contestar*, *escoger/elegir*, *alcanzar/lograr/conseguir*, *pedazo/trozo*, etc.) para averiguar si lo meta-lingüístico y lo abstracto pueden representar puntos de divergencia en la elección de uno u de otro verbo.

Para finalizar, el análisis revela disparidades en América en cuanto al uso de ambos verbos: Bolivia, Cuba, Puerto Rico y, en menor medida, Ecuador se destacan en América por un uso de ambos verbos similar al del español peninsular: uso más frecuente de *añadir* en proporciones generales; uso más frecuente de *añadir* en la primera acepción cuando el complemento del verbo es de tipo [IN abstracto]; uso más frecuente de *añadir* cuando el complemento del verbo es de tipo [contenido metalingüístico] (acepción 2). El análisis de las demás parejas (*responder/contestar*, *escoger/elegir*, *alcanzar/lograr/conseguir*, *pedazo/trozo*, etc.) también permitirá averiguar si esos tres países muestran un comportamiento igualmente diferente en comparación con los demás países latinoamericanos, lo que llevaría a interrogar el porqué de esta variación.

REFERENCIAS BIBLIOGRÁFICAS

- Anokhina, O. (2009). L'abstrait et le concret: approche linguistique de l'abstraction. *Ligeia*, 89-92, 242-249. <https://doi.org/10.3917/lige.089.0242>
- Asociación de Academias de la Lengua Española (2010a). *Diccionario de americanismos* (DicAm). <https://www.asale.org/damer/>
- Asociación de Academias de la Lengua Española (2010b). *Nueva gramática de la lengua española*. Espasa Libros.
- Bosque, I. & Demonte, V. (dirs.) (1999). *Gramática descriptiva de la lengua española*. Espasa Calpe.
- Buesa Oliver, T. & Enguita Utrilla, J. M. (1992). *Léxico del español de América: su elemento patrimonial e indígena*. MAPFRE.

- Company Company, C. (2007). *El siglo XVIII y la identidad lingüística de México*. Universidad Nacional Autónoma de México y Academia Mexicana de la Lengua. <https://n9.cl/1xru2>
- Chacón García, C. (2016). *Análisis demolingüístico del léxico hispánico: estudio aplicado a las "Nociones específicas" del "plan curricular del Instituto Cervantes"* [tesis doctoral. Universidad Nacional de Educación a Distancia]. E-Espacio UNED. <https://n9.cl/3jlf3>
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J. B. & Mével, J. P. (2002 [1994]). *Dictionnaire de linguistique*. Larousse.
- Grevisse, M. (2001 [1936]). *Le bon usage*. Duculot.
- Haensch, G. (2002). Español de América y español de Europa (Segunda parte). *Panace@*, 3(7), 37-64.
- Molero Pintado, A. (2003). *El español de España y el español de América: vocabulario comparado*. Ediciones SM.
- Molina Salinas, C. & Sierra Martínez, G. (2015). Hacia una normalización de la frecuencia de los corpus CREA y CORDE. *Revista Signos*, 89, 308-330. <https://doi.org/10.4067/S0718-09342015000300002>
- Moliner, M. (2004). *Diccionario de uso del español*. Gredos.
- Moreno de Alba, J. (1996). Léxico de las capitales americanas frente al léxico madrileño. *Lexis: Revista de lingüística y literatura*, 20(1-2), 487-501. <https://doi.org/10.18800/lexis.19960102.026>
- Real Academia Española (s. f.). *Corpus de Referencia del Español Actual*. <https://n9.cl/qj2ko>
- Real Academia Española (s. f.). *Corpus del Español del Siglo XXI*. <https://www.rae.es/CORPES/>
- Real Academia Española y Asociación de Academias de la Lengua Española (2024 [2014]). *Diccionario de la lengua española*. Vigésima tercera edición. Actualización 2024. <https://dle.rae.es>
- Rey-Debove, J. (1974). Autonymie et sémiotique littéraire. *Littérature*, 16, 107-116. <https://doi.org/10.3406/litt.1974.2020>
- Ueda, H. & Takagaki, T. (1993). *Varilex. Variación léxica del español del mundo: Mapas y estadísticas*. <https://h-ueda.sakura.ne.jp/varilex/>
- Weber, E. (2020). *El extranjero* de Albert Camus traducido por Pablo Montoya: ¿una traducción para lectores latinoamericanos? *Íkala, Revista de Lenguaje y Cultura*, 25(2), 513-532. <https://doi.org/10.17533/udea.ikala.v25n02a12>

ANEXOS

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	2878 42,4%	1925 28,4%	173 2,5%	1826 26,9%	3 0,05%	6779
Total	73,3%			29,3%	0,05%	
<i>agregar</i>	443 63,3%	156 22,3%	26 3,7%	58 8,3%	1 0,1%	700
Total	89,3			8,3%	0,1%	

TABLA 6: *Proporciones de uso de las distintas acepciones de añadir y agregar en España*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	210 68%	45 14,6%	6 1,9%	43 14%	1 0,3%	309
Total	84,5%			14%	0,3%	
<i>agregar</i>	837 54,3%	376 24,4%	38 2,5%	251 16%	2 0,1%	1540
Total	81,2%			16%	0,1%	

TABLA 7: *Proporciones de uso de las distintas acepciones de añadir y agregar en Argentina*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	33 33,3%	39 39,4%	3 3%	22 21,2%	0 0%	99
Total	75,5%			21,8%	0%	
<i>agregar</i>	22 32%	23 38,6%	5 7,5%	14 20%	0 0%	67
Total	78,1%			20%	0%	

TABLA 8: *Proporciones de uso de las distintas acepciones de añadir y agregar en Bolivia*

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	171 70,2%	38 15,5%	1 0,4%	32 13%	0 0%	245
Total	86,1%			13%	0%	
<i>agregar</i>	412 50,4%	207 25,3%	20 2,7%	160 19%	1 0,1%	817
Total	78,4%			19%	0,1%	

TABLA 9: *Proporciones de uso de las distintas acepciones de añadir y agregar en Chile*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	161 52,1%	80 25,9%	5 5%	55 17,8%	0 0%	309
Total	83%			17,8%	0%	
<i>agregar</i>	186 39,4%	141 29,9%	12 2,5%	119 25%	0 0%	472
Total	71,8%			25%	0%	

TABLA 10: *Proporciones de uso de las distintas acepciones de añadir y agregar en Colombia*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	17 35,4%	17 35,4%	0 0%	12 25%	0 0%	48
Total	83%			17,8%	0%	
<i>agregar</i>	57 39,6%	45 31,2%	1 0,7%	28 19,4%	0 0%	144
Total	71,5%			19,4%	0%	

TABLA 11: *Proporciones de uso de las distintas acepciones de añadir y agregar en Costa Rica*

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	59 29,5%	66 33%	2 1%	69 34,5%	0 0%	200
Total	63,5%			34,5%	0%	60,8%
<i>agregar</i>	31 24%	42 32,5%	3 2,3%	50 38,7%	0 0%	129
Total	58,8%			38,7%	0%	39,2%

TABLA 12: *Proporciones de uso de las distintas acepciones de añadir y agregar en Cuba*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	65 52,8%	30 24,4%	5 4%	18 14,6%	0 0%	123
Total	81,3%			14,6%	0%	167
<i>agregar</i>	126 75,4%	22 13,2%	2 1,2%	17 10,2%	0 0%	
Total	89,8%			10,2%	0%	

TABLA 13: *Proporciones de uso de las distintas acepciones de añadir y agregar en Ecuador*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	21 63%	6 18%	0 0%	5 15%	0 0%	33
Total	81%			15%	0%	100
<i>agregar</i>	35 35%	37 37%	5 5%	23 23%	0 0%	
Total	77%			23%	0%	

TABLA 14: *Proporciones de uso de las distintas acepciones de añadir y agregar en El Salvador*

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	20 35%	15 26,3%	4 7%	16 28%	0 0%	57
Total	68,3%			28%	0%	
<i>agregar</i>	41 31,8%	42 32,5%	8 6,2%	34 26,3%	1 0,8%	129
Total	70,5%			26,3%	0,8%	

TABLA 15: *Proporciones de uso de las distintas acepciones de añadir y agregar en Guatemala*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	46 71,9%	9 14%	0 0%	9 14%	0 0%	64
Total	85,9%			14%	0%	
<i>agregar</i>	75 50,7%	39 26,3%	2 1,3%	32 21,6%	0 0%	148
Total	78,3%			21,6%	0%	

TABLA 16: *Proporciones de uso de las distintas acepciones de añadir y agregar en Honduras*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	201 40%	143 28,5%	12 2,4%	142 28,9%	1 0,2%	502
Total	70,9%			28,9%	0,2%	
<i>agregar</i>	311 36,8%	254 30%	32 3,8%	231 25,2%	0 0%	844
Total	70,6%			25,2%	0%	

TABLA 17: *Proporciones de uso de las distintas acepciones de añadir y agregar en México*

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	13 36%	19 52,7%	0 0%	3 8,3%	0 0%	36
Total	88,7%			8,3%	0%	
<i>agregar</i>	30 33%	35 38,5%	3 3,3%	19 20,9%	0 0%	91
Total	74,8%			20,9%	0%	

TABLA 18: *Proporciones de uso de las distintas acepciones de añadir y agregar en Nicaragua*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	15 30,6%	16 32,6%	2% 4%	14 28,6%	0 0%	49
Total	67,3%			28,6%	0%	
<i>agregar</i>	60 65%	10 10,9%	2 2,2%	20 21,7%	0 0%	92
Total	78,1%			21,7%	0%	

TABLA 19: *Proporciones de uso de las distintas acepciones de añadir y agregar en Panamá*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	39 54,9%	13 18,3%	2 2,8%	17 23,9%	0 0%	71
Total	76%			24,3%	0%	
<i>agregar</i>	139 52,2%	69 26%	7 2,6%	48 18%	0 0%	266
Total	80,7%			18%	0%	

TABLA 20: *Proporciones de uso de las distintas acepciones de añadir y agregar en Paraguay*

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	160 56,3%	48 16,9%	5 1,7%	60 21,1%	0 0%	284
Total	75%			21,1%	0%	
<i>agregar</i>	399 70,7%	63 11,2%	8 1,4%	78 13,8%	1 0,8%	564
Total	83,3%			13,8%	0,8%	

TABLA 21: *Proporciones de uso de las distintas acepciones de añadir y agregar en Perú*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	40 36,4%	35 31,8%	10 9%	22 20%	0 0%	110 73,8%
Total	%			20%	0%	
<i>agregar</i>	12 30,8%	8 20,5%	0 0%	19 48,7%	0 0%	39 26,2%
Total	51,3%			48,7%	0%	

TABLA 22: *Proporciones de uso de las distintas acepciones de añadir y agregar en Puerto Rico*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	46 56,8%	23 28,4%	3 3,7%	9 11,1%	0 0%	81
Total	88,9%			20%	0%	
<i>agregar</i>	49 40,1%	38 31,1%	3 2,4%	32 26,2%	0 0%	122
Total	73,6%			26,2%	0%	

TABLA 23: *Proporciones de uso de las distintas acepciones de añadir y agregar en República Dominicana*

SECCIÓN: LEXICOLOGÍA
LA VARIACIÓN DIATÓPICA DEL LÉXICO: EL CASO DE *AGREGAR* Y *AÑADIR*
EN ESPAÑA E HISPANOAMÉRICA
Elodie Weber

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	15 36,6%	15 36,6%	0 0%	10 24,4%	0 0%	41
Total	88,9%			20%	0%	
<i>agregar</i>	233 54,4%	115 26,9%	6 %	72 16,8%	0 0%	428
Total	82,7%			16,8%	0%	

TABLA 24: *Proporciones de uso de las distintas acepciones de añadir y agregar en Uruguay*

	Acepción 1			Acepción 2	Acepción 3	Número de casos totales
	Ser concreto	Ser abstracto	Ser humano	Contenido metalingüístico	Ser humano	
<i>añadir</i>	217 67,4%	51 15,8%	5 1,5%	45 13,9%	0 0%	322
Total	84,7%			13,9%	0%	
<i>agregar</i>	225 53,4%	102 24,2%	10 2,4%	82 19,5%	1 0,2%	421
Total	80%			19,5%	0,2%	

TABLA 25: *Proporciones de uso de las distintas acepciones de añadir y agregar en Venezuela*



USO DE CORPUS PROPIOS EN LA ENSEÑANZA DEL LÉXICO FORMULAICO EN EL AULA DE ESPAÑOL COMO LENGUA EXTRANJERA (ELE)

USE OF SELF-COMPILED CORPORA FOR FORMULAIC VOCABULARY TEACHING IN THE SPANISH AS A FOREIGN LANGUAGE CLASSROOM (SFL)

Filip Zeman

Universidad Autónoma de Madrid

filip.zeman@uam.es

RESUMEN

Este artículo propone e ilustra con ejemplos una nueva técnica de enseñanza-aprendizaje del español como lengua extranjera basada en el enfoque léxico enriquecido por la creación y consulta de corpus propios. La propuesta se centra a la enseñanza del léxico, en concreto del léxico pluriverbal: partiendo de los avances en la enseñanza de léxico de segundas lenguas y de las aportaciones de la lingüística de corpus, presenta una serie de actividades con los corpus propios diseñadas con diferentes herramientas del sistema de consulta y gestión de corpus *Sketch Engine* para demostrar el potencial didáctico de las estrategias propuestas.

Palabras clave: corpus propio, ELE, enfoque léxico, lenguaje formulaico, *Sketch Engine*, estudiantes avanzados.

ABSTRACT

This article proposes and illustrates with examples a new Spanish as a foreign language teaching technique which is based on the Lexical Approach empowered by the creation and didactic use of self-compiled corpora. The proposed technique is aimed at the teaching of multi-word expressions. Our proposition derives from the latest acknowledgements in the areas of second-language vocabulary teaching and corpus linguistics. We will present a series of activities which are created around the use of self-compiled corpora. The activities make use of the different tools of the corpus manager and text analysis software *Sketch Engine*. Subsequently, the article examines the didactic potential of such activities.

Keywords: self-compiled corpus, SFL, lexical approach, formulaic language, *Sketch Engine*, advanced learners.

Recibido: 01-05-2024
Aceptado: 20-11-2024

DOI: <https://doi.org/10.17561/rilex.8.1.8906>



1. INTRODUCCIÓN

En el ámbito del español como lengua extranjera, la enseñanza del léxico ha experimentado un desarrollo notable mediante la introducción de nuevas maneras de entender y enseñar el léxico, la diversificación de las tareas, la introducción del factor afectivo, el desarrollo del aprendizaje autónomo y la atención a la pluriverbalidad. Una de las propuestas recientes en este ámbito consiste en aplicar los corpus lingüísticos en el aula. Propuestas de su uso en la enseñanza aparecieron junto con el surgimiento de la lingüística computacional como disciplina, pero las funcionalidades limitadas de las herramientas de consulta no permitieron una adaptación eficiente hasta hace relativamente poco. Gracias al desarrollo tecnológico, los corpus modernos se han convertido en sistemas de fácil acceso y consulta, que permiten anotar los datos y realizar computaciones estadísticas complejas de forma automática, y que presentan los resultados de búsqueda de forma clara y resumida al usuario. El presente trabajo sigue esta línea para proponer una nueva técnica de enseñanza del léxico del español con el uso de corpus propios. Teniendo en cuenta los resultados de los recientes estudios sobre la adquisición del léxico y los beneficios de la diversificación metodológica, presentaremos esta nueva técnica junto con una muestra de actividades de clase basadas en el *enfoque léxico* de Michael Lewis. Con ello pretendemos contribuir a que se superen las reticencias sobre el uso directo de los corpus en el aula de lenguas extranjeras. Este trabajo expone las bases teóricas de un futuro estudio empírico, en el que la técnica de enseñanza presentada se pondrá en práctica y en el que se medirá su efectividad.

Para contextualizar el trabajo, resumiremos en la sección 2 algunos de los avances en la enseñanza de léxico de las últimas décadas e introduciremos la teoría del enfoque léxico. También presentaremos la noción de *corpus lingüístico*, y revisaremos la historia y las formas de su aplicación en la enseñanza de lenguas extranjeras. En la sección 3, describiremos las herramientas y funcionalidades del programa *Sketch Engine* usadas en esta propuesta y

presentaremos una muestra de actividades de aula basadas en la consulta de corpus propios. Finalmente, en la sección 4 resumiremos las conclusiones de este estudio.

2. CONSIDERACIONES TEÓRICAS Y REVISIÓN BIBLIOGRÁFICA

Para contextualizar nuestra propuesta, en esta sección presentaremos brevemente su marco teórico y revisaremos los trabajos más relevantes de cada campo temático.

2.1. EL ENFOQUE LÉXICO DE MICHAEL LEWIS

La metodología de la enseñanza del léxico se encuentra hoy en día en una fase consolidada gracias a los métodos comunicativos de enseñanza y, especialmente, gracias al *enfoque léxico*. Este enfoque, que introdujo el lingüista Michael Lewis hace 30 años, supuso un cambio esencial en la percepción del léxico dentro y fuera del aula de lenguas extranjeras. Por primera vez se tomó conciencia de la naturaleza interconectada de las palabras dentro de un idioma. Se prestó atención a las expresiones complejas o *pluriverbales*, incluidas dentro del concepto de *unidades léxicas* (o *chunks* en inglés), tal y como lo interpretó Lewis (2008, p. 7). Con el término *unidad léxica* denominamos las “secuencias de palabras, continuas o discontinuas, que parecen prefabricadas, es decir, que se almacenan y se recuperan de la memoria como un todo, en lugar de generarse desde cero cada vez que se producen o están sujetas a las reglas de la gramática” (Wray, 2002, p. 9). La validez de la teoría de Lewis se confirmó con la llegada de la lingüística computacional. Los corpus permitieron, por primera vez, realizar un análisis estadístico rápido de la expresión lingüística humana y, aunque las estimaciones de las diferentes investigaciones relevantes varían en la extensión, indican que hasta la mitad de nuestra producción diaria es formulaica (Wray, 2002), es decir, “está formada por cadenas de palabras que, por su recurrencia, se recuperan de la memoria como un bloque” (Rufat & Jiménez Calderón, 2017, p. 48). El concepto de lo formulaico es muy abarcador: no solo incluye

las categorías conocidas de unidades léxicas pluriverbales (como las frases hechas o las colocaciones), sino también cualquier combinación de palabras relativamente frecuente y estable.

La pluriverbalidad para un aprendiz de lenguas extranjeras (LE) es clave para su fluidez expresiva y un procesamiento rápido y eficiente del lenguaje (Lewis, 2008, p. 15) tal y como lo procesan los hablantes nativos. Al no estar ocupada en producir combinaciones óptimas de palabras, la capacidad reflexiva del cerebro humano se puede destinar a otras tareas cognitivas –construir el contenido del enunciado, planificar el discurso, etc.–. El almacenamiento de lexis en bloques, más que en palabras sueltas, también corresponde mejor a la estructura del *lexicón mental* (Lewis, 2008, p. 78), el sistema complejo que usamos para almacenar las palabras y para relacionarlas en el plano horizontal (mediante conexiones sinonímicas o antonímicas, por ejemplo) y vertical (mediante vínculos hiperonímicos, meronímicos, etc.). Por último, el conocimiento de los bloques léxicos facilita la aplicación correcta de los principios *de idiomaticidad y selección libre*.

Estos dos principios, introducidos por un contemporáneo de Lewis, John Sinclair, son los que rigen la selección léxica (1991, pp. 109-10). Según la tesis de la selección libre, el hablante dispone de libertad casi absoluta a la hora de hablar y elegir palabras y estructuras (y sus combinaciones); las únicas restricciones que existen son las gramaticales. Sin embargo, los factores idiomáticos restringen estas posibilidades combinatorias en el uso real de la lengua. Por ejemplo, para despedirnos solemos recurrir a expresiones comúnmente usadas en nuestra comunidad lingüística, aunque en un principio la gramática podría generar muchas otras. Los hablantes nativos cultos tienen interiorizadas estas reglas, pero los aprendientes de segundas lenguas no, incluso cuando tienen un conocimiento léxico avanzado, por lo que a menudo *suenan a extranjero*. Al centrar la atención de los aprendientes en la existencia de los bloques léxicos, el enfoque léxico permite superar esta limitación.

Los corpus lingüísticos, que son el medio de nuestra propuesta y que introduciremos en la siguiente sección, aparecen escasamente en el enfoque léxico. A primera vista, este hecho puede parecer contradictorio: el enfoque léxico enfatiza el uso de textos auténticos y los corpus son la fuente más extensa de textos auténticos. No obstante, como detallaremos en la siguiente sección, en la época de la creación del enfoque léxico los corpus estaban empezando a transformarse en digitales, por lo que sus herramientas y funciones aún eran limitadas. Cuando Lewis menciona los corpus, los describe como una fuente útil para crear obras didácticas y lexicográficas de calidad. En otras palabras, Lewis les concibió como una fuente útil para expertos encargados de crear materiales de referencia –manuales de lengua, diccionarios, libros de referencia– (2008, p. 206). En ningún momento plantea su uso como herramientas de consulta manejadas por los aprendientes. Como veremos, hoy en día los corpus han evolucionado en cuanto a la cantidad de su contenido y la facilidad con la se acceden, por lo que se han convertido en herramientas útiles para el aprendiz.

2.2. LOS CORPUS LINGÜÍSTICOS

Los corpus son colecciones representativas de textos, delimitadas en función de unos criterios específicos. Desde su aparición en forma digital se ha abogado por su empleo en las clases de lengua extranjera, por ser una fuente de *input* auténtico.

Se han propuesto dos formas de su uso en el aula: la *directa* y la *indirecta* (Pérez Serrano, 2017, p. 77). La forma *indirecta* consiste en que el docente use los datos de corpus en la fase preparatoria de la enseñanza, para extraer textos o ejemplos de uso auténticos (cuya presencia es esencial en los enfoques modernos de enseñanza) y para recabar datos estadísticos; el uso indirecto incluye también consultas realizadas por el propio docente (en especial cuando no es hablante nativo) para aclarar las dudas que pueda tener sobre un fenómeno léxico o gramatical antes de presentarlo en clase. El uso indirecto no se limita a búsquedas realizadas por los docentes; las editoriales

recurren a los corpus para extraer datos auténticos para sus manuales de ELE. Del mismo modo, el uso indirecto implica no solo la extracción de producciones gramaticales, sino también la recopilación de datos negativos, que son esenciales para identificar problemas persistentes de la producción de los alumnos y para adaptar los materiales didácticos en consecuencia (sean estos los materiales de clase creados por los docentes o los manuales preparados por editoriales). Los corpus de aprendientes –por ejemplo, el *Corpus de aprendices de español como lengua extranjera*, CAES (Instituto Cervantes, 2014)– son una fuente importante de datos negativos.

El uso *directo* implica el manejo de los corpus por parte de los propios aprendientes tanto en el aula como fuera de él. No obstante, a diferencia de la forma indirecta, el uso directo no ha llegado a integrarse en la práctica cotidiana de la enseñanza. Existen varios trabajos que defienden el uso directo en clase, pero se trata de propuestas teórico-descriptivas (como es el caso de Buyse, 2017; Cruz Piñol, 2012; Elvira-García, 2021). Solo un puñado de trabajos proponen desarrollos prácticos en forma de actividades concretas (entre ellos Álvarez Cavanillas, 2017; Higuera, 2009; Pérez Serrano, 2017), pero ni estas propuestas prácticas lograron introducir la consulta directa como una práctica regular en la enseñanza. El manual de ELE *Frecuencias* (s. f.) de la editorial Edinumen puede considerarse como un avance en este sentido¹. Dicho manual está enriquecido con textos auténticos extraídos de los corpus del español actual de la RAE (uso indirecto), pero también incluye en el apéndice de su Guía didáctica actividades que ilustran el uso directo de los corpus de la RAE (Edinumen, s. f.). Se trata sin duda de una publicación pionera en este aspecto. Aun así, el hecho de que el uso directo esté limitado al apéndice indica que sigue siendo un recurso marginal y que el uso predominante es el indirecto.

¹ Agradezco esta observación a uno de los evaluadores.

Como se ve, los corpus (especialmente en su uso directo) no se han convertido todavía en herramientas de uso cotidiano en el aprendizaje de lenguas extranjeras pese a sus indudables beneficios y las numerosas publicaciones que abogan por su uso. Consideramos que esta situación puede deberse a dos factores importantes. Por un lado, la mayoría de los textos académicos que exploran su uso en clase son bastante anticuados (se han publicado durante el auge de la lingüística computacional en los años noventa y al principio del nuevo milenio) y operan con sistemas de consulta de esa época, que, desde el punto de vista de hoy, son muy simples. La mayoría de las propuestas propone actividades de consulta de *concordancias*, que son el resultado de una búsqueda en el corpus limitada a una línea de texto (véase Batten, Cornu & Engels, 1989; Gaskell & Cobb, 2004; Kettemann, 1995; Stevens, 1991, entre otros, como ejemplos de uso didáctico de las concordancias). Aunque las concordancias son una función básica y fácil de usar, las posibilidades de su explotación didáctica son muy restringidas: se limitan a la observación de los fragmentos textuales, reflexión sobre su contenido léxico o gramatical y la utilización de las conclusiones del análisis en alguna actividad adyacente.

Por otro lado, aunque los sistemas de consulta han ampliado sus funcionalidades desde los años noventa, siguen siendo herramientas dirigidas principalmente a expertos en la materia. Por lo tanto, presentan ciertas dificultades de consulta para usuarios con poca o ninguna formación lingüística (que es el caso de la mayoría de los aprendientes de ELE). En este sentido, en *CORPES XXI (Corpus del Español del Siglo XXI)* y en la versión anotada del *CREA (Corpus de Referencia del Español Actual)* de la RAE se ha mejorado sustancialmente la experiencia para el usuario desde la llegada de sus versiones 1.0. Aun así, para realizar una búsqueda simple, el usuario tiene que estar familiarizado, al menos, con los conceptos lingüísticos de *elemento gramatical* y *palabra ortográfica*, *lema* y *forma*. Consideramos que *CORPES XXI* sí tiene potencial didáctico. Para el tipo de contenidos que se abordan este trabajo,

las herramientas especialmente relevantes son *buscar por inventarios* (que permiten ver diferentes formas gramaticales de un lexema) y *coapariciones* (que muestran palabras que se combinan frecuentemente con el elemento buscado, de forma parecida a la de algunas de las herramientas de *Sketch Engine* que describiremos en la sección 3).

Los sistemas modernos de consulta y manejo de corpus, como *Sketch Engine*, con el que trabajaremos en la propuesta, dan un paso más allá. Disponen de una interfaz fácil de manejar, permiten recuperar los resultados de búsqueda en diferentes formatos (textual y gráfico) y disponen de una amplia gama de herramientas de consulta que facilitan el diseño de todo tipo de actividades. Presentaremos *Sketch Engine* con más detalle en la sección 3.

Hay que reconocer que la explotación didáctica incluso de los corpus más modernos presenta varios retos importantes, pero estos son superables en nuestra opinión. Como con cualquier forma nueva de trabajo, se requiere una introducción previa por parte del profesor. Naturalmente, esta introducción debe incluir los aspectos técnicos de su manejo —Boulton (2010, 2012) ha demostrado que para esto es suficiente con una breve introducción—, pero no puede limitarse únicamente a estos aspectos. Según el tipo de corpus y la clase de consulta que se haga, se pueden recuperar demasiados datos o demasiado pocos, y los estudiantes pueden encontrarse incluso con expresiones incorrectas o incompletas (Pérez Serrano, 2017, p. 80). Especialmente con usuarios novatos, la supervisión del profesor en la fase de análisis de datos y formulación de conclusiones es esencial para asegurar que las conclusiones que sacan sean correctas y para que no se desmotiven por la cantidad de datos que no saben interpretar. La capacidad de reflexión es especialmente importante con el uso de los corpus basados en la web, que, al tener muy limitada la capacidad de determinar la procedencia y la autoría de los textos, incluyen producciones de hablantes no nativos y muestras de lenguaje muy descuidado que pueden darse entre hablantes nativos (por ejemplo, en

foros o blogs, donde es posible encontrar muestras donde la gramática y la selección léxica estén algo descuidadas), pero que no es apropiado en la enseñanza del español como segunda lengua. Precisamente por la importancia de este tipo de capacidad reflexiva, consideramos que las actividades con corpus son más recomendables para estudiantes con niveles avanzados del idioma, que disponen de un criterio lingüístico desarrollado y saben interpretar los datos correctamente.

2.3. LOS ESTUDIANTES AVANZADOS

Se pueden considerar como *estudiantes avanzados* los discentes en la “etapa avanzada-superior o de uso competente de la lengua” (Instituto Cervantes, 2006), que en general corresponde a los niveles C1 y C2 del *Marco común europeo de referencia* (MCER) (Consejo de Europa, 2001). Si nos limitamos al área del léxico, podemos fijar dos criterios esenciales para medir el nivel de semejanza de la expresión discente a la expresión nativa: la *amplitud léxica* (el número de unidades léxicas que conoce el alumno) y la *profundidad léxica* (el conocimiento de las relaciones que existen entre una unidad y otra). A estos dos factores podemos añadir otros dos: la *precisión léxica* (capacidad de diferenciar entre sinónimos gracias al conocimiento de los matices de significado que los distinguen) y la *adecuación léxica*, que se consigue cuando se elige la voz más apropiada diacrítica y diafásicamente (Capel, 2012, p. 12).

A diferencia de los niveles inferiores (A1-B2), las características del vocabulario de los estudiantes avanzados están determinadas no solo por la instrucción en la LE, sino también por el nivel formativo general y por diversos factores socioculturales (que incluyen los intereses académicos y profesionales). La mayoría de los aprendientes no suelen llegar al nivel de los hablantes nativos, aunque en algunos casos esto puede ocurrir: en áreas específicas relacionadas con los intereses profesionales o académicos específicos del aprendiente, este puede llegar a sobrepasar el conocimiento de un hablante nativo no experto.

El mayor problema de la expresión léxica de estudiantes avanzados es que, incluso cuando alcanzan un tamaño de vocabulario de nivel nativo, frecuentemente conservan rasgos no nativos en su expresión, por lo que siguen *sonando a extranjero*. Esto se debe a un nivel más bajo de la profundidad léxica y un control deficiente de la idiomática. Estas características del conocimiento léxico de estudiantes avanzados se confirmaron, por ejemplo, en el estudio de Bogaards (2001), que reveló que los estudiantes de LE que en algunos casos superaron a sus pares nativos en el tamaño del léxico, se quedaron atrás en lo que se refiere al léxico coloquial, expresiones con carga cultural y el lenguaje formulaico. La asimilación de las reglas que operan la selección léxica nativa (y que determinan, hasta cierto punto, las características de las expresiones idiomáticas) es más difícil de conseguir con los métodos convencionales si la comparamos con el aprendizaje cuantitativo de léxico. Se consigue con la exposición de los estudiantes a la lengua a través de la lectura o el uso de materiales audiovisuales, siempre que el docente haga un esfuerzo adicional para que el aprendizaje no sea pasivo, incidental y, por ende, lento. A este grupo podemos también añadir el trabajo con los corpus, que tiene la ventaja de estimular una reflexión más consciente sobre las propiedades de las unidades léxicas.

2.4. LOS CORPUS PROPIOS

Como hemos adelantado, este estudio se centra en el uso de *corpus propios*, que en este trabajo definiremos como ‘colecciones de textos ajenos recopiladas por los propios alumnos de LE, quienes las usarán como parte de su aprendizaje’.

Los corpus propios pueden documentar eficientemente el uso de la lengua en un contexto muy acotado. Puesto que el temario de los cursos de idiomas extranjeros casi siempre se compone de temas acotados (correspondientes a las distintas unidades didácticas y campos nocionales), un corpus propio puede aportar al aprendiz datos más relevantes que un corpus general. En un corpus propio, los datos recogidos se pueden acotar en función de

numerosas variables durante el propio proceso de compilación y no *a posteriori*: el formato, el registro, los factores relacionados con su procedencia (autor o autores, lugar de producción y publicación, etc.), a veces incluso su finalidad. Esto permite un mayor nivel de control sobre el tipo de textos que serán recogidos y posteriormente analizados y evita problemas de interpretación de datos (contaminación de datos con producciones no cultas o no nativas a las que hemos aludido en la sección anterior, por ejemplo). Un control de este tipo no sería posible con las herramientas de filtración de los corpus convencionales.

Ya hemos comentado que solo existen unos pocos trabajos dedicados al uso directo de corpus en la enseñanza de segundas lenguas. Ninguno de estos trabajos se centra en los corpus propios. Los estudios que sí lo hacen analizan su uso en otras áreas afines a la enseñanza de lenguas extranjeras, como la traducción o la escritura académica. El artículo de Zhao y Shi (2015), por ejemplo, examina la creación de un corpus paralelo propio para la enseñanza de la traducción. En el estudio presentado en Pérez-Carrasco (2023), se diseñó y se empleó una secuencia didáctica para estudiantes de la traducción científico-técnica. Los alumnos compilaron sus propias colecciones de textos que después usaron en la traducción de textos científico-técnicos. Los trabajos de Tribble y Wingate (2013), Lee y Swales (2006) y Charles (2014) tratan sobre la escritura académica inglesa (*EAP: English for Academic Purposes*). Maia (2005) y Zanettin (2002) experimentaron con el uso de corpus propios en el aula de lenguas extranjeras, pero a diferencia de nuestro estudio no propusieron ninguna actividad concreta centrada en su uso. Los corpus propios en su investigación sirven como una herramienta de apoyo para realizar actividades convencionales, y en su uso se parecen a consultas de los diccionarios. El estudio de Smith (2011, 2020) se parece al nuestro en que se centra en las lenguas extranjeras, hace uso de corpus propios y utiliza *Sketch Engine*. No obstante, no propone actividades concretas. Hasta donde sepamos, la técnica que presentaremos en el siguiente apartado sería la primera de este tipo.

3. PROPUESTA DIDÁCTICA DE USO DE CORPUS PROPIOS

Basándonos en las características de la adquisición del léxico por parte de aprendientes avanzados que hemos revisado brevemente en la sección 2, hemos desarrollado una serie de actividades ejemplares con las que intentamos demostrar el potencial didáctico de nuestra técnica basada en el enfoque léxico y la consulta de los corpus propios. A continuación, presentaremos esta propuesta de una manera general y más bien teórica, centrándonos en las posibilidades que ofrecen los corpus propios y el sistema de consulta y gestión de corpus *Sketch Engine*, del que haremos uso para elaborar las diferentes actividades.

Sketch Engine fue creado en 2003 por la compañía Lexical Computing y se diseñó en colaboración con el equipo de *Natural Language Processing Centre* de la Universidad de Masaryk de Brno. A diferencia de los bancos de datos de la RAE que hemos mencionado en los apartados anteriores (*CREA* o *CORPES XXI*), *Sketch Engine* no es un corpus sino una *herramienta de consulta y gestión de corpus*. Como tal, permite consultar una multitud de corpus monolingües (el *BNC* o los corpus web *TenTen*, entre muchos otros) y corpus paralelos (por ejemplo, *Europarl*) a través de una interfaz de uso fácil y con una amplia gama de opciones de búsqueda. A efectos de este trabajo, es especialmente importante el hecho de que ofrezca la posibilidad de crear corpus propios, basados en textos en línea (con enlaces URL) o en archivos de texto subidos por el usuario. Los corpus propios se pueden compilar en más de 140 lenguas, pero el tipo de anotación automática textual que se ofrece es diferente para cada lengua. Para el español están disponibles todas las funciones importantes, entre ellas las siguientes:

1. Anotación morfosintáctica (*POS tagging*): un etiquetador automático (*tagger*) asigna a cada palabra una categoría gramatical y otros rasgos morfosintácticos –género, número, tiempo verbal, etc.– (Kilgarriff, Rychly, Smrz & Tugwell, 2004, p. 109)

2. Análisis sintáctico (*parsing*): es un proceso realizado por un programa automático (*parser*), que clasifica las unidades léxicas según su función sintáctica en la oración.
3. Lematización: cada palabra gráfica se relaciona con el lema que le corresponde.
4. Compilación automática de *Word Sketches* que resumen el comportamiento contextual de la palabra dada: identifican las unidades léxicas que la acompañan en diferentes posiciones sintácticas (Kilgarriff, Rychly, Smrz & Tugwell, 2004, p. 107).

Al tratarse de herramientas automáticas, no son necesarios conocimientos de lingüística computacional por parte del usuario, pero sí conviene que tenga conocimientos lingüísticos básicos y cierta capacidad de reflexión, porque estos programas en ocasiones se equivocan. En el caso del español, por ejemplo, es común que etiqueten como objeto directo un sujeto pospuesto; también se pueden confundir las categorías gramaticales de palabras homógrafas, por ejemplo. El usuario debería ser capaz de identificar estos fallos cuando analiza los datos.

A diferencia de otros sistemas de consulta, *Sketch Engine* clasifica las coapariciones de palabras (*Word Sketches*) según la función sintáctica que cumplen las palabras recuperadas con respecto al término de búsqueda (por ejemplo, si el término de búsqueda es un verbo, registra los sujetos y los objetos directos que aparecen con él en el contexto). En este sentido, *CORPES XXI* solo las clasifica según su categoría gramatical: por ejemplo, presenta los sustantivos que se combinan con el verbo sin diferenciar su función sintáctica con respecto a ese verbo. Otra ventaja que tiene *Sketch Engine* consiste en la métrica que usa para calcular la fuerza del vínculo en las coapariciones, que se llama *LogDice*. A diferencia de las fórmulas que –como MI score (*Mutual Information*, véase Gablasova, Brezina & McEnery, 2017), usado en *CORPES XXI*– solo tienen en cuenta la frecuencia de coaparición (comparan el número de veces que las palabras aparecen juntas con el número de veces que aparecen

por separado) y por tanto tienden a sobrevalorar algunas combinaciones hápax o erróneas, *LogDice* incluye en el cálculo la relación sintáctica entre las palabras, lo que se traduce en un cómputo más exacto. Volveremos sobre este tema en la sección 3.1.

Antes de usar cualquier herramienta tecnológica hay que introducirla debidamente. Este es el caso de los corpus digitales en general: en los estudios que hemos citado en la sección introductoria (Boulton, 2010, 2012), se hacía una introducción breve (de 5 a 10 minutos) sobre su manejo. *Sketch Engine* es un sistema más complejo que el que usó Boulton, por lo que en su caso la presentación previa debería ser algo más extensa. Como ha advertido uno de los evaluadores, el nivel de especialización de *Sketch Engine* puede ser más alto que el de *CORPES XXI* en al menos algunos aspectos, por ejemplo, en que *Sketch Engine* hace uso del lenguaje CQL (Corpus Query Language) para las búsquedas avanzadas. Por eso recomendamos que las consultas iniciales en *Sketch Engine* en el aula se limiten a las búsquedas categorizadas como básicas, en las que el usuario introduce solamente el término buscado y todos los parámetros adicionales se determinan de forma automática. En *CORPES XXI* no es posible este tipo de búsqueda simplificada. Consideramos que una introducción a las herramientas de *Sketch Engine* en versión básica se podría hacer en una hora, y que se podría segmentar en explicaciones breves de 5-10 minutos sobre cada una de las herramientas, realizadas en diferentes momentos del curso.

Para que las actividades que proponemos en este apartado se puedan llevar a cabo, los corpus propios que se usen deben ser suficientemente grandes, lo que es especialmente importante para las herramientas *Word Sketch*, *Word Sketch Diferencial* y *N-gramas*. Aunque resulta difícil fijar una cifra mínima por la cantidad de variables que existen, basándonos en las pruebas que hemos hecho, el tamaño mínimo recomendable sería de unas 20 000 palabras para temas muy acotados. Aunque el número parezca grande, gracias a la opción de compilar los corpus usando textos en línea (copiando

sus *URL*), la compilación es relativamente rápida: una estimación realista sería entre 10 y 15 minutos por un corpus temático.

En el siguiente subapartado (3.1.) introduciremos brevemente las herramientas de *Sketch Engine* de las que haremos uso, en la sección 3.2. describiremos el corpus propio de muestra que hemos compilado, y en la última sección (3.3.) presentaremos las actividades agrupadas según la herramienta usada en cada caso.

3.1. HERRAMIENTAS DE SKETCH ENGINE USADAS EN LA PROPUESTA

En este apartado describimos las herramientas *Concordancia*, *Word Sketch*, *Word Sketch Diferencial* y *N-gramas*².

Las concordancias son la función más básica de cualquier herramienta de consulta de corpus. Como hemos explicado en la sección 2.2., las concordancias son fragmentos de texto limitados a una línea, que son el resultado de una búsqueda en el corpus. Tras la definición de los parámetros de búsqueda, se presentan todas las apariciones de la expresión buscada en el formato KWIC (*Key Word in Context*) el elemento buscado aparece en el centro de la línea de texto (véase Figura 1)) o como frases sueltas. En la Figura 1 reproducimos un fragmento de las concordancias para el lema *violencia* a partir del corpus de muestra que presentaremos en el siguiente apartado.

La herramienta *Word Sketch* permite visualizar en forma de listas las palabras que cumplen ciertas funciones sintácticas con respecto al término buscado: sujeto, objeto, modificador, predicado selector, etc. La cantidad y tipo de datos que se muestran depende del tamaño del corpus y la categoría gramatical del término buscado. Por ejemplo, para los verbos y adjetivos se identifican como modificadores los adverbios, y para los sustantivos los adjetivos. La Figura 2 contiene el resumen de *Word Sketch* generado para el lema *denunciar*.

² En este trabajo usaremos la denominación oficial española de las herramientas de *Sketch Engine*. Únicamente adoptaremos nuestra propia traducción para la herramienta *Word Sketch Difference* –a la que nos referiremos como *Word Sketch Diferencial*– porque consideramos que la traducción oficial (*Diferencia Sketch*), no refleja de manera precisa la funcionalidad de la herramienta.

SECCIÓN: E/L2
USO DE CORPUS PROPIOS EN LA ENSEÑANZA DEL LÉXICO FORMULAICO EN
EL AULA DE ESPAÑOL COMO LENGUA EXTRANJERA
Filip Zeman



FIGURA 1: Fragmento de concordancias para el lema violencia

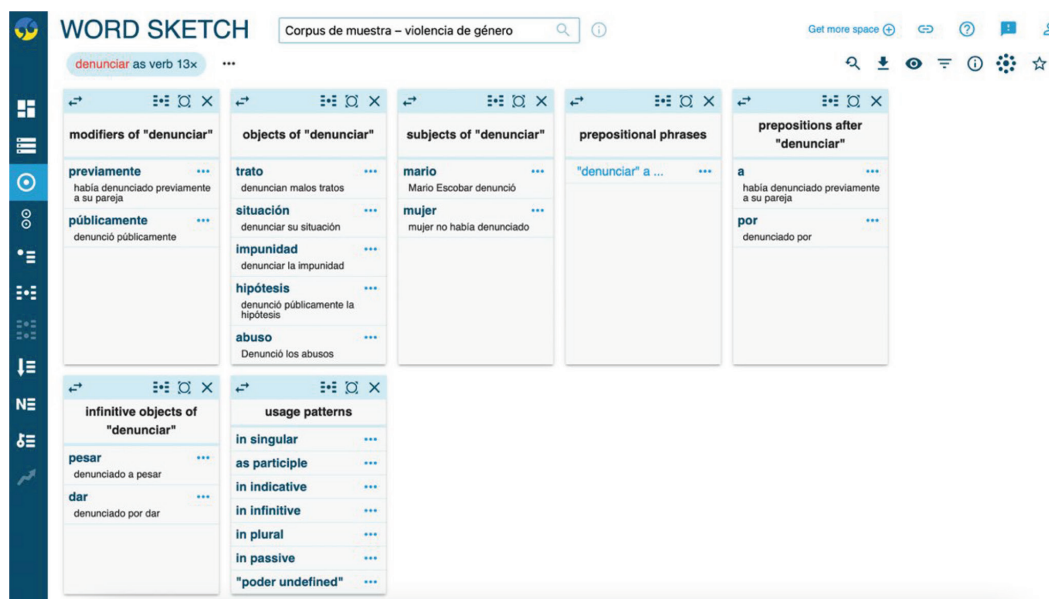


FIGURA 2: Resumen de Word Sketch para el lema denunciar

Además de listados de palabras, en *Word Sketch* se pueden ver representaciones gráficas, parecidas a mapas conceptuales, en las que distintas funciones sintácticas están incluidas en zonas marcadas con colores diferentes y donde también están gráficamente representadas la *frecuencia absoluta* de estas

combinaciones (cuántas veces en total aparecen en los datos) y la fuerza del vínculo existente entre dos palabras que coaparecen en el contexto (el colocativo y la base). Esta fuerza se mide a través del indicador *LogDice*, que ya hemos mencionado. *LogDice* compara la frecuencia de coaparición del colocativo y la base unidos por una relación sintáctica específica con la frecuencia de sus combinaciones con otras palabras (con la misma o distinta relación sintáctica). En la representación gráfica de *Word Sketch*, la fuerza de la coaparición está reflejada a través de la distancia entre la base y el colocativo³. La frecuencia absoluta se representa por el tamaño del círculo que rodea a los colocativos. La información que aportan *LogDice* y la frecuencia absoluta es muy útil para el aprendizaje del lenguaje formulaico y las expresiones idiomáticas. Con datos de la frecuencia absoluta de una expresión los estudiantes pueden deducir si es una expresión de uso real y si es común. El indicador *LogDice*, por su parte, les informa sobre el grado de fijación sintáctica de una combinación de palabras. En el caso de una expresión fija, el trabajo con corpus puede evitar que el alumno produzca combinaciones léxicas poco adecuadas.

Como ilustración, haremos una consulta de Word Sketch en el corpus web del español *esTenTen* de la palabra *caballo*. Entre las combinaciones con la frecuencia absoluta más alta encontramos, por ejemplo, *caballo negro* –frecuencia absoluta de 2500 casos– (Lexical Computing Ltd., 2021). A pesar de su frecuencia, no es una expresión fija: tanto la base como el colocativo pueden ser sustituidos por otras unidades (por ejemplo, como en *perro negro* o *caballo grande*). Esto también se refleja en la puntuación baja de *LogDice* de *caballo negro*: 5.2 (Lexical Computing Ltd., 2021). Por su parte, la expresión *caballo ganador* que aparece también en los resultados tiene una frecuencia absoluta parecida a la de *caballo negro*, pero su fijación interna es mucho mayor: en este caso, *LogDice* –8.2– (Lexical Computing Ltd., 2021)

³ En los listados de *Word Sketch*, el *LogDice* (o fuerza de coaparición) se indica de forma directa, a través de un número (hasta 14).

le informa al aprendiente de que se trata de una combinación relativamente estable que no admite mucha variación (no son comunes las combinaciones *perro ganador* o *caballo perdedor*, por ejemplo).

La herramienta *Word Sketch Diferencial* permite comparar los resúmenes de *Word Sketch* de dos palabras para ver en qué se parece su comportamiento contextual y en qué se diferencia. En la Figura 3, reproducimos la comparación de *Word Sketch Diferencial* para los sustantivos *cuerpo* y *cadáver*. La combinatoria de *cuerpo* está marcada en verde, la de *cadáver* en rojo, y en las franjas intermedias aparecen los elementos contextuales que estas palabras tienen en común.

Por último, la herramienta *N-gramas* facilita la búsqueda de expresiones pluriverbales, que a menudo no son composicionales. Estas se presentan en una lista junto con su frecuencia absoluta. La búsqueda se puede acotar por la extensión de la expresión (por el número de palabras que la forman) y otros criterios (como la secuencia de letras que debe contener la expresión). La Figura 4 contiene expresiones estables de tres y cuatro palabras detectadas en el corpus de muestra.

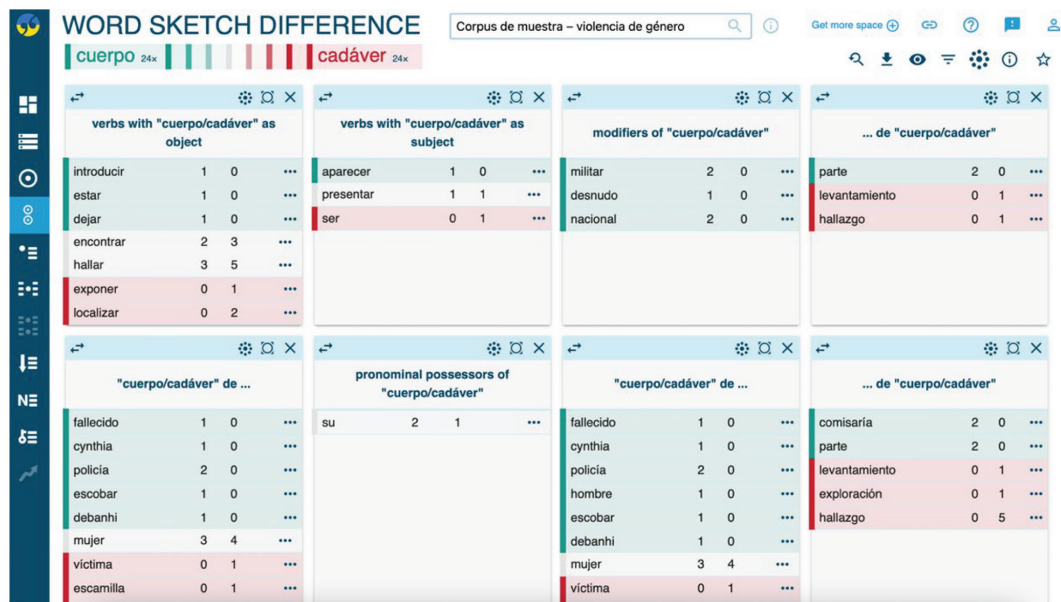


FIGURA 3: Comparación de la combinatoria de *cuerpo* y *cadáver* a través de *Word Sketch Diferencial*

N-GRAMS Corpus de muestra – violencia de género Get more space

3–4-grams, word (Items: 2,758 , total frequency: 7,814) 🔍 ⬇️ 👁️ ⚙️ ⌚ ☆

Word	Frequency ?	Word	Frequency ?	Word	Frequency ?
1 violencia de género	35 ...	18 a las víctimas	14 ...	35 tiene derecho a la	10 ...
2 de la violencia	24 ...	19 las víctimas de violencia	13 ...	36 de EL PAÍS	10 ...
3 la Guardia Civil	23 ...	20 de la Guardia	12 ...	37 la Policía Nacional	10 ...
4 Toda persona tiene	22 ...	21 en caso de	12 ...	38 contra las mujeres	10 ...
5 de violencia de	22 ...	22 el caso de	11 ...	39 por parte de	10 ...
6 tiene derecho a	22 ...	23 víctimas de violencia de	11 ...	40 en la factura	10 ...
7 persona tiene derecho	20 ...	24 de una mujer	11 ...	41 en el que	10 ...
8 de violencia de género	20 ...	25 el presunto autor	11 ...	42 a las víctimas de	10 ...
9 Toda persona tiene derecho	19 ...	26 de las mujeres	11 ...	43 atención a las víctimas	10 ...
10 de la mujer	18 ...	27 de la Policía	11 ...	44 atención a las	10 ...
11 de la víctima	18 ...	28 de atención a	11 ...	45 caso de violencia	10 ...
12 violencia de pareja	18 ...	29 de la Guardia Civil	11 ...	46 la violencia contra las	9 ...
13 de violencia machista	16 ...	30 derecho a la	11 ...	47 rastro en la factura	9 ...
14 persona tiene derecho a	16 ...	31 de Debanhi Escobar	11 ...	48 rastro en la	9 ...
15 la violencia de	15 ...	32 la violencia de pareja	11 ...	49 de arma blanca	9 ...

FIGURA 4: Expresiones estables de 3 y 4 palabras detectadas por N-gramas

Todas estas herramientas proporcionan la frecuencia (absoluta y/o relativa) de aparición de los resultados y otros datos estadísticos, indicativos, por ejemplo, de la fuerza de la relación contextual (*LogDice*) a la que nos hemos referido en la descripción de la herramienta *Word Sketch*.

3.2. CORPUS DE MUESTRA

Para ilustrar la aplicación de corpus propios en la enseñanza de ELE se ha compilado un corpus de muestra a través de *Sketch Engine* (opción “Buscar textos en la web”). Se ha elegido un tema de actualidad que podría tratarse en actividades ELE de nivel avanzado: la violencia de género. Se han recopilado textos de distintas fuentes, entre ellas artículos publicados en *El País*, *El Mundo*, *RTVE* y textos informativos del Gobierno de España. El corpus de muestra contiene un total de 30 000 palabras.

La creación de un corpus propio con textos sobre el tema tratado es el paso previo para la realización de las actividades que presentaremos a continuación. Puede ser un corpus diferente para cada alumno, pero parece más recomendable que sea el mismo para todo el grupo, basado en los mismos

archivos de texto o en las mismas páginas web. Así se agiliza la compilación y el profesor mantiene un mayor control sobre el trabajo de los alumnos. Además, *Sketch Engine* permite descargar y compartir un corpus propio, que se puede usar en su versión original por todos los alumnos o ser personalizado (por ejemplo, añadiendo materiales adicionales) en función de las actividades posteriores que se quieran plantear.

3.3. EJEMPLOS DE APLICACIÓN DIDÁCTICA DE LAS HERRAMIENTAS DE SKETCH ENGINE

En esta sección presentaremos propuestas concretas de aplicación didáctica de *Sketch Engine*. Como se verá, las tareas diseñadas a veces tratan aspectos relacionados con la organización del léxico. Por este motivo, y para no saturar a los alumnos con información sobre el manejo de los corpus, conviene distribuir el uso de las distintas herramientas entre temas o unidades didácticas diferentes. Para afianzar el aprendizaje de clase, recomendamos integrar este tipo de tareas en el trabajo autónomo de los alumnos (por ejemplo, los deberes de casa) y realizar su seguimiento individual.

3.3.1. Concordancias

Se pueden proponer dos formas de uso de las concordancias, a las que nos referiremos como *uso auxiliar* y *uso orientado*. El *uso auxiliar* consiste en el uso de las concordancias como una herramienta de consulta en actividades de clase no relacionadas con los corpus, como sustituto del diccionario. El corpus es una buena alternativa para los diccionarios monolingües y bilingües porque contiene muchos más datos en la lengua meta, como ejemplos y contextos de uso. Por ejemplo, si un alumno quiere ver con qué preposiciones se combina un verbo de régimen preposicional (p.ej., *dependen*), en Google tiene que introducir *dependen* (o “*dependen* + preposición”) y después revisar multitud de resultados para encontrar el dato deseado. En cambio, en las concordancias ve directamente los ejemplos de uso con la preposición. La búsqueda se puede afinar aún más con las opciones avanzadas, como *filter context* (filtrar el contexto) combinado con las etiquetas de categoría sintáctica (*POS context*). Así, se podrían buscar todos los ejemplos en los que

el verbo va seguido de una preposición dentro de un marco de proximidad (a una o más palabras de distancia con respecto al verbo). Del uso auxiliar ya han hablado varios autores (véase la sección 2), por lo que no vamos a dedicarle más espacio.

Las *tareas de uso orientado* se articulan alrededor de las concordancias. Se trataría especialmente de actividades que de alguna forma implican el análisis del léxico en el contexto: ejercicios de rellenar huecos, ejercicios léxico-gramaticales (p.ej., sobre cambios de significado asociados con los verbos *ser* y *estar*), etc. Para dar un ejemplo más concreto, nos detendremos aquí en las actividades de rellenar huecos, en las que se presenta un texto con algunos fragmentos omitidos e incluidos en una lista aparte. Sería recomendable elegir un texto que tenga el mismo tema que el corpus propio, pero que no esté incluido en dicho corpus. Conviene también que las palabras omitidas sean específicas del tema tratado y que los alumnos no las conozcan (es importante que entre ellas haya unidades léxicas pluriverbales). Como primer paso, se pide a los alumnos que elijan palabras correspondientes a cada hueco basándose en el texto de la tarea y, si tienen dudas, pueden recurrir a las concordancias como segundo paso. Tras buscar en concordancias una palabra o unidad léxica de la lista, se presenta al alumno un volumen del *input* mucho mayor. Resulta más fácil deducir el significado de las unidades léxicas desconocidas a partir de múltiples ejemplos de uso para luego poder posicionarlas dentro del texto de la tarea.

La segunda tarea que describimos es de tipo productivo y tiene como objetivo la consolidación de las estrategias de uso de los corpus por parte de los aprendientes. Se pide a los alumnos que redacten un texto de un género específico (por ejemplo, un artículo de opinión) en el que aparezcan determinadas unidades léxicas relacionadas con el tema asignado. Estas unidades léxicas (entre tres y cinco para cada estudiante) se asignan o se distribuyen por sorteo. Para usarlas adecuadamente, los alumnos tendrán que analizar

su uso. Para palabras previamente conocidas a nivel pasivo, es una manera de desarrollar la profundidad léxica, porque a través de las concordancias se aprenden nuevos matices semánticos y de uso.

3.3.2. *Word Sketch*

Como se ha explicado en la sección introductoria de este bloque, la herramienta *Word Sketch* compila listados de las palabras que coaparecen frecuentemente con la voz buscada. Antes de proponer actividades con *Word Sketch*, el docente debe ofrecer una breve introducción al manejo de esta herramienta. La propuesta que presentamos en esta sección hace uso de los mapas conceptuales para reflejar la combinatoria de las unidades léxicas. Este método de trabajo se promueve dentro del enfoque léxico y está justificado por las bases cognitivas de la organización y el procesamiento del léxico mental, que funciona como un sistema relacional complejo.

Para preparar la actividad que presentaremos a continuación, el docente debe elegir entre tres y cinco palabras (bases) que tengan cierta frecuencia en el corpus propio y que pertenezcan a una de las principales categorías gramaticales léxicas (sustantivos, verbos, adjetivos y adverbios), que son las que tienen mayor carga semántica y suelen tener restricciones de selección claramente definibles. Para hacerlo, el profesor puede usar otra función de *Sketch Engine*, *Word List* (listado de palabras), que extrae por orden de frecuencia absoluta todas las palabras (en nuestro caso, lemas nominales, adjetivales y verbales) que aparecen en el corpus. Para ilustrar nuestra propuesta de actividades, hemos elegido las voces *arrestar*, *consentimiento*, *denunciar* y *penal* del corpus de muestra. Esta pequeña lista se da a los alumnos para que analicen su comportamiento contextual con ayuda de *Word Sketch* y posteriormente hagan sus propios mapas conceptuales. Serían parecidos al que se presenta en la Figura 5 para el verbo *denunciar* (las palabras conectadas por líneas azules).

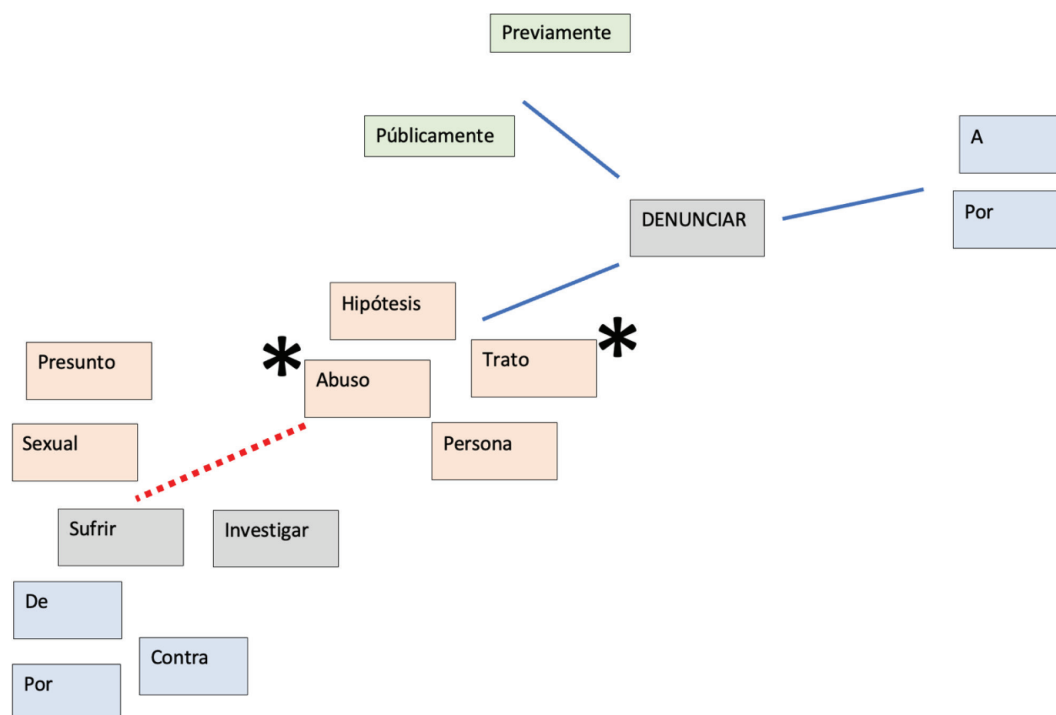


FIGURA 5: Mapa conceptual desarrollado basado en el comportamiento contextual del verbo denunciar

Siguiendo el formato de las visualizaciones gráficas de Word Sketch, que hemos mencionado en la sección 3.1., están marcadas con colores diferentes las palabras que cumplen diferentes funciones sintácticas con respecto a *denunciar*: los modificadores adverbiales (*previamente* y *públicamente*) aparecen en verde, los objetos directos nominales (*hipótesis*, *trato*, *abuso* y *persona*) en beige, y las preposiciones *a* y *por* en gris. Una vez diseñados los mapas conceptuales, se puede pedir a los alumnos que los comparen con las visualizaciones gráficas de Word Sketch y que marquen en sus mapas las combinaciones léxicas con niveles elevados de fijación (combinaciones no variables, a las que hemos aludido en la sección 3.1.). En el esquema ejemplar (figura 5), estas combinaciones fuertes se han marcado con el símbolo de asterisco: ambas combinaciones, *denunciar* + *abuso* y *denunciar* + *trato*, disponen de una calificación *LogDice* alta en el corpus de muestra (12,5 y 12,7, respectivamente).

En el siguiente paso de la actividad, se puede proponer a los alumnos expandir los mapas conceptuales añadiendo el contexto de algunas de las palabras incluidas en el mapa. En la Figura 5 ilustramos cómo se puede hacer con *abuso* (objeto directo de *denunciar* de la búsqueda original); la parte añadida se ha marcado con una línea discontinua roja. Además, si el corpus no es muy grande, el mapa se puede enriquecer con unidades léxicas de un corpus general. Así se fomentaría en los alumnos la capacidad de buscar ejemplos propios.

Por último, propondremos una actividad adicional que hace uso de un listado específico de *Word Sketch* (*usage patterns* ‘patrones de uso’) para los verbos y otras palabras relacionales, que resume su comportamiento gramatical (sus usos en indicativo o subjuntivo, su aparición en estructuras reflexivas, pasivas, etc.). Aunque nuestra propuesta trata de la enseñanza de léxico, este está, en el enfoque léxico, muy estrechamente vinculado con la gramática. En la actividad que proponemos, se trata precisamente de que los alumnos analicen los patrones de uso de la palabra buscada y los relacionen con las modulaciones semánticas que puede experimentar la palabra cuando se usa como parte de estos patrones. Se usarán dos herramientas de forma combinada: *Word Sketch* y las concordancias. En la fase preparatoria, el docente debe elegir entre tres y cinco palabras que presenten algún tipo de dificultad gramatical en su uso, como ser compatibles con el indicativo y el subjuntivo, o con los verbos *ser* y *estar*. Deben ser palabras relacionadas con el tema tratado y suficientemente frecuentes en el corpus. La tarea de los alumnos consiste en explicar las diferencias semánticas asociadas a estas alternancias sintácticas. Para ilustrarlo, tomaremos como ejemplo el verbo *detener* del corpus de muestra. Al consultar el listado *usage patterns*, los alumnos comprobarán que aparece en la construcción <ser + participio> (*ser detenido*) y <estar + participio> (*estar detenido*). Para explicar las diferencias semánticas entre ambas construcciones, tendrán que acceder a las concordancias (pueden hacerlo directamente desde *Word Sketch*) y, basándose en los ejemplos

de uso, llegar a la conclusión de que la pasiva *ser detenido* alude a un acto y la construcción atributiva *estar detenido* al estado que es el resultado de dicho acto.

3.3.3. *Word Sketch Diferencial*

La función de la herramienta *Word Sketch Diferencial* es contrastar el comportamiento gramatical de dos palabras. Como hemos visto en la sección 2.3., la precisión léxica es uno de los principales parámetros que definen el dominio de una LE. En niveles avanzados, se logra a través de una variedad de estrategias: actividades de sustitución de palabras comodín por terminología específica, ejercicios de diferenciación semántica de términos cognados y de cuasisinónimos, etc. La actividad que proponemos se centra precisamente en la comparación de sinónimos y cuasisinónimos en función de su comportamiento contextual.

Como con todas las herramientas de *Sketch Engine* mencionadas en esta sección, antes de la actividad el docente debe explicar para qué y cómo se usa *Word Sketch Diferencial*. Para detectar los (cuasi)sinónimos, el profesor puede hacer uso de la herramienta de frecuencia (*Wordlist*), que hemos introducido antes. Elegiría entre tres y cinco pares de sinónimos. En nuestro corpus de muestra, por ejemplo, aparecen las siguientes parejas: *crimen-delito*, *cuerpo-cadáver*, *hallar-encontrar* y *morir-fallecer*. Los alumnos tienen que introducir estas palabras en el buscador de *Word Sketch Diferencial* y analizar las diferencias que existen en cada par (preguntas guía en el ejercicio ayudarán a los alumnos enfocar su búsqueda, por ejemplo: *¿Cuál puede ser un crimen y un delito?*). En el análisis subsecuente se ve, por ejemplo, que los nombres *crimen* y *delito* se usan con modificadores diferentes: *machista* solo se usa con *crimen*, y *grave* preferentemente con *delito*. En el siguiente paso, se puede pedir a los alumnos que confirmen su análisis a través de un corpus de referencia de gran extensión (por ejemplo, *CORPES XXI* o *esTenTen18*).

Como ya hemos dicho, las actividades con *Word Sketch* y *Word Sketch Diferencial* son especialmente recomendables para el aprendizaje del lenguaje formulaico con un nivel alto de fijación: por un lado, contribuyen a ampliar el vocabulario de los aprendientes mediante la introducción de sinónimos y, por otro lado, les ayudan a detectar las diferencias combinatorias que existen entre los sinónimos. Desde la perspectiva de selección libre, no existe ninguna razón por la que no se pueda usar la combinación *delito machista*, por ejemplo: se trata de una combinación semánticamente válida y gramaticalmente correcta (hemos visto un caso parecido con *caballo ganador*). No obstante, no se da en el uso real, como confirman los datos.

3.3.4. *N-gramas*

La herramienta *N-gramas* detecta expresiones estables de palabras de diferente nivel de fijación y las ordena según su frecuencia de aparición. Puesto que *N-gramas* muestra la frecuencia absoluta de cada expresión detectada, el estudiante puede distinguir entre combinaciones de voces frecuentemente usadas por los hablantes nativos por un lado y los hápax o usos ocasionales (por ejemplo, invenciones de hablantes específicos) por otro lado. La búsqueda en *N-gramas* se puede acotar de varias maneras: por la secuencia de letras que debe contener la expresión, por el número de palabras gráficas, por frecuencia, etc. Si no se acota la búsqueda, se pueden encontrar expresiones previamente desconocidas. Obviamente, una búsqueda no restringida en un corpus general extenso (como *esTenTen18*) devolvería muchísimos resultados, únicamente clasificados por frecuencia. Por ello, su utilidad didáctica sería limitada. En tal caso, sería más eficiente utilizar las concordancias para buscar expresiones ya conocidas por los alumnos o previamente preparadas por el profesor. En cambio, con un corpus temáticamente acotado (como el nuestro) ambos tipos de búsqueda –acotada y sin acotar– serían viables y se podrían aplicar para extraer expresiones específicas de un tema

determinado dependiendo de si se quiere localizar expresiones ya conocidas o todas las expresiones, incluidas las desconocidas.

Como con la herramienta *Word Sketch*, en las dos actividades con *N-gramas* que proponemos aquí se hace uso de mapas conceptuales como método para representar y asimilar las relaciones léxicas, semánticas y asociativas, pero en este caso se pone énfasis en las expresiones pluriverbales (los resultados de búsquedas en *Word Sketch* y *Word Sketch Diferencial* siempre son unidades léxicas compuestas por dos palabras: la base y el colocativo; por lo contrario, *N-gramas* devuelve resultados multipalabra que en la búsqueda en *Word Sketch* no aparecerían). En la primera actividad se hace una búsqueda no restringida de expresiones con *N-gramas* en el corpus temáticamente acotado; los alumnos limitan la extensión de las expresiones a un mínimo de dos palabras y un máximo de cinco. En el caso de nuestro corpus de muestra, se extraerían expresiones como *violencia {de género / de pareja / contra las mujeres}*, *parte de lesiones*, *denuncia por violencia de género*, *presunto autor del crimen*, *igualdad de oportunidades*, etc. En el siguiente paso, el docente propondría a los alumnos diseñar un mapa conceptual (en este caso semántico) con estas expresiones: por ejemplo, el bloque representado por *violencia {de género / de pareja / contra las mujeres}* estaría relacionado por un vínculo causal con *parte de lesiones* y con un vínculo agentivo con *presunto autor del crimen*.

En la segunda actividad se haría uso de la búsqueda avanzada de *N-gramas*, para acotarla a expresiones que contengan una palabra concreta. Este paso permite extraer grupos acotados de expresiones, que podrían pasar desapercibidas en un listado general. En nuestro corpus, el sustantivo *violencia* aparece en las expresiones *violencia {de género / de pareja / contra las mujeres}*, *víctimas de violencia*, *caso de violencia*, *pacto contra la violencia de género*, etc. Los resultados de esta búsqueda se podrían combinar con el mapa conceptual más general, confeccionado en la primera actividad.

A diferencia de *Word Sketch*, *N-gramas* no clasifica los componentes de cada expresión por su función sintáctica. Es algo que se puede proponer como tarea a los alumnos, como un repaso de contenidos gramaticales previamente adquiridos.

4. CONCLUSIONES

Los corpus son una herramienta de potencial indudable para muchas de las ramas de la lingüística, entre ellas la enseñanza de lenguas extranjeras, pero por diversos motivos el auge de la lingüística computacional en los años noventa no propició su aplicación en la práctica docente. El objetivo de este trabajo ha sido precisamente ilustrar este potencial proponiendo una nueva forma de su uso en el aula de idiomas. Aquí nos hemos centrado en los beneficios que ofrece la consulta de corpus cuando se enmarca dentro del enfoque léxico de Lewis. En concreto, hemos argumentado que el trabajo con corpus puede beneficiar especialmente a los estudiantes avanzados, por las características de la adquisición léxica en este nivel de competencia.

Usando diferentes herramientas del sistema de consulta y gestión de corpus *Sketch Engine* (*Concordancia*, *Word Sketch*, *Word Sketch Diferencial* y *N-gramas*), hemos diseñado una muestra de actividades en las que se explota un corpus propio temáticamente acotado. Estas actividades se han centrado principalmente en las expresiones pluriverbales (su detección y sus propiedades sintácticas, entre ellas su grado de fijación), y los sinónimos y cuasisinónimos (su identificación en los textos y sus diferencias de uso dentro de combinaciones libres de palabras y dentro de expresiones pluriverbales). Hemos mostrado cómo se puede explotar la información estadística que proviene de los corpus (la frecuencia absoluta y la métrica de la fuerza de conexión contextual *LogDice*) para introducir en el aula estos y otros fenómenos léxicos, que se trabajan y se asimilan sobre todo en los niveles avanzados de competencia en segundas lenguas.

La eficacia de la técnica propuesta y de las actividades diseñadas habrá de ser confirmada en un estudio empírico amplio. El mismo estudio contribuirá a verificar la adecuación de los parámetros más técnicos, que aquí hemos fijado de forma tentativa, entre ellos el tamaño del corpus autocompilado y la forma de recolección de los datos. En este estudio, las expresiones pluriverbales se abordarán como una de las manifestaciones particulares de la combinatoria léxica, cuyo tratamiento en el aula de segundas lenguas no se suele abordar de forma sistemática a pesar de su importancia para poder alcanzar niveles altos de competencia léxica, sintáctica y discursiva.

FINANCIACIÓN

La publicación es parte del proyecto “Corpus de composicionalidad e informatividad léxica: anotación, análisis y aplicaciones (INFOLEXIS)” (IP Olga Batiukova, referencia PID2022-138135NB-I00), financiada por MCIN/AEI/10.13039/501100011033 y por el FSE+. Su realización ha sido financiada por un contrato predoctoral asociado a este proyecto.

La publicación es parte de la ayuda PREP2022-000882, financiada por MCIN/AEI/10.13039/501100011033 y por el FSE+.

REFERENCIAS BIBLIOGRÁFICAS

- Álvarez Cavanillas, J. (2017). Un enfoque léxico en los manuales ELE. En F. Herrera (ed.), *Enseñar léxico en el aula de español: el poder de las palabras* (pp. 57-69). Difusión.
- Batten, L., Cornu, A. & Engels, L. V. (1989). The use of concordances in vocabulary acquisition. En C. Laurent & M. Nordman (eds.), *Special language: from humans thinking to thinking machines* (pp. 452-467). Multilingual Matters.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition*, 23(3), 321-343. <https://doi.org/10.1017/S0272263101003011>
- Boulton, A. (2010). Data-driven learning: taking the computer out of the equation. *Language Learning*, 60(3), 534-572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Boulton, A. (2012). Hands-on / hands-off: alternative approaches to data-driven learning. En J. Thomas & A. Boulton (eds.), *Input, process and product: developments in teaching and language corpora* (pp. 153-169). Masaryk University Press.

- Buyse, K. (2017). Los corpus como herramientas de aprendizaje de léxico. En F. Herrera (ed.), *Enseñar léxico en el aula de español: el poder de las palabras* (pp. 121-140). Difusión.
- Capel, A. (2012). Completing the English vocabulary profile: C1 and C2 vocabulary. *English Profile Journal*, 3(1), 1-14. <https://doi.org/10.1017/S2041536212000013>
- Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, 35, 30-40. <https://doi.org/10.1016/j.esp.2013.11.004>
- Consejo de Europa (2001). *Marco Común Europeo de Referencia para las lenguas*. División de Lenguas Modernas.
- Cruz Piñol, M. (2012). *Lingüística de corpus y enseñanza del español como 2-L*. Editorial Arco Libros.
- Edinumen. (s.f.). *Frecuencias*. Edinumen. <https://edinumen.es/materiales/frecuencias/>
- Elvira-García, W. (ed.). (2021). *El uso de corpus en clase de ELE: la lengua real como modelo*. Difusión.
- Gablasova, D., Brezina, V. & McEnery, T. (2017). Collocations in corpus-based language learning research: identifying, comparing, and interpreting the evidence. *Language Learning*, 67(1), 155-179. <https://doi.org/10.1111/lang.12225>
- Gaskell, D. & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301-319. <https://doi.org/10.1016/j.system.2004.04.001>
- Higueras, M. (2009). Aprender y enseñar léxico. *MarcoELE*, 9, 111-26.
- Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes*. Instituto Cervantes. <https://n9.cl/ysfzp>
- Instituto Cervantes (2014). Corpus de aprendices del español: CAES (Versión 2.1.) [corpus]. Instituto Cervantes. <https://galvan.usc.es/caes/>
- Kettemann, B. (1995). On the use of concordancing in ELT. *AAA: Arbeiten aus Anglistik und Amerikanistik*, 20(1), 29-41.
- Kilgarriff, A., Rychly, P., Smrz, P. & Tugwell, D. (2004). The Sketch Engine. En G. Williams & S. Vessier (eds.) *Proceedings of the Eleventh EURALEX International Congress* (pp. 105-15). Université de Bretagne-Sud, Faculté des lettres et des sciences humaines. <https://n9.cl/gg423>
- Lee, D. & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: moving from available specialized corpora to self-compiled corpora. *ScienceDirect*, 25, 56-75. <https://doi.org/10.1016/j.esp.2005.02.010>
- Lewis, M. (2008). *Implementing the Lexical Approach: putting theory into practice*. Heinle.
- Lexical Computing Ltd. (2021). esTenTen (estenten18_fl5_1) [corpus]. Lexical Computing Ltd. <https://www.sketchengine.eu>

- Maia, B. (2005). Terminology and translation—bringing research and professional training together through technology. *Meta: Journal des traducteurs*, 50(4), s. p. <https://doi.org/10.7202/019921ar>
- Pérez-Carrasco, M. (2023). The nuts and bolts of corpora: secuencia didáctica para el empleo de corpus en la enseñanza de la traducción científico-técnica. *Hikma: Estudios de Traducción*, 22(1), 307-337. <https://doi.org/10.21071/hikma.v22i1.15374>
- Pérez Serrano, M. (2017). *La enseñanza-aprendizaje del vocabulario en ELE desde los enfoques léxicos*. Arco Libros-La Muralla.
- Real Academia Española. (s.f.). *Corpus del español del siglo XXI: CORPES XXI* (Versión 1.1.) [corpus]. Real Academia Española. <http://www.rae.es>
- Real Academia Española. (s.f.). Corpus de referencia del español actual: CREA (Versión 1.0.) [corpus]. Real Academia Española. <http://www.rae.es>
- Rufat, A. & Jiménez Calderón, F. (2017). Aplicaciones de enfoques léxicos a la enseñanza comunicativa. En F. Herrera (ed.), *Enseñar léxico en el aula de español: el poder de las palabras* (pp. 47-55). Difusión.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Smith, S. (2011). Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning*, 24(4), 291-316. <https://doi.org/10.1080/09588221.2011.557024>
- Smith, S. (2020). DIY corpora for Accounting & Finance vocabulary learning. *English for Specific Purposes*, 57, 1-12. <https://doi.org/10.1016/j.esp.2019.08.002>
- Stevens, V. (1991). Concordance-based vocabulary exercises: a viable alternative to gap-fillers. *Classroom Concordancing: English Language Research Journal*, 4, 47-63.
- Tribble, C. & Wingate, U. (2013). From text to corpus: a genre-based approach to academic literacy instruction. *ScienceDirect*, 41, 307-21. <https://doi.org/10.1016/j.system.2013.03.001>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519772>
- Zanettin, F. (2002). DIY corpora: the WWW and the translator. En B. Maia, J. Haller & M. Urlych (eds.), *Training the language services provider for the new millennium* (pp. 239-248). Universidad de Porto.
- Zhao, Y. & Shi, J. (2015). Self-compiled on-line parallel corpus in translation teaching. En W. Liu (ed.), *Conference on Education and Teaching in Colleges and Universities* (pp. 68-71). Atlantis Press. <https://doi.org/10.2991/cetcu-15.2016.20>



EL LÉXICO ELE EN LOS MODELOS DE LENGUAJE

ELE LEXICON IN LANGUAGE MODELS

María Victoria Cantero Romero
Universidad de Jaén
vc Cantero@ujaen.es

RESUMEN

Este trabajo se encuentra enmarcado en el vocabulario generado por los grandes modelos de lenguaje como *ChatGPT*, y en concreto, en el especializado en español como lengua extranjera. Para ello, se le ha solicitado a dos grandes modelos de lenguaje, *ChatGPT* y *LlaMa* que genere textos de comprensión escrita para estudiantes de nivel de español desde A1 hasta B2. Para ello, es necesario el uso de una entrada específica que genere la respuesta más cercana a nuestra finalidad. Tras la petición, se han analizado las muestras siguiendo las referencias del apartado de nociones específicas del *Plan Curricular del Instituto Cervantes* (2006). Asimismo, se ha realizado una comparación entre ambos modelos, esperando encontrar diferencias entre las respuestas proporcionadas. Con este análisis, hemos querido comprobar si las muestras textuales generadas siguen los estándares del *Plan Curricular del Instituto Cervantes* (2006) con respecto al vocabulario. En conclusión, se ha comprobado que los modelos de lenguaje pueden ser de utilidad para la elaboración de materiales didácticos, no obstante, requieren de una revisión manual.

Palabras clave: Español Lengua Extranjera (ELE), vocabulario, comprensión lectora, modelos de lenguaje.

ABSTRACT

This work is framed in the vocabulary generated by large language models such as *ChatGPT*, and specifically, in the one specialized in Spanish as a foreign language. For this purpose, two large language models, *ChatGPT* and *LlaMa*, have been asked to generate written comprehension texts for students of Spanish level from A1 to B2. For this purpose, it is necessary to use a specific input that generates the closest response to our purpose. After the request, the samples have been analyzed following the references of the specific notions section of the *Plan Curricular del Instituto Cervantes* (2006). Likewise, a comparison has been made between both models, hoping to find differences between the answers provided. With this analysis, we wanted to check whether the textual samples generated follow the standards of the *Plan Curricular del Instituto Cervantes* (2006) with respect to vocabulary. In conclusion, it has been proven that language models can be useful for the development of teaching materials, however, they require manual revision.

Keywords: Spanish Foreign Language (SFL), vocabulary, reading comprehension, language models.

Recibido: 07-10-2024
Aceptado: 22-11-2024



1. INTRODUCCIÓN

El estudio del vocabulario en la enseñanza de español como lengua extranjera (ELE) ha sido un área de interés clave para los docentes, quienes buscan optimizar el proceso de adquisición léxica en los estudiantes. En este sentido, la Inteligencia Artificial (IA), y más específicamente los Grandes Modelos de Lenguaje (LLMs), se presentan como una herramienta valiosa para el diseño de actividades y la mejora de las competencias comunicativas en el aula de ELE.

La irrupción de *ChatGPT* en diciembre de 2022 supuso un hito significativo en el ámbito del procesamiento del lenguaje natural y su impacto se ha hecho sentir también en el campo de la enseñanza de idiomas (Jiménez Linares, López-Gómez, Martín-Baos, Romero & Serrano-Guerrero, 2023, p. 113). Este modelo ha comenzado a ser utilizado por docentes como recurso para la creación de actividades (Fredholm, 2024, p. 8), aunque su uso entre los estudiantes ha planteado desafíos, al ser empleado en algunos casos como un sustituto del esfuerzo personal en la producción escrita.

Este artículo parte de una experiencia en la que un grupo de estudiantes de ELE utilizó *ChatGPT* para redactar sus tareas sin participar activamente en el proceso de escritura. Tras observar que el nivel de vocabulario empleado superaba el esperado para su nivel de competencia lingüística, se identificó la necesidad de investigar el papel de la IA en la enseñanza del español, particularmente en el ámbito del vocabulario y la adecuación a los niveles establecidos por el *Plan Curricular del Instituto Cervantes* (PCIC).

En particular, este estudio se enfocará en la evaluación del vocabulario generado por los LLM en textos diseñados para la comprensión lectora en niveles A1/A2 y B1/B2. El estudio del léxico en estos niveles es esencial, ya que el dominio de un vocabulario adecuado es un componente central en el desarrollo de la competencia lingüística de los estudiantes de ELE (Moreno Villanueva, 2021, p. 14). La adecuación de los textos a los niveles de los estudiantes no solo garantiza una progresión efectiva en el aprendizaje,

sino que también permite a los estudiantes enfrentarse a textos que reflejan de manera realista el uso del español en distintos contextos comunicativos (Bartol Hernández, 2010, p. 85). Asimismo, es importante la realización de un estudio de los textos de comprensión lectora, ya que se trata de una competencia básica (Núñez Delgado & Rodríguez Guerrero, 2017, p. 34). Como nos indican Muñoz-Basols y Fuertes Gutiérrez (2024, p. 351) las herramientas de IA nos pueden facilitar la evaluación del progreso en la adquisición del vocabulario.

Sin embargo, es necesario que los textos generados sean revisados cuidadosamente para asegurar que respetan los principios lingüísticos y pedagógicos adecuados para la enseñanza del español (Ribes-Lafoz & Navarro Colorado, 2023, pp. 1264-1265). En este sentido, el análisis del vocabulario en los textos generados por IA permitirá determinar si estos modelos pueden ser una herramienta útil en la creación de actividades didácticas que refuercen las competencias lingüísticas de los estudiantes.

Por tanto, el presente estudio no solo busca explorar la viabilidad del uso de LLM en la enseñanza de ELE, sino también aportar nuevas perspectivas sobre cómo la IA puede contribuir al estudio del vocabulario en contextos de aprendizaje de lenguas. La adecuada integración de estas tecnologías en el aula no solo puede optimizar el trabajo docente, sino también enriquecer la experiencia de los estudiantes, motivándolos a interactuar de manera más efectiva con el español como lengua extranjera.

2. PREGUNTAS DE INVESTIGACIÓN

La primera cuestión se enfoca en el vocabulario generado por *ChatGPT* y *LlaMa* en función de los niveles de competencia de los estudiantes. ¿Cómo se relaciona la calidad y complejidad del vocabulario proporcionado por *ChatGPT* y *LlaMa* con los niveles de competencia lingüística de los estudiantes de ELE? Esta pregunta implica examinar la calidad y complejidad del vocabulario proporcionado por la IA en relación con los niveles de aprendizaje

de español como lengua extranjera. Se evaluará si las palabras y expresiones seleccionadas son adecuadas para facilitar la comprensión y adquisición de léxico, fomentando un aprendizaje progresivo y ajustado a las necesidades de cada nivel.

A continuación, nos cuestionamos si el vocabulario y las estructuras generadas por los modelos de lenguaje están alineados con los descriptores y directrices del *Plan Curricular del Instituto Cervantes* para los niveles A1-B2. Se analizará la correspondencia entre el léxico generado, los descriptores oficiales y los conocimientos de nuestro alumnado, garantizando que los términos sean adecuados para los niveles de competencia A1-B2 y respeten los criterios pedagógicos establecidos para la enseñanza del español como lengua extranjera.

Como última pregunta de investigación, nos planteamos ¿Cuál es el impacto de los modelos de lenguaje en el aprendizaje del español como lengua extranjera en términos de generación de materiales didácticos, adaptación a las necesidades del aula y motivación de los estudiantes? Para dar respuesta a esta cuestión se explorará la efectividad de los modelos para generar materiales didácticos útiles, adaptados y motivadores que faciliten el aprendizaje y la participación activa de los estudiantes, y se medirá su impacto en el aula de ELE.

3. METODOLOGÍA

En esta sección abordaremos varios aspectos clave. En primer lugar, se describirán los modelos de lenguaje utilizados en el estudio. A continuación, se ofrecerá una breve explicación de las nociones específicas del *Plan Curricular del Instituto Cervantes*.

3.1. MODELOS DE LENGUAJE UTILIZADOS

Para este estudio, realizamos una comparación entre dos modelos de generación de texto: ChatGPT (versión GPT-3.5) desarrollado por OpenAI, y LLaMA (versión 3.1.8b) desarrollado por Meta AI. Ambos modelos,

basados en la arquitectura *Transformer*, utilizan mecanismos de autoatención para procesar y generar lenguaje natural, simulando la producción textual humana.

La elección de *ChatGPT* se justificó no solo por su facilidad de uso, sino también por su amplio alcance: en abril de 2024, *ChatGPT* contaba con 180.5 millones de usuarios activos (ChatGPT: Usuarios y Estadísticas, 2024), lo que evidencia su relevancia y adopción global. Elegido por su accesibilidad y popularidad entre usuarios de diferentes ámbitos, incluidos docentes y estudiantes. Este modelo ha ganado reconocimiento por su capacidad para generar textos coherentes y realistas, sin necesidad de instalación o descarga, lo que facilita su implementación en contextos educativos (Ribes-Lafoz & Navarro Colorado, 2023, pp. 1264-1265).

Por otro lado, *LLaMA* es un modelo diseñado como una herramienta de investigación para optimizar el rendimiento de los grandes modelos de lenguaje. *LLaMA* se seleccionó por su eficiencia en la generación de texto con menos parámetros y su capacidad para analizar el lenguaje de manera precisa (Fernández, 2023).

3.2. EL LÉXICO EN EL PLAN CURRICULAR DEL INSTITUTO CERVANTES

A continuación, vamos a exponer un resumen de las nociones específicas de los niveles de A1 hasta B2 que encontramos en el *Plan Curricular del Instituto Cervantes* (2006). Estas nociones específicas son los componentes relacionados con el significado.

En el nivel A1, encontramos los siguientes apartados con ejemplos:

1. Individuo: dimensión física, esta dimensión se refiere a las descripciones físicas de las personas, su cuerpo y su estado de salud. En el nivel A1, los estudiantes aprenden vocabulario relacionado con las partes del cuerpo, como *cabeza*, *pierna*, *brazo*, así como expresiones simples para hablar del estado físico de las personas, como *estar cansado* o *tener hambre*. Esto permite a los estudiantes describir a otros o a sí mismos en términos básicos.

2. Individuo: dimensión perceptiva y anímica, en este ámbito, los estudiantes aprenden a hablar sobre sus emociones y percepciones. En el nivel A1, se introducen palabras básicas para describir estados de ánimo, como *feliz, triste, enfadado* o *nervioso*. También se cubren expresiones relacionadas con los cinco sentidos, como *ver, oír, sentir*, lo que permite a los estudiantes expresar sus experiencias sensoriales de manera sencilla.
3. Identidad personal: este apartado se centra en la capacidad de los estudiantes para hablar sobre sí mismos de manera básica. En el nivel A1, los estudiantes aprenden a dar información personal como su nombre, edad, nacionalidad, y ocupación. También se introducen los términos necesarios para hablar sobre aspectos de la vida personal como *el estado civil (soltero, casado)* y la dirección de residencia. El objetivo es que puedan presentarse y proporcionar detalles esenciales en interacciones cotidianas.
4. En cuanto a las relaciones personales, el vocabulario cubre los términos más comunes para referirse a miembros de la familia (*padre, madre, hermano*), amigos y conocidos. Los estudiantes aprenden a describir relaciones básicas y a hablar de su círculo social utilizando frases sencillas. Además, pueden describir la relación que tienen con personas en contextos formales e informales.
5. Alimentación: este apartado aborda todo lo relacionado con la comida y las bebidas. En el nivel A1, los estudiantes aprenden los nombres de alimentos básicos, como *pan, agua, carne, fruta*, así como verbos esenciales como *comer* y *beber*. También se introducen estructuras para pedir comida en un restaurante o describir una comida diaria. Esta área es fundamental para desenvolverse en contextos como ir a un restaurante o hacer la compra.
6. El ámbito educativo se enfoca en proporcionar el vocabulario básico para hablar sobre las instituciones académicas, las asignaturas y las actividades en el aula. En el nivel A1, los estudiantes aprenden palabras

como *profesor, clase, examen y deberes*. Esto les permite hablar sobre su experiencia en el aula y su vida académica de manera sencilla, facilitando la descripción de actividades como *estudiar, aprender y hacer un examen*.

7. Trabajo: en este apartado, se introducen nociones relacionadas con las profesiones más comunes y las acciones vinculadas al trabajo. Los estudiantes aprenden a describir su ocupación con palabras como *médico, profesor, taxista* y estructuras básicas como *trabajar en* o *ser*. Además, pueden hablar sobre situaciones laborales, como *tener trabajo* o *estar en paro* proporcionando el vocabulario esencial para describir su situación laboral.
8. El vocabulario relacionado con el ocio cubre una amplia gama de actividades recreativas, desde deportes hasta entretenimiento cultural. En el nivel A1, los estudiantes aprenden a hablar de actividades como *ver una película, leer un libro, escuchar música y viajar*. Esto les permite describir cómo pasan su tiempo libre y hablar sobre planes y preferencias recreativas de manera básica.
9. Información y medios de comunicación: este apartado, se introduce el vocabulario necesario para hablar de los medios de comunicación, como *televisión, radio, teléfono y correo electrónico*. Los estudiantes también aprenden a describir acciones como *ver la televisión, leer el periódico* o *escribir un mensaje electrónico*. Esto les proporciona las herramientas lingüísticas para interactuar con medios de comunicación en contextos cotidianos.
10. El vocabulario relacionado con la vivienda incluye los tipos de residencias y las partes de una casa. En el nivel A1, los estudiantes aprenden a describir su hogar con palabras como *casa, piso, habitación, cocina y baño*. También se introduce vocabulario relacionado con la compra y el alquiler de viviendas, lo que les permite hablar sobre el tipo de lugar donde viven y las características de este.

11. En el ámbito de los servicios, los estudiantes aprenden términos relacionados con el correo, los bancos, y los servicios médicos. Palabras como *carta, sello, banco, dinero, hospital* y *policía* son esenciales en este nivel para que los estudiantes puedan desenvolverse en situaciones cotidianas como enviar una carta o visitar al médico.
12. Compras, tiendas y establecimientos. Este apartado se centra en las interacciones comerciales más básicas, como comprar en una tienda. Los estudiantes aprenden vocabulario relacionado con tiendas (*mercado, supermercado, tienda*) y productos comunes, como *pantalones, camisa, precio, dinero*. Esto les permite realizar compras sencillas y describir lo que necesitan o desean adquirir.
13. En el área de la salud, los estudiantes aprenden a describir su estado físico, como *estar enfermo* o *tener fiebre*. También adquieren vocabulario relacionado con el cuidado personal y la higiene, como *lavarse las manos, bañarse y ducharse*. Este vocabulario es crucial para describir situaciones relacionadas con la salud y el bienestar.
14. Viajes, alojamiento y transporte: en este ámbito, se introduce el vocabulario relacionado con los viajes, el transporte y el alojamiento. Palabras como *aeropuerto, tren, autobús, hotel* y *billete* son esenciales para que los estudiantes puedan describir y entender situaciones de viaje. También se cubren actividades comunes como *comprar un billete* o *viajar en avión*.
15. El vocabulario económico básico incluye términos como *dinero, precio, tarjeta, cambiar dinero*. En este nivel, los estudiantes aprenden a describir transacciones comerciales sencillas, como pagar en una tienda o cambiar dinero en un banco.
16. Ciencia y tecnología. Este apartado introduce términos básicos relacionados con la tecnología y la informática, como *ordenador, Internet, programa* y *navegar por Internet*. También se cubre el vocabulario

necesario para describir acciones sencillas en el ámbito de la tecnología, como *enviar un correo electrónico* o *usar un portátil*.

17. Gobierno, política y sociedad. En este ámbito, se cubren nociones relacionadas con las instituciones políticas y el gobierno, como *país*, *capital*, *presidente*, *rey*, *reina*. Los estudiantes aprenden a describir su lugar de origen y a reconocer las estructuras políticas más básicas.
18. Actividades artísticas. Este apartado se centra en las actividades culturales, como el cine, la música y las exposiciones de arte. Los estudiantes aprenden palabras como *película*, *concierto*, *museo*, *teatro* y pueden describir sus preferencias artísticas o hablar de una actividad cultural que les interese.
19. Religión y filosofía. Aquí, el vocabulario se relaciona con las principales religiones, como *cristiano*, *judío*, *musulmán*, y los lugares de culto, como *iglesia*, *mezquita* y *sinagoga*. Esto permite a los estudiantes describir su religión o referirse a aspectos culturales religiosos básicos.
20. Geografía y naturaleza. El vocabulario geográfico cubre palabras como *país*, *ciudad*, *montaña*, *río*, *playa*, *mapa*. Los estudiantes aprenden a describir ubicaciones geográficas y características naturales básicas, lo que les permite hablar de su entorno y su contexto geográfico de manera sencilla.

En el nivel A2, en el apartado de nociones específicas encontramos:

1. Individuo: Dimensión física, se incluyen partes del cuerpo más detalladas como *espalda*, *estómago*, *oído*. También se describen características físicas sencillas como *ser calvo* y *llevar gafas*.
2. Individuo: Dimensión perceptiva y anímica, se utilizan adjetivos para describir la personalidad, tales como *optimista*, *reservado*, *nervioso*. En cuanto a los estados anímicos, se incluyen expresiones como *estar preocupado* y *odiar algo*.

3. Identidad personal: se abordan aspectos de la identidad personal más completos, como *apellidarse y tener la nacionalidad*. Además, se mencionan objetos personales como *agenda, maletín*.
4. Relaciones personales: se amplía el vocabulario relacionado con la familia, incluyendo términos como *hijo único, exmarido*. También se abordan las relaciones laborales con términos como *socio, compañero de trabajo*.
5. Alimentación: se presentan comidas más complejas, como *merluza, gambas, pasta*. También se incluyen bebidas comunes como *zumos de naranja, lata o relacionadas como tener sed*.
6. Educación: se mencionan asignaturas escolares como *Matemáticas, Historia, Medicina*. Además, se incluyen actividades académicas como *repasar un ejercicio, memorizar, aprobar*.
7. Trabajo: se amplía el vocabulario de profesiones con términos como *dependiente, funcionario*. También se incluyen actividades laborales comunes, como *contestar el teléfono, preparar una factura*.
8. Ocio: se mencionan actividades recreativas como *ir de paseo, ir de compras, tener una cita*. Además, se incluyen deportes como *montar en bicicleta o jugar a las cartas*.
9. Información y medios de comunicación: se incluyen medios de comunicación más amplios como *navegar por Internet, hablar/llamar por teléfono*. También se abordan formas de comunicación escrita, como *escribir una carta, enviar un mensaje electrónico*.
10. Vivienda: se mencionan tipos de viviendas más específicos como *chale, piso de estudiantes, propietario*. También se describen condiciones de vivienda como *piso amueblado, en buen estado*.
11. Servicios: se incluyen servicios adicionales como *Cruz Roja, Seguridad Social*. Además, se mencionan servicios financieros, como *cajero automático*.

12. Compras, tiendas y establecimientos: se abordan tiendas especializadas como *frutería, carnicería, panadería*. También se mencionan productos comunes como *ropa de hombre, ropa de mujer, bolsos*. Y pagos como *dejar propina, con IVA*.
13. Salud e higiene: se incluyen enfermedades y dolencias como *tener alergia, doler la espalda*. También se abordan aspectos de la higiene personal, como *cepillarse los dientes, peinarse*.
14. Viajes, alojamiento y transporte: se presentan tipos de alojamiento como *media pensión, pensión completa*. Además, se mencionan tipos de transporte como *ciclista, coger un taxi, coger el avión*.
15. Economía e industria: se mencionan transacciones cotidianas como *pagar en efectivo, publicidad*. También se incluyen tipos de empresa, como *fábrica, negocio, despacho*.
16. Ciencia y tecnología: se incluyen tecnologías más avanzadas como *informática, navegar por internet, virus*. También se mencionan acciones relacionadas con *matemáticas*.
17. Gobierno, política y sociedad: se incluyen instituciones políticas como *gobierno, ministro, ayuntamiento*. Además, se mencionan términos relacionados con la comunidad, como *vecino, ciudadano*.
18. Actividades artísticas: se abordan aspectos del arte, incluyendo *museo, exposición, monumento*. También se incluyen géneros musicales como *flamenco, rock, pop*.
19. Religión y filosofía: se incluyen prácticas religiosas como ir a la iglesia, celebrar Navidad.
20. Geografía y naturaleza: se mencionan elementos naturales como *árbol, flor, pino*. También se incluyen términos relacionados con el clima, como *lluvia, nieve, sol*.

Para el nivel B1, el *Plan Curricular del Instituto Cervantes* nos indica lo siguiente:

1. Individuo: Dimensión física en el nivel B1, los estudiantes aprenden términos relacionados con partes del cuerpo como *músculo, hueso, piel, corazón, pulmón, cuello, hombros, pecho, cintura, barriga, rodilla, tobillo, codo, muñeca, frente, mejilla, barbilla, ceja y pestaña*.
2. Individuo: Dimensión perceptiva y anímica, se trabajan conceptos emocionales y actitudinales como *el optimismo, pesimismo, tranquilidad, paciencia, sinceridad, sentido del humor, alegría, admiración, alivio, emoción, felicidad, orgullo, angustia, enfado, temor, tristeza. Estas emociones y actitudes permiten a los estudiantes expresarse con mayor profundidad en contextos cotidianos y personales*.
3. Identidad personal: se introducen nociones relacionadas con la identidad y el estado civil, como *cambiar de nombre, llamar por el nombre o apellido, tratar de tú o usted, vivir en las afueras o en el centro, recién nacido, anciano, varón, hembra, boda civil o religiosa, divorcio y separación*.
4. Relaciones personales: los términos relacionados con las relaciones familiares y personales incluyen *madre soltera, niño adoptado, convivir con la pareja, discutir con los padres, cónyuge, cuñado, suegro, yerno, nuera, familiar cercano o lejano, familia numerosa, educar a un hijo, tener novio o pareja, salir con alguien, ligar*.
5. Alimentación: se abordan conceptos como *alimentación sana y equilibrada, calorías, proteínas, alimentos sin conservantes, bajo en calorías, productos frescos o congelados, marisco, lentejas, guisantes, leche entera o descremada, chuleta de cerdo, aceite de oliva, freír, hervir y cocer*.
6. En el ámbito educativo, los estudiantes aprenden vocabulario como *colegio público o privado, guardería, curso elemental, educación primaria y secundaria, asignatura obligatoria, matrícula, beca, asignatura trimestral, curso de formación continua, hacer un intercambio, examen, hacer un control y corregir un examen*.

7. Los términos relacionados con el mundo laboral incluyen *director de RRHH, empleado de banca, sustituto, trabajar para una empresa, tener un empleo en la administración, bufete, taller, casco, botas, hacer un pedido, trabajar en equipo, firmar un contrato, negociar un contrato y despedir*.
8. Se introducen actividades de ocio como *ir de camping, hacer turismo, inscribirse en un club, tener tiempo libre, salir de copas, ser aficionado a, coleccionar sellos, fiesta de disfraces, parque temático, espectador, galería de arte, poner una película y hacer cola*.
9. Información y medios de comunicación: este campo abarca términos como *medios de comunicación, dar información, estar informado, buzón, destinatario, carta formal o personal, contestador automático, llamada, línea telefónica, hacer una llamada, periodismo, titular y suscribirse a un periódico*.
10. Los términos relacionados con la vivienda incluyen *casero, propietario, visitar un piso, pagar el alquiler, pagar el recibo de la luz, empresa de mudanzas, hacer una mudanza, bloque de pisos, vivienda de alquiler, portero, administrador y buena o mala distribución*.
11. En el ámbito de los servicios, los estudiantes aprenden vocabulario como *cartero, buzón de correos, giro postal, cuenta corriente, cheques de viaje, ingresar dinero, pedir un préstamo, clínica, urgencias, vigilante, policía municipal y seguridad ciudadana*.
12. El vocabulario de compras incluye términos como *salida de emergencia, descuento, garantía, rebajas, establecimiento, cliente, consumidor, ropa de verano, zapatos planos, coser, que la ropa quede bien o mal, carro de la compra y producto natural*.
13. Salud e higiene: se introducen conceptos relacionados con la salud, como *seguro médico, tener fiebre, estar sano, diagnóstico, ataque al corazón, enfermedad leve o grave, contagiarse, perder el apetito, quemadura, corte, tener cicatriz, adelgazar, estar mareado, tener náuseas y hacerse una revisión*.

14. El vocabulario de viajes incluye términos como *equipaje, cambiar un billete, hacer turismo, viaje de trabajo, registrar las maletas, hostel, tienda de campaña, albergue, habitación completa, consigna, autovía y autopista de peaje*.
15. Economía e industria: en este campo, los términos incluyen *cuenta bancaria, pedir un crédito, ingresar dinero, talonario de cheques, pagar a plazos, presupuesto, hacer una transferencia, cobrar un cheque, hipoteca y caja fuerte*.
16. El vocabulario tecnológico abarca términos como *ratón, programa, antivirus, consultar una página web, descargar un archivo, abrir una cuenta de correo, adjuntar un documento y virus informático*.
17. Gobierno, política y sociedad, los estudiantes aprenden nociones relacionadas con la burocracia y el gobierno, como *DNI, pasaporte, documento caducado, funcionario, burocracia y tener validez*.
18. Se trabajan conceptos artísticos como *galería de arte, función de teatro, representación teatral, espectáculo deportivo, estrenar y cancelar un espectáculo*.
19. Los términos religiosos incluyen boda religiosa, misa y ceremonia.
20. En el ámbito de la geografía, se introducen términos como *mapa, plano turístico, travesía y ruta*.

Por otro lado, con respecto al nivel B2 encontramos lo siguiente:

1. Individuo: dimensión física, encontramos términos como *cerebro, hígado, riñón, esqueleto, columna, costilla, nervio, articulación, arteria, tendón*. También se incluyen términos sobre las funciones corporales más avanzadas, como *respirar, llorar, tocar, besar, abrazar*. En cuanto a la postura, se utilizan términos como *incorporarse, alzarse, agacharse, cruzar los brazos/las piernas*.
2. Individuo: dimensión perceptiva y anímica, se incorporan adjetivos más complejos como *arrogante, ambicioso, introvertido, vago*, y se exploran emociones más profundas como *admiración, alivio*,

angustia, fascinación, resignación. También se amplía el vocabulario para describir sentimientos con matices, como *dar asco/pánico/terror, sentirse animado/agobiado, estar furioso/histérico*.

3. Identidad personal: se incluyen términos más detallados como ser virgen, hacer el amor, tener relaciones sexuales, así como aspectos formales de la identidad como Documento Nacional de Identidad (DNI), *validez, fecha de expiración*. También se consideran aspectos burocráticos como *renovar o solicitar documentos*.
4. Relaciones personales: se amplía el vocabulario relacionado con la familia y la sociedad, incluyendo términos como *familiar cercano/lejano, madre de alquiler, niño probeta, familia adoptiva*. En cuanto a las relaciones interpersonales, se incluyen conceptos como *tener una relación de amor/trabajo, salir/romper con alguien, tener un romance o una aventura*.
5. Alimentación: se introduce vocabulario relacionado con dietas y productos específicos, como *producto orgánico, alimento bajo en calorías, rico en proteínas, legumbres, frutos secos*. En la preparación de alimentos, se incluyen acciones más complejas como *rebozar, rallar, exprimir, dorar la cebolla*.
6. Educación: en este nivel se detallan más términos académicos como *congreso, seminario, créditos, carrera, máster, doctorado*. Se incluye vocabulario relacionado con los estudios universitarios y profesionales, como *solicitar una beca, obtener créditos, licenciarse, doctorarse*.
7. Trabajo: se amplía el vocabulario laboral con términos como *contrato temporal, empleo cualificado, trabajo por cuenta propia/ajena, subir el sueldo, firmar un presupuesto*. También se incluyen aspectos más avanzados de la carrera profesional como *negociar un contrato, trabajar en equipo, coordinar un departamento*.
8. Ocio: en este nivel se incorporan términos más detallados como *resolver pasatiempos, organizar una exposición, asistir a una representación*.

- teatral, participar en un campeonato.* Además, se añaden actividades de ocio más complejas como *hacer una marcha, hacer alpinismo.*
9. Información y medios de comunicación: vocabulario relacionado con los medios de comunicación más avanzados como *editorial, redactor jefe, corresponsal, reportaje, suplemento cultural.* También se incluyen términos sobre tecnología de la información como *descargarse archivos, participar en una videoconferencia.*
10. Vocabulario más específico sobre tipos de viviendas y sus características como *vivienda de segunda mano, con buena/mala distribución, bien/mal orientada.* También se incluyen términos sobre construcción y reforma, como *cemento, ladrillo, reformar una vivienda.*
11. Servicios: se incluyen términos más avanzados sobre servicios bancarios y financieros como *hipoteca, crédito, talonario de cheques, pedir un préstamo.* También se amplía el vocabulario relacionado con servicios de seguridad, como *policía municipal, inspector de policía, servicio de vigilancia.*
12. Compras, tiendas y establecimientos: se introducen conceptos más avanzados como *comprar por catálogo o por Internet, condiciones de financiación, pagar a plazos.* En cuanto a la ropa y complementos, se incluyen términos como *ropa de temporada, estar pasado de moda, limpiar en seco.*
13. Salud e higiene: se incluyen términos relacionados con enfermedades y condiciones médicas más complejas como *diabetes, bronquitis, neumonía, depresión, enfermedad terminal.* También se amplía el vocabulario sobre tratamientos médicos como *hacer una transfusión, escayolar, tener una cicatriz.*
14. Viajes, alojamiento y transporte: vocabulario relacionado con tipos de viajes más complejos como *viaje oficial, viaje de negocios, turismo de aventura.* También se añaden términos relacionados con el transporte como *hacer escala, sala de embarque, transporte aéreo.*

15. Economía e industria: términos más avanzados de economía como *ingresos periódicos, inversor, acciones, abrir una cuenta corriente, solicitar un crédito*. También se incluyen términos industriales relacionados con la gestión económica de empresas.
16. Ciencia y tecnología: se amplía el vocabulario científico y tecnológico, incluyendo términos como *navegar en Internet, descargar archivos, antivirus, servidor, proveedor de acceso*. También se incluyen términos relacionados con la ciencia como *ecología, energía renovable*.
17. Gobierno, política y sociedad: vocabulario sobre política y sociedad como *gobierno, elecciones, ley, votar*. Se añaden términos más formales sobre los derechos ciudadanos y la burocracia, como *validar documentos, renovar el DNI, hacer trámites*.
18. Actividades artísticas: vocabulario relacionado con las artes, incluyendo términos como *montar una exposición, organizar un festival, participar en una obra de teatro*. También se introducen términos sobre eventos culturales como *estrenar, cancelar un espectáculo*.
19. Religión y filosofía: se amplía el vocabulario sobre las creencias y religiones con términos como *misa, confirmación, bodas de oro/plata, banquete nupcial*.
20. Geografía y naturaleza: se añaden términos más avanzados para describir el entorno natural, como *escalada de montaña, descenso de barrancos, ecología, energía renovable*. También se incluye vocabulario relacionado con la planificación de viajes y el transporte, como *recorrer el océano, ruta turística, cruzar fronteras*.

3.3. "PROMPTS" O ENTRADAS UTILIZADAS

Denominamos *prompt* al texto de entrada que utilizamos para realizar solicitudes a los modelos de lenguaje. Es esencial utilizarlos de manera efectiva a la hora de trabajar con los modelos de lenguaje, para ello hay que tener en cuenta una serie de información que proporcionarle al modelo: una

instrucción concreta, un contexto con información complementaria, la pregunta y el tipo de respuesta que buscamos (Giray, 2023, p. 2630).

Teniendo en cuenta esas indicaciones, el *prompt* utilizado en este estudio sería el siguiente: “Eres un docente de español como lengua extranjera y vas a preparar un texto de comprensión lectora para estudiantes de nivel XX del Marco Curricular del Instituto Cervantes”.

4. RESULTADOS

4.1. TEXTOS GENERADOS

A continuación, presentamos los textos generados por el modelo en respuesta a las consultas realizadas, correspondientes a los niveles de A1 a B2.

Textos de nivel A1. Generado por *ChatGPT*:

(1) La familia Pérez

La familia Pérez vive en una casa grande en el campo. Los Pérez son cinco: papá, mamá, Marta, Pedro y el perro Toby. Todos los días, papá y Pedro van al trabajo en el campo. Mamá se queda en casa y cuida de Marta, que es una niña pequeña. Toby es un perro muy cariñoso y siempre juega con los niños.

Los Pérez son una familia muy unida. Los domingos les gusta salir juntos a pasear por el campo y disfrutar del aire fresco. A veces, van al río y se bañan todos juntos. Les encanta pasar tiempo en familia y hacer actividades al aire libre.

Un día, los Pérez decidieron hacer una barbacoa en el jardín. Papá encendió el fuego, mamá preparó la comida y Marta y Pedro recogieron leña. Toby estaba muy emocionado y no paraba de mover la cola. Todos se sentaron alrededor de la mesa y disfrutaron de una deliciosa comida al aire libre.

Los Pérez son una familia feliz y siempre están juntos en los momentos importantes. Les gusta compartir momentos especiales y hacer cosas divertidas en familia. (OpenAI, 2024)

Generado por *LlaMa*:

- (2) Hola, soy Lucía. Vivo en Madrid. Me gusta leer libros de aventuras. Mi favorito es "El Quijote" de Miguel de Cervantes. Es un libro muy divertido y emocionante. ¿Sabías que es el libro más antiguo de España? ¡Es verdad! Lo escribió en el siglo XVII. Me gusta leerlo en el parque o en la biblioteca. A veces, me imagino que soy Sancho Panza, el compañero de Don Quijote. ¡Es divertido! (MetaAI, 2024)

Podemos observar que los textos tienen una longitud distinta, siendo el generado por *LlaMa* menor, con unas 70 palabras y el texto generado por *ChatGPT* cuenta con unas 186 palabras. Si tomamos de ejemplo la comprensión lectora del Diploma de Español como Lengua Extranjera (DELE), en el nivel A1 tenemos textos de entre 20 y 175 palabras (Instituto Cervantes, 2019, pp. 8-9). Por lo que ambos textos podrían encontrarse dentro de las medidas del DELE, a pesar de que el texto de *ChatGPT* supere brevemente el número de palabras.

Textos de nivel A2. Generado por *ChatGPT*:

- (3) El Parque Nacional

El Parque Nacional es un lugar hermoso y tranquilo donde la naturaleza es la protagonista. En este parque, los visitantes pueden disfrutar de largas caminatas por senderos rodeados de árboles, cascadas y animales silvestres. También se pueden observar aves exóticas y plantas únicas en su hábitat natural.

Los guardaparques se encargan de proteger y preservar la flora y fauna del parque, así como de mantener limpias las áreas de descanso y los senderos. Además, organizan actividades educativas para enseñar a los visitantes la importancia de cuidar el medio ambiente y respetar la vida silvestre.

Si decides visitar el Parque Nacional, recuerda llevar agua, protector solar y ropa cómoda para disfrutar al máximo de la experiencia. ¡No olvides tu cámara para capturar los momentos más especiales de tu visita! (OpenAI, 2024)

Generado por *LlaMa*:

(4) Hoy he ido a la playa con mis amigos. El sol brillaba en el cielo. Me puse el protector solar y la gorra para no quemarme. Me gustan las playas con arena blanca y el mar azul. Allí había muchas personas nadando y jugando en el agua.

Me senté en una sombrilla con mis amigos. Comimos sandwiches y bebimos agua. Después de comer, fuimos a nadar. Me gustó mucho el agua fría. Me sentí cansado, así que volví a la sombrilla a descansar.
(MetaAI, 2024)

En estas muestras podemos apreciar que el texto generado por *LlaMa* vuelve a ser el de menor tamaño con unas 83 palabras, frente al generado por *ChatGPT* con 128 palabras. Si los comparamos con los textos de comprensión lectora del DELE A2, en este, podemos encontrar una variedad de textos de entre 50 y 80 palabras o textos entre 375 y 425 palabras (Instituto Cervantes, 2022, pp. 8-9). Por lo que podemos decir que las muestras de los modelos podrían encajar en los textos de las tareas, siempre que no se tomasen como textos únicos de la tarea, en especial el generado por *LlaMa*, ya que cuentan con un número menor de palabras.

Textos de nivel B1. Generado por *ChatGPT*:

(5) La Fiesta de la *Vendimia*

En el pueblo de Villafranca, cada año se celebra la Fiesta de la Vendimia en honor a la *cosecha* de uvas. Durante esta festividad, los habitantes del pueblo y visitantes de otras *regiones* se reúnen para participar en diversas actividades relacionadas con la vendimia.

La fiesta comienza con un desfile de carrozas decoradas con *racimos* de uvas y personas vestidas con trajes tradicionales. Luego, se lleva a cabo la tradicional *pisada* de uvas, donde los participantes aplastan las uvas con los pies para extraer el *mosto*.

Además, hay concursos de cata de vinos, donde los expertos y aficionados prueban diferentes variedades de vino producidas en la región. También se organizan *espectáculos de música* y baile folclórico para animar la celebración.

La Fiesta de la Vendimia es una oportunidad para disfrutar de la *cultura local*, degustar deliciosos *platos típicos* y conocer el proceso de elaboración del vino. ¡Una experiencia única que no te puedes perder! (OpenAI, 2024)

Generado por *LlaMa*:

- (6) Hoy he decidido *explorar* la ciudad. Me levanté temprano y tomé un café en el bar de la esquina. Después de desayunar, fui al *museo de arte* para ver la exposición de *pinturas del siglo XIX*. Me gustó mucho la obra de Velázquez. Me sentí fascinado por *la técnica y la creatividad* de los artistas.

Luego, me dirigí al parque central para hacer un paseo. El sol brillaba en el cielo y la gente se reía y hablaba en las bancas. Me senté en una de ellas y observé a los niños jugando en el parque. Me recordó a mis propios días de niño.

Después de un rato, decidí ir al centro comercial para hacer algunas compras. Compré un regalo para mi hermana y un libro de historia que había leído horas atrás pero que quería re leer de nuevo. (MetaAI, 2024)

Con respecto a la muestra del nivel B1, podemos destacar que los textos generados por los modelos *ChatGPT* y *LlaMa* cuentan con una extensión de 154 y 140 palabras respectivamente. Teniendo de referencia los textos de las tareas de comprensión escrita del DELE B1 (Instituto Cervantes, 2014a, pp. 8-9), estos textos podrían ser utilizados en tareas en las que los alumnos deban leer múltiples textos, ya que se encuadran en la extensión de este

tipo de tareas (tarea 3, 100-120 palabras por texto, tres textos en la tarea), sin embargo, se aleja de las tareas que cuentan con un único texto, ya que estas cuentan con un total de 400 y 450 palabras.

Textos de nivel B2. Generado por *ChatGPT*:

(7) El Arte Urbano: Expresión Creativa en la Ciudad

El arte urbano, también conocido como arte callejero o street art, es una forma de expresión artística que se desarrolla en espacios públicos como calles, paredes y edificios. A través de murales, grafitis, estenciles y otras técnicas, los artistas urbanos transmiten mensajes y emociones al público en general.

Este tipo de arte ha cobrado gran relevancia en las últimas décadas, convirtiéndose en una manifestación cultural y social que transforma el entorno urbano y despierta reflexiones en la sociedad. Muchos artistas urbanos utilizan sus obras para abordar temas como la injusticia social, la diversidad cultural, el medio ambiente y la libertad de expresión.

Además de embellecer las ciudades, el arte urbano invita a la reflexión y promueve el diálogo entre diferentes grupos sociales. Algunas ciudades han promovido el arte urbano como una forma de revitalizar espacios abandonados y fomentar el turismo cultural.

A través del arte urbano, los artistas encuentran una plataforma para expresar sus ideas y conectar con la comunidad de una manera única y directa. ¡Descubre la creatividad y la diversidad del arte urbano en tu ciudad! (OpenAI, 2024)

Generado por *LlaMa*:

(8) La ciudad de Madrid es la capital de España y es conocida por su rica historia, su arquitectura y su cultura. La ciudad tiene un clima mediterráneo, con veranos cálidos y inviernos suaves. La gente de Madrid es amable y acogedora, y se puede disfrutar de la ciudad en cualquier época del año.

Una de las atracciones turísticas más famosas de Madrid es el Museo del Prado, que alberga obras maestras de pintores como Goya, Velázquez y El Greco. Otro lugar emblemático es el Palacio Real, que ha sido la residencia de la familia real española durante siglos. La noche en Madrid es también muy animada, con bares y restaurantes en los que se puede disfrutar de la música en vivo y la gastronomía local.

Sin embargo, la ciudad también tiene sus problemas, como la contaminación y la congestión del tráfico. Los madrileños están trabajando para reducir el impacto ambiental de la ciudad y hacerla más sostenible. Algunos de los proyectos más interesantes incluyen la creación de espacios verdes en el centro de la ciudad y la implementación de un sistema de transporte público eficiente. (MetaAI, 2024)

Por último, en el nivel B2, si consultamos el número de palabras en las tareas del DELE B2, en tareas en las que se incluye más de un texto, se encuentra entre 130 y 150 palabras por texto, mientras que en las tareas donde hay un único texto, cuenta entre 400 y 450 palabras (Instituto Cervantes, 2014b, pp. 8-9). En las muestras obtenidas por los modelos, *ChatGPT* nos presenta un texto de 186 palabras y *LlaMa* un texto de 185 palabras. Ambos textos se encontrarían lejanos a los utilizados en las tareas del DELE de texto único.

4.2. ANÁLISIS CONTRASTIVO DEL PLAN CURRICULAR DEL INSTITUTO CERVANTES

Una vez que los modelos nos han proporcionado las distintas muestras textuales, nos disponemos a realizar el análisis de estos con respecto a las nociones específicas del Plan Curricular del Instituto Cervantes.

En el análisis de los textos del nivel A1, podemos observar diferencias entre los producidos por *ChatGPT* y *LlaMa*. En primer lugar, el texto generado por *ChatGPT* contiene un vocabulario simple y acorde con el nivel A1, lo que sugiere un buen ajuste inicial a los requisitos de este nivel. Sin embargo, es importante señalar la presencia de algunos términos y construcciones que superan las competencias esperadas para este nivel. Por ejemplo,

el uso de la expresión *pasear por el campo* se podría clasificar dentro del nivel A2 de acuerdo con el *Plan Curricular del Instituto Cervantes* (PCIC). De igual manera, el verbo *encantar* también es propio de un nivel superior, ya que su uso es más complejo que el esperado en A1. Además, es relevante destacar la inclusión de tiempos verbales en pasado, lo cual no corresponde al utilizado en este nivel.

Por otro lado, el texto generado por *LlaMa* también presenta algunos desajustes en relación con el nivel A1. Aunque se observa una tendencia a utilizar vocabulario sencillo, encontramos el uso del verbo *gustar*, que según el PCIC se clasifica como propio del nivel A2. Asimismo, en el texto se identifican términos como *antiguo*, *siglo* y *biblioteca*, que no se ajustan adecuadamente al nivel A1, ya que son palabras que requieren un conocimiento léxico más avanzado. Estos términos sugieren un léxico más abstracto y relacionado con conceptos temporales e históricos que normalmente se introducen en niveles posteriores del aprendizaje de una lengua.

Por lo tanto, si bien ambos textos presentan elementos que se ajustan al nivel A1, tanto el generado por *ChatGPT* como el de *LlaMa* contienen vocabulario que excede las competencias esperadas para este nivel.

En cuanto a los textos correspondientes al nivel A2, observamos diferencias significativas entre los textos generados por *ChatGPT* y *LlaMa*. Por un lado, el texto producido por *ChatGPT* contiene, en su mayoría, términos que efectivamente se corresponden con el nivel A2, lo que cumple parcialmente con las expectativas para este nivel. No obstante, es importante destacar que también se identifican elementos de vocabulario que exceden las limitaciones de este nivel. Por ejemplo, términos como *senderos*, *cascada*, *silvestre* o *aves exóticas* se encuentran dentro de un léxico más complejo y podrían catalogarse como propios de niveles superiores debido a

la especificidad y el detalle que requieren, tanto en comprensión como en producción.

Por otro lado, el texto generado por *LlaMa* presenta una estructura más sencilla y ajustada al nivel A2. Se observa un uso predominante de términos más accesibles y frecuentes en dicho nivel, tales como *playa*, *cielo*, *gustar*, *nadar* o *estar cansado*. Este tipo de vocabulario es más coherente con los objetivos de un nivel elemental, en el que se espera que los aprendices se familiaricen con palabras y expresiones comunes en contextos cotidianos y de fácil comprensión.

A partir de esta comparación, es posible señalar que, en lo que respecta a la fidelidad con el nivel A2, el texto generado por *LlaMa* se aproxima de manera más efectiva a las expectativas del nivel que se le ha solicitado. La simplicidad de los términos utilizados en *LlaMa* refleja una adecuación más precisa a las competencias lingüísticas que se espera que los aprendices adquieran en este estadio del aprendizaje de una lengua extranjera. Por lo tanto, si bien el texto de *ChatGPT* incluye vocabulario relevante, su nivel de complejidad podría generar dificultades para estudiantes en el nivel A2, lo que sugiere que la generación de textos por parte de este modelo puede requerir ajustes para cumplir de manera más exacta con los objetivos de este nivel.

En el análisis del nivel B1, se destacan diversos aspectos relevantes en los textos generados por *ChatGPT* y *LlaMa*. En primer lugar, el texto elaborado por *ChatGPT* muestra una notable correspondencia con las competencias lingüísticas y gramaticales esperadas para el nivel B1. La mayor parte del vocabulario empleado se ajusta adecuadamente a las exigencias de este nivel, lo que refleja una alineación con los parámetros del PCIC. No obstante, se han identificado algunos términos que podrían ser considerados más avanzados y propios del nivel B2, tales como *vendimia*, *cata de vinos*,

y el verbo *degustar*. Estas palabras están asociadas con un vocabulario más especializado y contextos culturales concretos, como el de la gastronomía y el vino, lo que las hace más apropiadas para un nivel superior debido a su complejidad y especificidad.

Por su parte, el texto generado por *LlaMa* también incorpora varios elementos léxicos y expresiones que se ajustan al nivel B1, como las frases *me recordó* y *me sentí fascinado*, que son adecuadas para describir emociones y sensaciones de manera más precisa, acorde con la capacidad expresiva que se espera en este nivel intermedio. Sin embargo, es importante subrayar que el texto de *LlaMa* incluye vocabulario propio de niveles más básicos, como el nivel A1, con ejemplos como *ir al parque*. Este tipo de léxico es demasiado elemental y no refleja el grado de complejidad esperado en un texto de nivel B1. Además, se observan términos más cercanos al nivel B2, como *exposición de pinturas* y expresiones como *técnica y creatividad de los artistas*, que requieren un conocimiento más abstracto y técnico relacionado con el arte y la cultura, lo que supera la competencia lingüística del nivel B1.

En conclusión, al comparar los textos generados por *ChatGPT* y *LlaMa* para el nivel B1, se puede afirmar que el texto de *ChatGPT* se ajusta más consistentemente a los criterios establecidos por el PCIC, subrayadas en el apartado anterior, podemos observar términos que se encontrarían en el nivel B1 relacionadas con alimentación y actividades artísticas. Aunque ambos textos presentan elementos válidos para el nivel B1, el de *ChatGPT* muestra una mayor coherencia y uniformidad en el uso del vocabulario, mientras que el texto de *LlaMa* presenta una fluctuación más notable, al incluir términos que oscilan entre niveles inferiores y superiores a B1, lo que afecta su adecuación

a los estándares requeridos para este nivel lingüístico. Al igual que en el texto de *ChatGPT*, se puede observar los términos de B1 subrayados en el texto generado por *LlaMa*.

Con respecto al nivel B2, el texto generado por *ChatGPT* se encuentra más orientado a nivel B2, con términos como *arte callejero* o un listado de técnicas artísticas como *murales*, *grafitis* o *esténciles*. Con respecto al texto de *LlaMa*, encontramos también léxico del nivel B2 como *la residencia de la familia real durante siglos*, *sostenible* y *eficiente*.

En ambos textos, encontramos asimismo términos de nivel B1, pero con predominio del nivel B2, por lo que podemos decir que ambos modelos presentan una adecuación con el nivel que se le ha solicitado.

4.3. COMPLEJIDAD Y VARIEDAD LÉXICA

Para analizar la complejidad y la variedad léxica, se ha utilizado la herramienta *Textflow*, desarrollado por el grupo de Sistemas Inteligentes de Acceso a la Información (s. f). A continuación, presentamos una tabla con los resultados generados:

Como podemos observar en la tabla 1, con respecto a la complejidad de frases, el modelo *LlaMa* presenta un número más elevado en comparación con el texto del mismo nivel generado por *ChatGPT*. No obstante, en el resto de métricas, el modelo *ChatGPT* presenta una complejidad superior, como en el uso de signos de puntuación y de palabras.

Con respecto a la variedad léxica, podemos ver que es similar en ambos modelos, con una variedad más elevada en los textos generados por *LlaMa*. Sin embargo, en los niveles A1 y A2 la diferencia entre ambos modelos es más pronunciada.

Nivel	Modelo	Compleji-dad de la frase	Compleji-dad signos de puntuación	Compleji-dad de palabras	Compleji-dad de palabras extrañas	Compleji-dad de sílabas	Diversidad léxica simple
A1	LLaMA	10	19	70	15	129	0.09375
A1	GPT-3.5	15	24	192	34	348	0.033497536945812800
A2	LLaMA	10	12	85	25	145	0.07356321839
A2	GPT-3.5	7	14	136	46	285	0.0430622009
B1	LLaMA	11	14	143	33	248	0.04663923182
B1	GPT-3.5	8	16	165	46	338	0.03744939271
B2	LLaMA	9	18	188	43	368	0.03130755064
B2	GPT-3.5	8	21	191	52	418	0.03091060985

TABLA 1: Resultados Textflow de complejidad y variedad léxica

5. CONCLUSIONES

Como conclusiones, podemos determinar varios aspectos.

En primer lugar, hay que señalar que ambos modelos no solo nos han proporcionado el texto solicitado, sino que nos han dado una serie de preguntas para evaluar la comprensión lectora de nuestros estudiantes, dichas preguntas nos sirven de objeto de análisis en estudios posteriores.

En segundo lugar, hemos de remarcar la necesidad de realizar una entrada adecuada al modelo, para que este se ajuste a la respuesta que necesitamos. Como podemos comprobar los textos generados por los modelos son textos narrativos, sin embargo, sería interesante añadir la modalidad textual en la entrada siguiendo el PCIC, como por ejemplo en nivel A1 un folleto con información turística. No obstante, con una entrada adecuada estos materiales generados pueden adaptarse a los distintos niveles y, asimismo, motivar a los alumnos, a la lectura, ya que se puede solicitar textos de interés para ellos.

En tercer lugar, hay que destacar que el modelo *ChatGPT* muestra un vocabulario más adecuado al nivel solicitado en los niveles intermedios B1 y B2, mientras que *LlaMa* no se ajusta a los estándares de estos niveles, dándonos un vocabulario de nivel inferior y superior. En los niveles básicos A1 y A2, ambos niveles nos devuelven un texto con vocabulario superior al solicitado, por lo que requieren de un ajuste.

En conclusión, los modelos de lenguaje pueden servirnos de apoyo a los docentes a la hora de elaborar material, no obstante, hay que realizar una revisión al mismo, puesto que, como hemos comprobado, no siempre se ajusta a los estándares del *Plan Curricular del Instituto Cervantes* y sería necesario una revisión manual de los mismos.

REFERENCIAS BIBLIOGRÁFICAS

- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección de vocabulario. En R. M.^a Castañer & V. Lagüéns (eds.), *De moneda nunca usada: estudios dedicados a José M.^a Enguita Utrilla* (pp. 85-107). Instituto Fernando El Católico-CSIC.

- ChatGPT: usuarios y estadísticas (2024). (2024, 17 mayo). <https://n9.cl/sdwf7>
- Fernández, Y. (2023, 21 abril). Qué es LLaMA, cómo funciona y qué sabemos de la inteligencia artificial de Meta. *Xataka*. <https://n9.cl/u2bnb>
- Fredholm, M. (2024). ChatGPT: Aplicación y percepción de su uso en el aula de español como lengua moderna: Un estudio sobre la producción escrita en la escuela básica en Suecia [Proyecto Fin de Carrera. Universidad de Dalarna. Escuela de Formación del profesorado]. DiVA. Dalarna University's publications database. <https://n9.cl/2z8jn>
- Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals Of Biomedical Engineering*, 51(12), 2629-2633. <https://doi.org/10.1007/s10439-023-03272-4>
- Instituto Cervantes (2006) *Plan curricular del Instituto Cervantes*. Instituto Cervantes. <https://n9.cl/6sl6a>
- Instituto Cervantes (2014a). Guía del examen DELE B1. Instituto Cervantes. <https://n9.cl/psafgs>
- Instituto Cervantes (2014b). Guía del examen DELE B2. Instituto Cervantes. <https://n9.cl/nkc3n>
- Instituto Cervantes (2019). Guía del examen DELE A1. Instituto Cervantes. <https://n9.cl/wgp3t>
- Instituto Cervantes (2022). Guía del examen DELE A2. Instituto Cervantes. <https://n9.cl/7282m6>
- Jiménez Linares, L., López-Gómez, J. L., Martín-Baos, J. Á., Romero, F. P. & Serrano-Guerrero J. (2023). ChatGPT: reflexiones sobre la irrupción de la inteligencia artificial generativa en la docencia universitaria. En J. A. Cruz Lemus, N. Medina Medina & M.^a J. Rodríguez Fórtiz (eds.), *Actas de las XXIX Jornadas sobre la Enseñanza Universitaria de la Informática* (pp. 113-120). Asociación de Enseñantes Universitarios de la Informática. <https://n9.cl/kex4sy>
- Meta AI (2024). LLaMA (versión 3.8b) [Modelo de lenguaje de gran tamaño]. <https://llama.meta.com/>
- Moreno Villanueva, J. A. (2021). Léxico disponible y evaluación del vocabulario en estudiantes sinohablantes de ELE. *RLA. Revista de Lingüística Teórica y Aplicada*, 59(2), 13-42. <https://doi.org/10.29393/rla59-9ldjm10009>
- Muñoz-Basols, J. & Fuertes Gutiérrez, M. (2024). Oportunidades de la Inteligencia Artificial (IA) en la enseñanza y el aprendizaje de lenguas. En J. Muñoz-Basols, M. Fuertes Gutiérrez & L. Cerezo (eds.), *La enseñanza del español mediada por tecnología: de la justicia social a la Inteligencia Artificial (IA)* (pp. 343-365). Routledge. <https://doi.org/10.4324/9781003146391>
- Núñez Delgado, M.^a P. & Rodríguez Guerrero, B. (2017). La comprensión lectora en ELE. Análisis de manuales de B1 con base en su tipología textual. *Revista Internacional de Lenguas Extranjeras*, 1(6), 31-54. <https://doi.org/10.17345/rile6.31-54>
- OpenAI. (2024). ChatGPT (versión GPT-3.5) [Modelo de lenguaje de gran tamaño]. <https://chat.openai.com/chat>

- Ribes Lafoz, M.^a A. & Navarro Colorado, B. (2023). Aprovechamiento de ChatGPT en la enseñanza de lengua extranjera en educación superior. En D. Ortega-Sánchez & A. López-Padrón (eds.), *Educación y Sociedad: Claves Interdisciplinarias* (pp. 1264-1271). Octaedro. <https://n9.cl/r9msog>
- Sistemas Inteligentes de Acceso a la Información (SINAI). (s. f.). GitHub - sinai-uja/textflow. GitHub. <https://n9.cl/rcfudb>

