Uso de corpus propios en la enseñanza del español como lengua extranjera

Filip Zeman

Máster en Lengua Española: Investigación y Prácticas Profesionales

MÁSTERES DE LA UAM

2021-2022

Facultad de Filosofía y Letras

UAM Universidad Autónoma de Madrid

MÁSTER EN LENGUA ESPAÑOLA: INVESTIGACIÓN Y PRÁCTICAS PROFESIONALES

Uso de corpus propios en la enseñanza del español como lengua extranjera

Filip Zeman

Tutora: Dra. Olga Batiukova

13 de junio de 2022

Índice

0. Agradecimientos

En primer lugar, quiero darle las gracias a mi tutora, la Dra. Olga Batiukova, quién me ayudó desde el inicio en la realización de este trabajo.

Agradezco también a Sergio Palacios (el director de la academia de español Inhispania) y a Lucía Maya (profesora de esta academia) por facilitarme la realización de la encuesta para este estudio, y a todos los estudiantes que han participado en la encuesta por su esfuerzo. Me han proporcionado información muy valiosa, sobre la que se basa la propuesta que presento en este trabajo.

Por último, quiero expresar mi gratitud a los profesores del Máster de Lengua Española: Investigación y Prácticas Profesionales de la Universidad Autónoma de Madrid, quienes me formaron en la materia y me proporcionaron los conocimientos necesarios para poder elaborar este Trabajo de Fin de Máster.

1. Introducción

En el ámbito del español como segunda lengua, la enseñanza del léxico ha experimentado un desarrollo notable en las últimas décadas y se ha beneficiado de la introducción de los enfoques modernos de enseñanza. Se han introducido nuevas maneras de entender y enseñar el léxico, de diversificar las tareas, de introducir el factor afectivo y de desarrollar el aprendizaje autónomo. Este trabajo sigue la línea de estos nuevos enfoques: teniendo en cuenta los resultados de los recientes estudios sobre la adquisición del léxico y los beneficios de la diversificación metodológica, en él se formulará una novedosa propuesta de enseñanza del léxico a aprendientes del español de nivel avanzado a través de la creación y consulta de corpus lingüísticos propios.

El trabajo se organiza del siguiente modo. En la sección 2 se introducen los aspectos básicos de la enseñanza y el aprendizaje del léxico tal y como se formulan dentro del *enfoque léxico*. En la misma sección se presenta la noción de *corpus lingüístico* y se revisan los estudios que han tratado sobre el uso de corpus en el aula de lenguas extranjeras (LE). La sección 3 está centrada en lo que sabemos hoy en día sobre el uso de corpus propios en la enseñanza de idiomas y el aprendizaje del léxico en niveles avanzados de competencia. En la sección 4 profundizamos en estas cuestiones a través de los resultados de la encuesta sobre la adquisición del léxico diseñada para este trabajo. Por último, en la sección 5 presentamos la propuesta didáctica de uso de corpus propios con ayuda del sistema de consulta y gestión de corpus Sketch Engine.

2. Consideraciones teóricas generales

2.1. El enfoque léxico

El método conocido como el *enfoque léxico*, desarrollado por Michael Lewis, ha marcado la enseñanza del léxico en los últimos treinta años. Pese a que se han modificado o abandonado algunos de sus postulados originales, las tesis principales del enfoque, que enfatizan la necesidad de trazar una estrategia compleja de enseñanza del léxico, han sido reconocidas universalmente y siguen vigentes en el presente. Cualquier estudio sobre la enseñanza-aprendizaje del léxico y, por extensión, cualquier actividad docente de este ámbito, se inspira en este método. En este sentido, la propuesta de uso de los corpus propios para la enseñanza del léxico que se desarrolla en el presente trabajo no es una excepción.

El enfoque léxico fue introducido por Michael Lewis en sus dos publicaciones: *The Lexical Approach* (Lewis, 1993) y *Implementing the Lexical Approach* (Lewis, 1997). Uno de los aciertos de Lewis fue reconocer la existencia de *unidades léxicas* complejas y hablar de estas en vez de *palabras* (Pérez Serrano, 2017: 9). Wray define las unidades léxicas como "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" (2002: 9). La metodología de Lewis no constituye un mero cambio terminológico: implica una transición en la percepción del léxico desde la concepción tradicional de palabras aisladas que se insertan dentro de los huecos de los patrones estructurales (Pérez Serrano, 2017: 9, 70) hacía una visión en la que el léxico es un sistema relacional de inmensa complejidad.

Según la hipótesis de Lewis, el cerebro humano almacena no solo palabras simples sino también unidades de mayor extensión, que se denominan *bloques léxicos* (o *chunks* en inglés). Esto significa que la mente humana almacena locuciones como *claro que sí* como una unidad, en vez de recordar cada palabra de forma aislada; en la recepción o producción, la forma se extrae de la memoria como una unidad indivisible. Los beneficios para el aprendiente de lenguas extranjeras son numerosos. En primer lugar, los bloques léxicos son claves para la fluidez expresiva, porque permiten procesar el léxico de forma más rápida y eficiente. Los hablantes nativos se expresan de esta forma automatizada y recuperan los bloques prefabricados de la memoria como si fueran números de teléfono (Pérez Serrano, 2017: 25). En segundo lugar, este tipo de almacenamiento léxico encaja mejor con la idea de un *lexicón mental* que es una red de ideas interconectadas y no una lista de palabras. El lexicón mental no solo incluye las unidades léxicas, sino también información sobre todas las relaciones que existen entre una palabra y otra, sean esas horizontales (sinonimia, antonimia) o verticales (hiperonimia, hiponimia, meronimia). Un modelo del léxico que refleje las conexiones entre las palabras y sus combinaciones es más adecuado para dar cuenta de la retención durante el aprendizaje.

La fijación de los rasgos semánticos y sintácticos de una palabra se conoce como *idiomaticidad*, término introducido por Sinclair dos años antes de que surgiera el enfoque léxico de Lewis. La idiomaticidad funciona en el sentido opuesto a la *selección libre*. Según la tesis de la selección libre, el hablante dispone de libertad casi absoluta a la hora de hablar y elegir palabras y estructuras; las únicas restricciones que existen son las gramaticales. Sin embargo, los factores idiomáticos restringen estas posibilidades combinatorias. Por ejemplo, para despedirnos solemos recurrir a expresiones comúnmente usadas en nuestra comunidad lingüística, aunque en un principio la gramática podría generar muchas otras.

La aparición de propuestas innovadoras de Sinclair y Lewis fue propiciada por el auge de la lingüística de corpus en la década de los noventa. Los corpus, en su forma digitalizada, han permitido demostrar que existe una tendencia evidente a que ciertas palabras coaparezcan en un contexto específico (Pérez Serrano, 2017). Esta idea fue el punto de partida de la tesis de idiomaticidad de Sinclair y sirvió como base para la definición del concepto de unidad léxica en términos de Lewis. Los estudios posteriores confirmaron la idoneidad de estas propuestas: hoy, gracias al análisis de corpus, ya sabemos que casi la mitad de nuestra producción lingüística 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 y Jiménez Calderón, 2017: 48). La naturaleza formulaica de la lengua se refleja no solo en las expresiones fijas –por ejemplo, las locuciones– sino también en unidades léxicas y combinaciones de palabras con un grado de fijación más bajo. Retomaremos estas consideraciones en las secciones 3 y 5.

Otro de los grandes aciertos de la propuesta de Lewis ha sido su insistencia en la importancia de la autonomía del aprendiz. Partía de la idea de que existen tantas unidades léxicas que es imposible enseñarlas todas en clase. Si en una hora de clase los alumnos suelen aprender entre diez y doce palabras (Chamorro, 2017: 71) y el vocabulario de un estudiante de nivel C1 debe contener unas 10 000 unidades léxicas (según el *MCER*), su enseñanza requeriría unas 1000 horas o incluso más, si tenemos en cuenta la polisemia. Por la cantidad de unidades léxicas que existen, Lewis argumentaba que tenía más sentido entrenar al aprendiz para que piense *sobre* el léxico: identifique y segmente los bloques y después deduzca su significado y función. Según el autor, esta estrategia se puede explicar en un tiempo reducido y el alumno puede ponerla en práctica cada vez que se encuentre con una unidad léxica desconocida. Hoy en día, el aprendizaje autónomo sigue gozando de mucho apoyo y la mayoría de las publicaciones sobre la enseñanza del léxico hacen referencia a él. Es una estrategia que se usa ampliamente en la práctica. Por ejemplo, los exámenes de inglés de la Universidad de Cambridge, más específicamente los del nivel más alto –*CAE* (*Cambridge Advanced Certificante in English*) y *CPE* (*Cambridge Certificate of Proficiency in English*), correspondientes a los niveles C1 y C2 del *MCER*– incluyen tareas en las que se comprueban si el alumno es capaz de realizar el análisis léxico de forma autónoma. Este tipo de tareas contiene léxico que los alumnos no conocen, para valorar su capacidad de determinar su función o significado.

Otra aportación del enfoque léxico está relacionada con la polisemia y la *enseñanza cualitativa* del léxico (Álvarez Cavanillas, 2017). En los métodos de enseñanza precedentes

(hasta el método estructural) se ponía énfasis en la *enseñanza cuantitativa* de la lengua, cuyo fin es facilitar una preparación rápida para un alto número de aprendientes para que estos puedan desenvolverse en situaciones comunicativas básicas. Por lo tanto, se enseñaba a usar palabras generales e imprecisas y se toleraba la vaguedad expresiva (Pérez Serrano, 2017: 31). Dentro del enfoque léxico se aboga por una *enseñanza cualitativa*: se presta mucha atención a la precisión en el uso léxico (como un factor clave para alcanzar el dominio en una lengua extranjera, como veremos en el apartado 3.2.) y son frecuentes los ejercicios léxicos en los que los alumnos tienen que sustituir palabras de sentido general por unidades léxicas más específicas.

El último elemento que trajo consigo el enfoque léxico es la propia conciencia de lo que significa aprender una unidad léxica. Es quizás su aportación más relevante, pero la hemos dejado para el final para poder contextualizarla en el marco de sus planteamientos teóricos básicos. Ya se considera anticuada e inútil la concepción según la cual el léxico es una simple lista de palabras y conocer una palabra equivale a conocer su significante, significado y, en el caso de LE, su traducción a la lengua materna del alumno (Tarrés, 2017: 35). Como hemos visto, la realidad es mucho más compleja porque las palabras o unidades léxicas se relacionan entre sí en el plano paradigmático y el sintagmático, y existen rasgos inherentes a cada palabra, en principio ajenos a su significado, que determinan su uso (los factores diatópicos, diafásicos y diastráctos, entre otros). También es necesario recordar que existen unidades léxicas que no pueden traducirse sin más a la lengua materna (LM) del alumno (por inexistencia de un equivalente directo, por diferencias socioculturales, etc.). Bogaards concibe el aprendizaje de una unidad léxica como un proceso que incluye los siguientes factores:

> aprender una nueva forma para un nuevo significado; aprender un nuevo significado para una forma ya conocida; aprender un nuevo significado para una combinación de formas ya conocidas; aprender relaciones semánticas entre unidades léxicas […]; aprender relaciones morfológicas entre unidades léxicas; aprender los usos gramaticales correctos de las unidades léxicas; aprender las colocaciones; aprender el uso pragmático apropiado. (Bogaards, 2001: 327-8).

Marta Higueras desarrolla la propuesta de Bogaards en forma de la siguiente lista, que abarca otros tipos de información léxica:

1. La denotación y referencia.
2. El sonido o grafía, según el canal.
3. Las estructuras sintácticas en las que aparece.

4. Las peculiaridades morfológicas.
5. Las relaciones paradigmáticas con las unidades que podrían haber aparecido en su lugar.
6. Las combinaciones sintagmáticas o palabras con las que normalmente se asocia.
7. El registro y si tiene más probabilidades de aparecer en la lengua hablada o escrita o ambas.
8. El contenido cultural.
9. Los usos metafóricos.
10. La frecuencia de uso.
11. La pertenencia a expresiones institucionalizadas. (Higueras, 2009: 114)

En el siguiente subapartado, nos referiremos a las críticas que se han hecho del enfoque léxico durante los treinta años de su existencia.

2.1.1. Críticas al enfoque léxico

Con el paso del tiempo, varias de las ideas propuestas por Lewis han sido revisadas. Algunas críticas tienen que ver con el papel secundario que tiene el profesor dentro de este modelo debido a importancia otorgada a la autonomía del alumno (Pérez Serrano, 2017). Otras están relacionadas con la disociación que introduce Lewis entre el léxico y la gramática: propone entrenar a los estudiantes para que segmenten los bloques léxicos, pero no para que piensen sobre su estructura gramatical. Han sido también muy criticadas las premisas sobre la adquisición del léxico de las que parte Lewis para justificar su enfoque. Como hemos mencionado, el autor propone que el docente dedique el tiempo de clase a enseñar a los alumnos a identificar y segmentar las unidades léxicas. Los estudiantes deben anotar en su cuaderno las unidades léxicas junto con la información sobre las relaciones que surgen entre ellas, y para afianzar la asimilación se recurre a la exposición (el *aprendizaje incidental*). En esto los corpus juegan un papel importante, al ser una fuente de muestras del lenguaje natural. Sin embargo, la idea de asimilación por exposición es problemática, sobre todo en el caso de las unidades léxicas pluriverbales, porque se recuerdan con mayor dificultad que las palabras sueltas (Pérez Serrano, 2017). Además, la realidad del aula frecuentemente no permite (incluso con los recursos de los corpus) una exposición suficiente al input, para que se logre la retención. Se especula que un alumno tiene que encontrarse con una palabra o unidad léxica entre 6 y 10 veces para asegurar la retención (Pérez Serrano, 2017: 50). En una enseñanza pasiva, el número de encuentros necesarios sería aún más alto. Por lo contrario, la retención se producirá más rápidamente si el alumno tiene que realizar alguna operación cognitiva con la unidad léxica adquirida, como en

el método inductivo de la asimilación del léxico que se propone dentro del método comunicativo. En el planteamiento original de Lewis, este tipo de actividades no estaban previstas y fueron introducidas más tarde.

Otro problema importante de la propuesta de Lewis es la falta de especificidad sobre el tipo de léxico que se debería enseñar: se nombra vagamente la utilidad como el único criterio. Este problema se lo plantean la mayoría de los estudiosos que analizan el aprendizaje de léxico: véase Rufat y Jiménez Calderón (2017), Troitiño (2017), Cruz Piñol (2012), Serradilla Castaño (2014), etc. Las respuestas que se ofrecen se suelen enmarcar dentro la perspectiva cuantitativa o la cualitativa. La primera "busca las palabras más rentables para cada tipo y nivel de enseñanza […] a través del análisis de frecuencia de los corpus de la lengua" (Troitiño, 2017: 147). *A priori*, es una estrategia válida: como ha demostrado Nation (2006: 13-15), las 1000 palabras más frecuentes en inglés permiten entender un 74% de textos en esta lengua. No obstante, la perspectiva cuantitativa también presenta ciertos retos. Por ejemplo, Serradilla Castaño, en su trabajo centrado en la fraseología, reconoce la utilidad del corpus CREA, pero menciona al mismo tiempo que existen expresiones frecuentemente usadas que aparecen poco reflejadas allí, por lo que resultan indispensables la experiencia y el conocimiento del profesor (2014: 81). En la misma línea, varios autores hablan de los listados de frecuencias como una herramienta útil, pero recomiendan cautela a la hora de usar sus datos. Cruz Piñol (2012: 109), por ejemplo, lo ilustra a través del ejemplo de los pronombres personales *yo* y *vosotros*. En el listado de frecuencias de cualquier corpus del español el pronombre *yo* tiene una frecuencia mucho mayor que *vosotros*, pero esto no significa se deba enseñar cada uno en un nivel diferente de la lengua. Por su parte, la estrategia cualitativa –que defienden Troitiño (2017) o Rufat y Calderón (2017)–, consiste en enseñar las unidades léxicas más rentables para el aprendiente en función de sus intereses y objetivos comunicativos. En este caso, el proceso de recopilar una base léxica coherente es aún más complicado.

El enfoque léxico se encuentra, hoy en día, en una fase ya consolidada: se reconoce como una aportación importante a la enseñanza de LE, pero en sus treinta años de historia ya ha sido sustancialmente revisado en algunos de los aspectos esenciales que acabamos de comentar. Aun así, todavía quedan muchas áreas por explorar. Mientras que los fundamentos teóricos de este modelo se han discutido extensamente, faltan trabajos que exploren su aplicación en la práctica docente. Uno de los aspectos que queda por explorar es precisamente el que abordaremos en este trabajo: la relación del enfoque léxico con la lingüística de corpus, a la que dedicaremos el siguiente apartado.

2.2. Los corpus lingüísticos

Los corpus se suelen definir como colecciones representativas de textos, delimitadas en función de unos criterios explícitos. Pueden ser creados por instituciones académicas, y también por personas o grupos de personas interesadas. En otras palabras, cualquiera puede crear su propio corpus, aunque este término suele aplicarse a proyectos desarrollados por entidades académicas o empresariales. Existen corpus escritos y orales, y aquí nos referiremos principalmente a los escritos.

Los corpus varían según su tamaño, el tipo de metadatos textuales y extratextuales que poseen, y según sus ámbitos de aplicación, entre otros parámetros. Ahora nos detendremos brevemente en ellos.

En cuanto al primer parámetro, el tamaño, existen corpus grandes y pequeños, en función de la parcela de los datos lingüísticos que se quiera abarcar y su finalidad. Para que un corpus se pueda considerar como una muestra representativa del lenguaje, cuyo análisis arroje resultados válidos, debe tener un tamaño mínimo. Este tamaño mínimo es muy difícil de fijar, puesto que en él influyen muchas variables y porque cualquier texto o colección de textos (desde un artículo hasta una antología o incluso el conjunto de todo lo que está publicado en internet) puede ser convertido en un corpus. Por ejemplo, Lehrberger y Bourbeau (1988: 141) opinan que un corpus de un tema muy acotado debe tener como mínimo 10 000 palabras.

La mayoría de los corpus son más extensos. Por ejemplo, los corpus académicos del español –CREA, CORDE y CORPES XXI– tienen 154 000 000 (Real Academia Española [RAE], 2008), 250 000 000 (RAE, 2014) y 397 594 000 (RAE, 2021b) palabras, respectivamente. Los corpus web, integrados por textos publicados en línea, tienen un tamaño muy superior: por ejemplo, el corpus web esTenTen18 (Kilgarriff y Renau, 2013) ofrece casi 17 mil millones de palabras. Además, si adoptamos la idea, que no carece de validez, de que todos los textos accesibles a través de un buscador como Google se pueden considerar como un corpus, este tendría un tamaño imposible de determinar.

Las finalidades de los corpus, el segundo criterio establecido, pueden ser tan variadas como su extensión. En general, se pueden distinguir dos factores que las determinan: la naturaleza de los textos que se incluyen y el uso que se hace del corpus. Por un lado, el criterio seguido en la selección de los textos determina cuáles serán las posibilidades de uso del corpus (en función del tipo de textos, un corpus puede ser general o especializado, diacrónico o sincrónico, etc.). Por otro lado, la función prevista del corpus determina qué textos serán incluidos durante su creación.

Al hablar sobre el tercer parámetro que define la naturaleza de un corpus, la anotación, nos referimos a los metadatos lingüísticos (normalmente de tipo gramatical) y extralingüísticos con los que está enriquecido. Corpus no anotados son meras recopilaciones de textos, lo que limita significativamente sus aplicaciones: al no contar con lematizadores, solo permiten búsquedas de formas de palabras específicas y análisis básicos de frecuencia. Son compatibles con el formato KWIC (del inglés *Keyword in Context*), que consiste en mostrar la palabra buscada en el centro de la línea de texto que la contiene y permite observar y analizar su cotexto. Por tanto, los corpus no anotados pueden servir para muchos propósitos, pero su utilidad es limitada porque exigen un esfuerzo adicional por parte del usuario a la hora de interpretar los datos.

La anotación extratextual de un corpus aporta metadatos sobre la localización del texto (en qué libro o página web aparece), el momento en el que fue creado, el nombre de su autor, su género textual, etc. La anotación textual incluye la lematización (qué formas puede tener el lema), el etiquetado morfosintáctico (la categoría sintáctica o *POS* y otros rasgos morfosintácticos) y el análisis sintáctico (qué función cumple la palabra dentro de la oración). Obviamente, los corpus anotados permiten búsquedas mucho más sofisticadas y cómodas. Si volvemos a fijarnos en los corpus académicos, CORDE no contiene metadatos textuales pero sí extratextuales. CREA tiene una versión no anotada y otra, anotada con el sistema de codificación desarrollado para CORPES. El sucesor del CREA, CORPES XXI, sí contiene un lematizador y etiquetas morfosintácticas.

2.2.1. Los corpus en la enseñanza de lenguas

Desde el auge de la lingüística de corpus en los años noventa surgieron voces que promovían el uso de esta herramienta en el aula de LE. Se suelen diferenciar dos maneras de usar los corpus en clases de idiomas: la directa y la indirecta (Pérez Serrano, 2017: 77). La forma indirecta consiste en usar los datos de corpus en la fase preparatoria de la enseñanza, para extraer ejemplos auténticos de uso y para recabar datos estadísticos. Desde la aparición de los enfoques por tareas y proyectos se valora positivamente la presencia de textos auténticos en la enseñanza, por lo que los corpus se convierten en una herramienta indispensable. A través de KWIC, el docente puede extraer una lista de ejemplos para su posterior análisis por parte de los estudiantes. A partir de los metadatos extratextuales el docente puede recuperar textos completos que contengan el elemento lingüístico deseado y ofrecerlos como material de lectura a los alumnos. Puesto que las editoriales recurren a los corpus para la creación de manuales de

LE, el profesor termina usando los datos de corpus incluso si nunca ha estado en contacto directo con ellos. El corpus es importante también como fuente de referencia para 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 de los corpus, además de aportar muestras de textos reales (Buyse, 2017), facilita un aprendizaje autónomo e inductivo (Ferrando, 2017). Con el uso de los corpus cambia también el papel del profesor y, por extensión, del manual: estos dejan de ser la única fuente de conocimiento en el aula porque los corpus proporcionan grandes cantidades de información (Pérez Serrano, 2017: 79). Además, conviene señalar que la naturaleza panhispánica de la mayoría de los corpus del español –que contienen muestras del lenguaje muy variadas desde el punto de vista diatópico– favorece el aprendizaje intracultural.

La forma directa de usar los corpus en la enseñanza de LE ha sido menos desarrollada desde el punto de vista teórico y práctico. El uso directo implica el manejo de los corpus por parte de los propios aprendientes tanto en el aula como fuera de él. Pérez Serrano (2017: 77) distingue entre dos subcategorías dentro de este uso: el *uso directo mediado* y el *uso directo estricto*. En el caso del uso mediado, los estudiantes utilizan el corpus bajo la supervisión del profesor para resolver dudas lingüísticas relacionadas con las actividades planificadas. El uso estricto se da cuando los alumnos realizan consultas en el corpus de manera independiente para resolver sus dudas, sin control por parte del profesor.

2.2.2. Uso directo de corpus en el aula de LE

Ahora nos referiremos brevemente a los estudios que examinaron la validez del uso directo de los corpus en el aula. En su estudio del 2010, Boulton ha sometido a 62 estudiantes a una prueba de uso de los corpus a través de la consulta de concordancias. En la tarea, los estudiantes tenían que contestar a preguntas relacionadas con cinco expresiones lingüísticas apoyándose únicamente en unas concordancias impresas. En la siguiente tarea había otras cinco expresiones similares, pero la fuente de consulta fueron los diccionarios. Los resultados indicaron que los alumnos aprendían al menos de manera tan efectiva con las concordancias como con el diccionario. Se sacaron dos conclusiones significativas. La primera es que, tras una breve explicación por parte del profesor sobre el funcionamiento de los corpus (de 5-10 minutos), los estudiantes eran capaces de analizar los datos por sí mismos y sacar conclusiones válidas. La segunda conclusión, más general, es que el uso de los corpus para adquirir el léxico es una estrategia válida.

En un estudio posterior, Boulton (2012) comparó el uso de datos extraídos de los corpus y presentados a los alumnos en papel (como se había hecho en Boulton 2010) con el uso de los corpus en línea por los propios estudiantes. El marco temporal de este segundo estudio ha sido más extenso y la tipología de ejercicios más variada: a lo largo de diez semanas, se realizaron diversos ejercicios léxicos y gramaticales, impresos y en línea. Los resultados indicaron que resultan más eficaces los datos de corpus presentados en papel (por ejemplo, las listas de concordancias imprimidas), pero los alumnos prefieren el trabajo en línea. Se comprobó también que, aunque el grupo en general no tuvo problemas para manejar el corpus, había diferencias significativas entre alumnos, lo que implica que el aprendizaje a través de los corpus no sería igualmente apropiado para todos los aprendices. Se trata de una conclusión importante y que sigue la línea de los enfoques modernos de la enseñanza, que ponen énfasis en las diferencias en los estilos de aprendizaje de diferentes estudiantes. Teniendo esto en cuenta, se puede decir que introducir los corpus como una alternativa a los métodos de aprendizaje tradicionales es una opción válida, pero siempre hay que tener en cuenta el estilo de aprendizaje de cada alumno y no esperar que sea una herramienta igual de eficiente para todos.

En la sección anterior hemos constatado que el aprendizaje a través de los corpus facilita un aprendizaje autónomo e inductivo. En cuanto a la autonomía, se supone que, una vez que los estudiantes se familiaricen con el uso de los corpus, podrán realizar consultas para aclarar sus dudas. Después de todo, es algo que ya están haciendo de alguna forma (Cruz Piñol, 2012: 107): muchos o incluso la mayoría de los estudiantes hacen búsquedas de palabras y frases en Google y otros buscadores para verificar su corrección o indagar en su uso (volveremos sobre esta consideración en la sección 4.4., en relación con la encuesta sobre la adquisición léxica que hemos realizado a un grupo de estudiantes avanzados de ELE). Una búsqueda en un corpus no sería muy diferente en lo esencial y le aportaría mucha más información útil al aprendiz.

En cuanto al aprendizaje inductivo del léxico (en el que una unidad léxica se asimila por la exposición durante un tiempo relativamente prolongado), se puede afirmar que los corpus constituyen una de las pocas formas que existen para ponerlo en práctica. En el aprendizaje inductivo de la gramática, resulta fácil introducir un mismo fenómeno (por ejemplo, un tiempo verbal) repetidamente dentro de un limitado período de tiempo y mantener la variedad, pero volver varias veces sobre la misma unidad léxica para asegurar la retención puede resultar repetitivo y agobiante. Los corpus ayudan a mitigar este inconveniente.

2.2.3. Inconvenientes del uso directo de los corpus

No obstante, el manejo directo de los corpus (propios o ajenos) en el aula presenta ciertos retos. Uno de estos retos es la barrera tecnológica que sigue existiendo (Pérez Serrano, 2017), aunque ya no es tan infranqueable como hace unos años. Los autores que la mencionan se refieren generalmente a la falta de equipos informáticos en el aula: hace quince años el uso de los corpus en el aula estaba sujeto a la disponibilidad de ordenadores en el centro de enseñanza, pero hoy en día los estudiantes suelen venir a clase con sus propios dispositivos con conexión a internet. Aun así, sigue habiendo una barrera –aunque en una forma distinta– porque la interfaz de usuario que tienen muchos corpus no es fácil de manejar (en nuestra opinión, este sería el caso de los corpus académicos). Es cierto que cualquier corpus se debe introducir después de una explicación por parte del profesor sobre su naturaleza y sus funcionalidades básicas (Cabot, 2017), pero si, además, la interfaz de usuario no es intuitiva, la fase de explicación y entrenamiento se alarga.

Otro reto importante asociado al manejo de los corpus por parte de aprendientes de LE tiene que ver con su capacidad para interpretar los datos extraídos: "Puede que el alumno encuentre demasiados datos, o demasiado pocos o que le sea difícil diferenciar entre lo correcto e incorrecto" (Pérez Serrano, 2017: 80). Especialmente con usuarios-estudiantes novatos la supervisión del profesor en la fase de análisis de datos y formulación de conclusiones es esencial para asegurar que las conclusiones que saca el estudiante sean correctas y para que no se desmotive por la cantidad de datos que no sabe interpretar. Esto es especialmente relevante con los corpus avanzados (o herramientas de consulta avanzadas), que proporcionan una cantidad y variedad de datos enorme que puede desconcertar a usuarios principiantes. Por lo tanto, existe un consenso en que los corpus se deben introducir preferentemente con alumnos familiarizados con las nuevas tecnologías y en niveles avanzados de dominio de la lengua meta. En estos niveles los alumnos suelen tener una conciencia lingüística más desarrollada y pueden sacar más provecho a los metadatos textuales (por ejemplo, el etiquetado morfosintáctico). Además, son capaces de analizar en profundidad los datos encontrados en el corpus, por ejemplo, diferenciar entre expresiones cultas y coloquiales o analizar las restricciones selectivas de los predicados.

Para un uso efectivo de los corpus uno tiene que ser consciente de sus limitaciones, por ejemplo, su grado de representatividad como una muestra del lenguaje. Ningún corpus es ideal en este sentido. Los corpus grandes suelen ser los más representativos, pero el tamaño no es el único criterio importante. Los corpus académicos, por ejemplo, solo incluyen textos escritos

cultos o semicultos (principalmente de libros, periódicos y revistas). Al no incluir el registro coloquial, transmiten una imagen incompleta de la lengua. Los corpus académicos tampoco son muy equilibrados en cuanto a las variedades geográficas, porque incluyen un porcentaje mucho mayor de textos generados en España que en América Latina (aunque hay que decir que la Academia está intentando cambiarlo). Los corpus web (por ejemplo, esTenTen) suelen ser más equilibrados en este sentido, pero presentan otros problemas. 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 puede darse entre hablantes nativos (por ejemplo, en los foros o blogs, donde no se cuida la gramática o la selección léxica), pero que no es apropiado en la enseñanza de español como segunda lengua.

2.2.4. Propuestas de uso de los corpus en la enseñanza

Como hemos mencionado, el número de estudios que ponen en práctica el enfoque léxico es muy bajo. Son menos aún los trabajos que indagan en el desarrollo práctico del uso de los corpus, sobre todo si comparamos su número con el de los estudios dedicados a otros ámbitos de la enseñanza de segundas lenguas. Los trabajos que sí lo hacen se centran predominantemente en el inglés. Cobb y Boulton (2015) analizaron las referencias existentes y concluyeron que el 95% de los estudios están escritos en inglés y examinan el uso de los corpus del inglés. Solo encontraron cinco estudios en francés, que analizan corpus del francés, y dieron por hecho que se daría una correspondencia similar para otras lenguas. Hay muy pocos trabajos escritos en español sobre el uso de corpus y estos suelen ser de tipo teórico-descriptivo (como es el caso de Buyse, 2017 o Piñol, 2012). Solo un puñado de trabajos proponen desarrollos prácticos en forma de actividades concretas (por ejemplo, Higueras, 2009; Pérez Serrano, 2017; Cavanillas, 2017).

Conviene mencionar que la mayoría de los estudios sobre el uso de corpus se centran predominantemente en la explotación de las concordancias. Es verdad que las concordancias son una herramienta básica, fácil de usar y que se puede utilizar para muchos tipos de actividades. Sin embargo, los corpus actuales ofrecen muchas otras funcionalidades que se pueden explotar con fines didácticos, lo que es aún más evidente para sistemas avanzados de consulta de corpus (como Sketch Engine).

3. Consideraciones teóricas específicas

3.1. Los corpus propios en la enseñanza de lenguas extranjeras

Este trabajo constituye una propuesta de uso de los corpus en la enseñanza de lenguas. Más específicamente, se centra en el uso de *corpus propios*. En el contexto de este trabajo, los corpus propios son corpus recopilados por los propios alumnos de LE, quienes los usarán como parte de su aprendizaje. Los corpus propios pueden contener textos producidos por los propios alumnos o por otros autores. En este trabajo hablaremos del segundo grupo (corpus de textos de otros autores).

Los corpus propios de textos ajenos pueden documentar eficientemente el uso de la lengua en un contexto muy acotado, lo que presenta muchas ventajas frente a los corpus convencionales. Puesto que el temario de los cursos de idiomas extranjeros casi siempre se compone de temas acotados, 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. El propio proceso de creación del corpus puede ser gratificante para el alumno y positivo para su autoestima en la medida en que tiene que usar su propio criterio para seleccionar materiales adecuados.

Como hemos dicho en la sección 2.5., los trabajos que examinan la aplicación del enfoque léxico son pocos y aún son menos los estudios que tratan sobre la explotación de los corpus en LE. Con respecto a los trabajos que analizan el uso de los corpus propios, tenemos que decir que son casi inexistentes si nos limitamos al ámbito de la enseñanza de LE. Sí hay estudios –no muchos– que analizan el uso de corpus propios en otras áreas afines, 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, que abarca el aprendizaje de lenguas, pero con una finalidad diferente (traducción como tal). Los trabajos de Tribble y Wingate (2013) y Lee y Swales (2006) tratan sobre la escritura académica inglesa (*EAP: English for Academic Purposes*). Tribble y Wingate (2013) proponen un curso de escritura académica en inglés basado en la creación y consulta de corpus propios. Lee y Swales (2006) informan sobre los resultados de un experimento práctico centrado en la misma área. Este segundo estudio es más relevante para el presente trabajo que Tribble y Wingate (2013), puesto

que indagó sobre la enseñanza de escritura académica a estudiantes no nativos. Los participantes de este estudio hicieron un curso de inglés académico de nivel universitario, durante el que usaron corpus existentes y crearon también dos corpus propios –uno de artículos académicos de su ámbito y el otro de su propia escritura– que después compararon. En línea con la observación que hemos hecho antes sobre el efecto gratificante que este tipo de tareas tienen sobre los alumnos, los participantes del estudio valoraron positivamente la práctica descrita: señalaron que los dotó de confianza y que la prefieren a un trabajo basado en la consulta de manuales o gramáticas.

En conclusión, aunque el uso de corpus propios en el contexto específico de la enseñanza de LE no se ha tratado hasta donde sepamos, se pueden extrapolar a este ámbito las siguientes conclusiones formuladas en trabajos sobre disciplinas afines, como la traducción o la escritura académica:

1. La creación de corpus propios es una actividad de la que los estudiantes disfrutan independientemente del tema que se trate en clase.
2. El uso de corpus propios es más rentable cuando el aprendiz trabaja sobre temas acotados (por ejemplo, los estudiantes del estudio de Lee y Swales, 2006 crearon corpus de escritura académica médica).
3. La posibilidad de contrastar escritura propia con la ajena que ofrecen los corpus propios no resulta desmotivadora para el aprendiz. Es más, este tipo de análisis comparativo tiene un efecto enriquecedor para su competencia léxica y gramatical.

3.2. Aprendizaje del léxico en niveles avanzados de competencia

Este trabajo se centra específicamente en los estudiantes con un nivel avanzado de segunda lengua (a partir de ahora nos referiremos a ellos como *estudiantes avanzados*). En este último bloque teórico definiremos las características de este grupo discente y su manera de aprender el léxico en comparación con los estudiantes de nivel inicial o intermedio.

Aplicaremos este término para referirnos, de forma general, a los estudiantes cuyo nivel de español según el *Marco común europeo de referencia (MCER)* es C1 y C2. La adaptación española del *Marco*, el *Plan curricular del Instituto Cervantes*, denomina este grupo como "etapa avanzada-superior o de uso competente de la lengua" (Instituto Cervantes, 2006). Algunos de los equivalentes de la denominación *estudiante avanzado* son *high-level proficiency learner* (Bardel, 2016: 75) o *mastery learner* (Capel, 2012: 12). Bardel (2016) distingue tres subcategorías dentro del amplio concepto de *estudiante avanzado* en orden descendente de

dominio de la LE: hablantes de nivel nativo (*nativelike*), hablantes casi nativos (*near-native*) y hablantes avanzados (*advanced*). Un hablante de nivel nativo usa la LE como si fuese un nativo, sin que los fallos de su expresión se puedan detectar en un meticuloso escrutinio. Un hablante casi nativo se expresa como un nativo en situaciones comunicativas cotidianas, pero sí se pueden detectar algunas desviaciones con respecto a los nativos en un examen cuidadoso. Un hablante avanzado se expresa de manera parecida a los nativos, pero los rasgos no nativos son perceptibles en toda su expresión escrita u oral. En este trabajo se considerará todo el grupo de *estudiantes avanzados*, con sus tres subcategorías.

Para fijar lo que constituye un nivel avanzado de competencia léxica, se hace referencia al número de unidades léxicas que conoce el alumno y también a algunos factores cualitativos: la *profundidad léxica* (noción que explicaremos más adelante), el nivel de especificidad del léxico usado y la autonomía de aprendizaje. Empezaremos por el factor cuantitativo o el número de palabras conocidas. En la sección 2.1. de este trabajo, dedicada al enfoque léxico, se ha comentado que un estudiante de nivel C1 domina, según el MCER, unas 10 000 unidades léxicas. Vamos a matizar esta estimación. A la hora de cuantificar el vocabulario de un estudiante, conviene hablar no solo de las *unidades léxicas* sino también de *familias léxicas*. Como hemos dicho antes, el cerebro del hablante almacena el léxico en el lexicón mental. Palabras relacionadas morfológicamente –derivadas de una misma base morfológica– constituyen una familia léxica. Por ejemplo, *proponer, propuesta* y *proposición* son tres unidades léxicas diferentes integradas dentro de la misma familia léxica. Si el número que estipula el MCER (10 000) se aplicara a las familias léxicas, estaríamos hablando de todas las palabras derivadas a partir de 10 000 bases morfológicas, a las que habría que añadir las formas de estas palabras si incluyéramos la morfología flexiva.

Con respecto al tamaño del léxico, hay que señalar asimismo que resulta difícil de cuantificar, tanto para las lenguas maternas como para las extranjeras. Para una lengua materna, Pustejovsky y Batiukova (2019) lo cifran en unas 250 000 *entradas* léxicas para el inglés, que abarcan el vocabulario activo y pasivo. La dificultad para determinar el tamaño del vocabulario de un hablante deriva, al menos en parte, de la falta de un método fiable y sencillo para poder medirlo: se necesita una amplia gama de pruebas (Bardel, 2016). A esto habría que añadir que se trata de un nivel del lenguaje con la mucha variación, condicionado por factores extralingüísticos (la situación comunicativa), socioculturales (pertenencia de los hablantes a diferentes grupos sociales según su nivel económico y educativo), psicológicos (las restricciones impuestas por la memoria), etc. (Bardel, 2016: 76). De ahí la diferencia de resultados incluso en estudios que aplican métodos similares. En el estudio de Goulden *et al.*

(1990), veinte hablantes nativos graduados realizaron una prueba de conocimientos léxicos (marcaron las voces conocidas en una lista de 500 unidades léxicas). El resultado fue que un hablante nativo educado conoce hasta 20 000 familias léxicas. En el estudio de Milton (2009), donde se usó la misma prueba que en Goulden *et al.* (1990), se habla de un promedio de 9000 familias.

Naturalmente, el léxico de hablantes no nativos incluye un número más bajo de familias léxicas, porque incluso en los niveles avanzados normalmente no se llega al nivel de un nativo. En su descripción del perfil léxico que deben poseer los estudiantes que se examinan para que se les certifique el nivel C2 de inglés (el más alto de los seis niveles que hay, véase Capel, 2012), la Universidad de Cambridge estipula como número meta 7000 familias léxicas. Al mismo tiempo, se reconoce que la competencia léxica de algunos de los estudiantes es superior y se acerca a la de los hablantes nativos. En cambio, Nation (2006) en su estudio del vocabulario inglés de estudiantes universitarios de grado fija el número entre 8000 y 9000 familias. Como se ve, dar con una cifra definitiva es imposible: un estudiante avanzado de nivel (cuasi)nativo conocería entre 7000 y 20 000 familias léxicas. Independientemente de cuál sea el número exacto, podemos decir que el nivel avanzado es el que se define por una mayor variedad y tamaño del léxico, y por su cercanía al dominio nativo. 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, así como por factores socioculturales (que incluyen los intereses académicos y profesionales).

Uno de los factores cualitativos que se invocan para definir el concepto de estudiante avanzado es la *profundidad léxica*, que usaremos como equivalente del término inglés *lexical depth* (Bardel, 2016: 82). La profundidad léxica se puede relacionar con el complejo conjunto de informaciones que integra una unidad léxica: su forma (sonido o grafía), su comportamiento sintáctico, su significado denotativo y connotativo, y las circunstancias asociadas a su uso (registro, frecuencia, etc.). Tanto el tamaño del léxico como su profundidad se incrementan en niveles avanzados de competencia: la exposición a un idioma y su aprendizaje redundan en que, por un lado, el léxico sea más rico y variado y, por el otro lado, en que se profundice en el conocimiento de cada unidad léxica. Un conocimiento más amplio y profundo permite, a su vez, mejorar la conciencia lingüística (capacidad de pensar sobre la lengua) y desarrollar el criterio lingüístico, que distinguen a los estudiantes avanzados.

En el plano de la producción, los aprendientes avanzados seleccionan las palabras (y expresiones complejas lexicalizadas) con gran *precisión léxica*. Los estudiantes saben diferenciar entre sinónimos porque conocen los matices de significado que los distinguen, no

incurren en la vaguedad y ambigüedad, son conscientes de la existencia de la polisemia, etc. Otro factor clave es la *adecuación léxica*, que se consigue cuando se elige la voz más apropiada distrática y diafásicamente (Capel, 2012: 12). En ocasiones, un estudiante avanzado puede incluso llegar a superar a los hablantes nativos en cuanto a la competencia léxica, si se pone suficiente énfasis en su enseñanza. En su estudio, Bogaards (2000) comparó en una prueba a los estudiantes avanzados y los hablantes nativos, y los primeros superaron a los segundos en tareas de relación semántica (sinonimia, hiperonimia, etc.).

Una vez trazado el perfil formativo de los estudiantes avanzados, nos detendremos ahora en la propia enseñanza de idiomas en esos niveles. En general, se puede decir que el aprendizaje en niveles avanzados es un proceso más autónomo, aleatorio y especializado que en los niveles inferiores. Una mayor autonomía se debe a dos causas fundamentales. En primer lugar, el número de alumnos en cursos de idiomas de nivel avanzado suele ser relativamente bajo, y la instrucción se desarrolla muchas veces en el marco de clases particulares (para prepararse para un examen, por ejemplo) o de manera autónoma. Los niveles avanzados se tardan en alcanzar: se llega a ellos normalmente al final de las etapas educativas obligatorias o incluso posobligatorias, es decir, cuando ya ha terminado la fase más intensa de aprendizaje de LE. Además, pese a numerosos cambios introducidos en las últimas décadas para abandonar el aprendizaje cuantitativo, la estructura de la enseñanza de idiomas sigue conservando rasgos piramidales: cuanto más alto sea el nivel, menos posibilidades de alcanzarlo se ofrecen y a un menor número de estudiantes. En segundo lugar, el aprendizaje avanzado es más autónomo en cualquier área de conocimiento. Si nos ceñimos a la enseñanza de LE, al tener un criterio lingüístico más desarrollado, los estudiantes avanzados recurren más a la deducción (por ejemplo, para procesar las unidades léxicas en el contexto –véase Capel (2012)– y son capaces de procesar contenidos lingüísticos que en los niveles inferiores pasan desapercibidos.

El segundo factor que hemos mencionado es la naturaleza aleatoria o improvisada del aprendizaje avanzado, que está relacionada con la autonomía del aprendizaje. La aleatoriedad consiste en que el aprendizaje se produce en contextos no relacionados con la enseñanza de idiomas. Un ejemplo claro es la educación universitaria, en la que muchos estudiantes reciben instrucción en una lengua extranjera, sea cual sea su formato: además de asignaturas obligatorias de idiomas extranjeros, es cada vez más común cursar asignaturas no lingüísticas en una lengua extranjera. Fuera del ámbito universitario, los aprendientes avanzados también suelen consumir materiales en la LE (principalmente a través de internet). A menudo, por poseer ya un nivel avanzado, estos alumnos ni siquiera distinguen entre contenidos en su lengua materna y en la LE (textos, vídeos, películas, series, etc.). El idioma extranjero deja de ser una

finalidad en sí misma y pasa a ser un vehículo, por lo que el aprendizaje se convierte en un efecto colateral y suele ser de tipo inductivo (recuérdese lo que hemos comentado en la sección 2.1. sobre la teoría del aprendizaje inductivo dentro del enfoque léxico de Lewis). La aleatoriedad tiene que ver con la autonomía: cuando no se consigue procesar contenidos léxicos nuevos a partir de conocimientos adquiridos previamente (por deducción), se recurre a la búsqueda autónoma en un diccionario, un corpus u otras fuentes en soporte digital o físico.

El tercer y último factor presente en el aprendizaje en niveles avanzados es su carácter especializado. La instrucción en estos niveles se suele impartir casi siempre en ciclos educativos posobligatorios para grupos reducidos de alumnos. A menudo se desarrolla dentro de cursos de LE como lengua de especialidad, relacionados con la actividad laboral o académica de los alumnos, en los que el léxico especializado y la terminología juegan un papel importante (Capel, 2012). En muchos casos, este idioma es el inglés, qué es la lengua vehicular de la enseñanza universitaria y el mundo científico-técnico. En menor medida, también cumple esta función el español. El léxico que se aprende en estos contextos es el especializado o técnico, y los conocimientos de los estudiantes superan los que pueda tener un hablante nativo no especialista (Bardel, 2016).

En conclusión, un estudiante avanzado de LE puede acercarse a los hablantes nativos en cuanto a sus conocimientos léxicos o incluso igualarlos en este sentido. Si se considera la profundidad léxica y el contexto de las lenguas de especialidad, los aprendientes de LE a menudo superan a los hablantes nativos. Pero hay áreas, desatendidas en la instrucción formal, en las que los alumnos avanzados no llegan al nivel nativo. Por ejemplo, el estudio de Bogaards (2000) reveló que los estudiantes de LE que en algunos casos superan a sus pares nativos se quedan atrás en lo que se refiere al léxico coloquial, expresiones con carga cultural y el lenguaje formulaico (expresiones fijas, colocaciones). En nuestra propuesta didáctica expuesta en la sección 5.4., veremos precisamente ejemplos de actividades con corpus cuyo objetivo es hacer conscientes a los alumnos de la importancia del lenguaje formulaico y ayudarlos a practicarlo.

4. Encuesta sobre la adquisición del léxico

En esta sección presentamos la encuesta sobre la adquisición del léxico español por parte de aprendientes de nivel avanzado que se ha diseñado específicamente para este estudio, sus objetivos, el perfil de los alumnos que han participado y los resultados de la encuesta.

4.1. Objetivos

Esta encuesta tiene dos objetivos fundamentales. El primero es confirmar las conclusiones generales hechas en estudios precedentes sobre sobre el estilo de aprendizaje de los estudiantes avanzados, y sobre el uso de los corpus y otras fuentes de consulta en el proceso de aprendizaje (cfr. secciones 2 y 3). El segundo consiste en recaban información adicional para la elaboración de la propuesta didáctica presentada en la sección 5.

4.2. Estructura y contenido

El cuestionario consta de 15 preguntas divididas en cuatro bloques. Las preguntas se presentan aquí en forma resumida y el cuestionario completo se incluye como Anexo 1. En el bloque 1 (preguntas 1-5) se recopilan los datos personales de los alumnos: el sexo, la edad, la nacionalidad, la lengua materna y el dominio de otras lenguas extranjeras. El bloque 2 (preguntas 6-8) se centra en cuestiones relativas a la instrucción en ELE recibida por los encuestados (su tipo y duración). El bloque 3 (pregunta 9) está articulado alrededor de las estrategias empleadas por el aprendiente cuando se encuentra con léxico desconocido. El cuarto y último bloque (preguntas 10-15) se centra en las herramientas de consulta del léxico disponibles a los aprendices de ELE y el uso que hacen de ellas. Las preguntas hacen referencia a cuatro categorías generales de recursos: los diccionarios monolingües (pregunta 10), los diccionarios bilingües (pregunta 11), los traductores basados en corpus paralelos (pregunta 12) y los corpus (pregunta 13). Para cada grupo de recursos, los alumnos tenían que indicar si lo usan o no y, si la respuesta es afirmativa, indicar cómo lo usan. En la pregunta 15 se valora la utilidad de cada herramienta en una escala de 1 (poco útil) a 5 (muy útil). Se ha incluido aparte esta pregunta para poder abarcar los casos en los que el alumno no usa una herramienta concreta, pero es consciente de su utilidad, o sí la usa, pero no la valora muy bien. La pregunta 14 es abierta: da la posibilidad de mencionar recursos léxicos no incluidos en los puntos 10-13 y 15.

4.3. Perfil de los encuestados

En la encuesta han participado 13 estudiantes de español como LE. Ciertamente, el tamaño de muestra es pequeño e insuficiente para poder sacar conclusiones estadísticamente significativas, pero sí permite detectar las tendencias más robustas. El principal criterio de

selección de participantes ha sido su nivel de español, puesto que este estudio está centrado principalmente en estudiantes avanzados. El nivel de español no se ha determinado en función del tiempo de aprendizaje o de una certificación oficial. Sí se ha tenido en cuenta el nivel de las clases de español que han cursado y si tenían formación sólida relacionada con la lengua española (como estudios de grado o máster en lingüística hispánica o filología española). El grupo resultante es bastante heterogéneo en cuanto a la nacionalidad (británicos, chinos, coreanos, etc.) y la edad. Resumimos sus características en la Tabla 1. Además, los encuestados provienen de distintos contextos de aprendizaje (estudiantes de cursos intensivos de la academia Inhispania, alumnos extranjeros en la UAM, estudiantes de español en otras universidades europeas, etc.).

1. Sexo	**2. Edad**	**3. Tipo de instrucción en ELE recibida**	**4. Forma de aprendizaje en el presente**	**5. Duración de aprendizaje (años)**
54% Mujer	77% 21-25	46% – Antes de la universidad	54% – Asistencia a clases de español	8% 0-1
46% Hombre	23% Más de 25	53% – En la universidad	61% – Aprendizaje autónomo	8% 2
		46% – En una escuela de idiomas	15% – No estoy aprendiendo	61.5% 3-5
		31% – De manera autónoma		23% Más de 8

Tabla 1. Perfil de los encuestados

4.4. Análisis de los resultados

En esta sección se analizarán aquellos de los datos obtenidos que son relevantes a efectos de este trabajo y se comentarán las implicaciones que tienen para la propuesta didáctica. Los resultados se recopilan en cuatro tablas (Tablas 2, 3, 4 y 5).

La tabla 2 muestra las respuestas a la pregunta 9 del cuestionario (*¿Qué es lo primero que haces cuando te encuentras con una palabra española que no conoces?*). La tabla 3 presenta el porcentaje de alumnos que usan diferentes fuentes de consulta. La tabla 4 registra la puntuación media (de 1 a 5) asignada a cada grupo de fuentes de consulta por los encuestados en función de su utilidad. La tabla 5 recoge las respuestas a la única pregunta abierta del cuestionario, en la que los encuestados podían incluir otras fuentes de consulta del léxico que usan; para cada fuente, se incluye el número de veces que se cita.

Estrategia de adquisición del léxico	Porcentaje de alumnos que la aplican
Intento deducir el significado	69,2%
Busco la traducción a mi LM	23,3%
Busco la definición en español	7,7%

Tabla 2. Estrategias de adquisición de léxico desconocido

Categoría de fuentes de consulta	Porcentaje de alumnos que la usan
Diccionarios monolingües	61,5%
Diccionarios bilingües	100%
Traductores basados en corpus paralelos	61,5%
Corpus	7,7%

Tabla 3. Uso de distintas fuentes de consulta

Categoría de fuentes de consulta	Puntuación media asignada por los encuestados
Diccionarios monolingües (DLE)	4,10
WordReference	3,25
Linguee y similares	3,70
Google translate	3,15

Tabla 4. Valoración de la utilidad de diferentes fuentes de consulta

Otros recursos léxicos usados	Número de menciones
SpanishDict	3
Google	3
Diccionario Naver	1

Tabla 5. Otros recursos utilizados

La tabla 2 muestra una prevalencia bastante marcada de la opción 3 (*intento deducir el significado*), con un 69% de respuestas positivas. Se trata de una característica esencial de la adquisición del léxico en niveles avanzados, que tiene una relación directa con la autonomía de aprendizaje. La autonomía se manifiesta en la capacidad de reflexión sobre unidades léxicas desconocidas, y su desarrollo depende de dos factores fundamentales: el tiempo de aprendizaje (cuanto más prolongado sea, más maduro es el criterio lingüístico) y la importancia que se otorga a una reflexión independiente durante la enseñanza. La segunda opción más frecuente (*busco traducción a mi LM*), elegida por tres de los encuestados, implica una autonomía mucho

menor. Aunque sería necesaria una muestra mucho más amplia para determinar qué rasgos del perfil de los estudiantes subyacen en esta elección, podemos relacionar estos resultados con los dos factores que acabamos de citar. Dos de los tres aprendientes que eligieron esta opción llevaban poco tiempo estudiando el español (1-2 años). Aunque han llegado a un nivel avanzado gracias a los cursos intensivos, es posible que su criterio lingüístico y, con él, su capacidad de deducción, no se haya desarrollado suficientemente. La tercera estudiante es de origen chino y su respuesta se puede explicar por el hecho de que, en su país, la metodología de la enseñanza de LE no ha experimentado un cambio tan drástico como en otras partes del mundo y sigue conservando muchas de las características de la enseñanza tradicional. La preferencia por una adquisición basada en la deducción por parte de la mayoría de los alumnos avanzados justifica el uso de corpus con este grupo de alumnos en la medida en que implica que saben determinar qué significado expresan las unidades léxicas en el contexto.

Según los resultados recogidos en la Tabla 3, los diccionarios bilingües son la fuente de consulta usada por todos los encuestados. Esto contradice los postulados metodológicos del enfoque léxico, según los cuales se debe disuadir a los aprendices de depender demasiado de los diccionarios bilingües (especialmente en niveles avanzados), e insistir en las ventajas de otras fuentes de consulta, especialmente los diccionarios monolingües, porque estos contribuyen al desarrollo del criterio lingüístico y al aprendizaje incidental del léxico durante la consulta. La popularidad de los diccionarios bilingües se debe, en nuestra opinión, a dos factores: su facilidad de uso y el hábito creado en etapas anteriores de aprendizaje. En cuanto al primer factor, el uso de diccionarios bilingües no requiere conocimientos lingüísticos ni técnicos: a través de la voz consultada se accede directamente a su equivalente en otra lengua. El segundo factor hace referencia al traslado de estrategias asimiladas por el estudiante en el nivel principiante e intermedio a su aprendizaje cuando llega al nivel avanzado. En efecto, los diccionarios bilingües son una de las herramientas más usadas en los niveles inferiores porque se basan en un modelo muy simplificado del léxico, que consiste en una lista de palabras que tienen sus equivalentes en otras lenguas. Esto crea un hábito procedimental que persiste en los niveles avanzados, aunque entonces la visión del nivel léxico que tienen los aprendientes ya es mucho más multifacética. Las otras herramientas de consulta tienen una frecuencia de uso mucho menor: un 61,5% de los encuestados usan diccionarios monolingües y traductores basados en corpus paralelos, y un 7,7% usan corpus (aunque, como veremos más adelante, el porcentaje real para los corpus paralelos sería un poco más alto).

A primera vista, el poco uso que hacen de los corpus lingüísticos los estudiantes encuestados dificulta la justificación de una propuesta como la que hacemos en este trabajo,

pero antes de sacar conclusiones conviene que nos detengamos en la valoración de la utilidad de diferentes fuentes de consulta, cuyos resultados están resumidos en la Tabla 4. Como se puede ver, los diccionarios bilingües, la fuente más usada, obtienen la valoración más baja (3,15 en el caso de Google Translate y 3,25 para WordReference). Los superan, por unas décimas, los traductores basados en corpus paralelos (con un 3,7) y, con un margen más amplio, los diccionarios monolingües (con un 4,10). Se puede concluir que, aunque los alumnos usan frecuentemente los diccionarios bilingües, son conscientes de sus limitaciones y también de las ventajas que presentan las otras herramientas.

Los corpus paralelos proporcionan información muy parecida a la de los diccionarios bilingües en cuanto a la traducción, y además aportan ejemplos y el contexto de uso, y la posibilidad de búsqueda de unidades léxicas complejas. Los ejemplos de uso funcionan como elementos descodificadores y codificadores a la vez: por un lado, ayudan a descodificar el significado de una voz y captar sus diferentes matices, y por el otro proporcionan información adicional necesaria para el uso activo de la palabra en el contexto. Otra ventaja de los corpus paralelos consiste en que permiten buscar unidades léxicas complejas y combinaciones estables de palabras, y por tanto reflejan mejor los aspectos formulaicos del lenguaje. Los diccionarios bilingües han ido mejorando en este sentido y ahora incluyen algunas expresiones con altos niveles de fijación e idiomaticidad, pero se quedan cortos a la hora de reflejar combinaciones de palabras que son estables pero composicionalmente transparentes, como las colocaciones.

Por último, analizamos los datos de la tabla 5 (otros recursos utilizados). Hay dos fuentes de consulta que tienen más de una mención: SpanishDict y Google reciben tres menciones cada uno. El diccionario SpanishDict es un diccionario basado en varios corpus paralelos que se parece a Linguee y Reverso en cuanto a sus funciones y uso. A diferencia de estos, solo incluye corpus paralelos español-inglés, pero ofrece también otras funciones (como unidades didácticas, que no se comentarán aquí por su poca relevancia para este trabajo). Su principal ventaja frente a Linguee es la variedad y el tamaño de los corpus paralelos en los que se basa. Linguee toma datos predominantemente del corpus paralelo Europarl y se nutre, además, de otros sitios web cuya calidad de contenidos no siempre es la deseable: el proceso automático de equiparación de textos no se verifica por expertos humanos. SpanishDict, además de Europarl, incluye el corpus Paracrawl, cofinanciado por la UE. Paracrawl es más grande que Europarl: para el español, Europarl incluye 54 806 927 palabras (Koehn, 2005) y Paracrawl 4 374 060 920 palabras (Broader/Continued Web-Scale Provision of Parallel Corpora for European Languages, 2019). SpanishDict es, entonces, una versión más fiable de Linguee en lo que se refiere a la pareja español-inglés. El hecho de que algunos de los encuestados hayan

respondido negativamente a la pregunta sobre el uso de corpus paralelos, pero después han mencionado SpanishDict en la pregunta 14 implica que los traductores basados en corpus paralelos son una herramienta útil para ellos, aunque no sean conscientes de están usando realmente los corpus paralelos.

El segundo recurso que ha sido mencionado con frecuencia en el cuestionario es el buscador de Google. Es un dato significativo por dos razones. Por un lado, confirma que, como se ha comentado en la sección 2.2., los estudiantes usan los buscadores para comprobar cómo se escriben y se usan las palabras. En este sentido, Google funciona de forma parecida a los diccionarios monolingües (suele mostrar datos en la lengua extranjera, por lo que se aprende de forma inductiva) pero, a diferencia de estos, permite buscar unidades léxicas complejas. Por otro lado, como se ha argumentado aquí, Google se puede equiparar a un sistema de consulta de corpus, si se entiende por corpus todo lo publicado en internet. El tipo de búsqueda es similar al que se puede hacer en un corpus, aunque las opciones para definir y acotar los parámetros de búsqueda son limitadas.

4.5. Conclusiones de la encuesta

En resumen, los datos de la encuesta confirman que los corpus ya han entrado en la enseñanza de LE, aunque de forma indirecta. El uso de los corpus lingüísticos, entendidos en el sentido estricto, es casi inexistente, por lo que se puede concluir que los resultados de los trabajos que examinaron el uso de concordancias y otras funciones de los corpus en la enseñanza de LE no han sido transferidos a la práctica. La utilización indirecta de los corpus abarca el uso de los traductores basados en corpus paralelos y el uso de Google como herramienta de consulta. Los corpus paralelos gozan de menos uso que los diccionarios bilingües, pero los alumnos valoran positivamente las informaciones adicionales que estos aportan: los ejemplos y contextos de uso y la posibilidad de buscar unidades léxicas complejas. Google se usa como la herramienta de consulta del corpus más grande del mundo. En lo que se refiere a la adquisición del significado de palabras desconocidas, hemos comprobado que los estudiantes avanzados suelen ser capaces de deducirlo a partir del contexto (o al menos lo intentan). Por tanto, aunque casi no existe un uso directo de los corpus por parte del grupo de aprendientes que nos interesa, los datos obtenidos confirman que tiene mucho potencial: se da un aprendizaje autónomo y deductivo, y se hace uso de herramientas que proporcionan contexto y ejemplos.

5. Propuesta didáctica de uso de corpus propios

Basándonos en las características de la adquisición del léxico por parte de aprendientes avanzados de LE perfiladas en las secciones 2, 3 y 4 de este trabajo, hemos desarrollado una propuesta didáctica de uso de corpus propios en la enseñanza de ELE. En todas las propuestas se usará el sistema de consulta y gestión de corpus Sketch Engine, que presentaremos a continuación. Proponemos las actividades con corpus propios para estudiantes avanzados porque, como hemos visto en 3.2., la naturaleza específica de los corpus propios se corresponde con las características del aprendizaje avanzado (sea este un aprendizaje de temas específicos dentro del ámbito de la lengua general o un aprendizaje del español con fines específicos).

5.1. Descripción general de Sketch Engine

Sketch Engine es un sistema de consulta y gestión de corpus, creado en 2003 por la compañía Lexical Computing. Se creó en colaboración con el equipo de *Natural Language Processing Centre* de la Universidad de Masaryk de Brno. Hoy en día, la empresa proporciona *software* a algunas de las instituciones más prestigiosas del ámbito lexicográfico y lexicológico, por ejemplo, las editoriales Collins, Marriam-Webster y Oxford University Press. Sketch Engine permite consultar corpus monolingües (como el famoso British National Corpus 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 o en archivos de texto subidos por el usuario. Se pueden compilar corpus en más de 140 lenguas, pero el tipo de anotación automática textual que se ofrece (a la que hemos eludido en el apartado 2.2.1.) 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 *et al*., 2004: 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*[1], que resumen el comportamiento contextual de la palabra dada: identifican las unidades léxicas que la acompañan en diferentes posiciones sintácticas (Kilgarriff *et al.*, 2004: 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 muy común que etiqueten como objeto directo un sujeto pospuesto; también se pueden confundir categorías gramaticales de palabras homógrafas. El usuario debería ser capaz de identificar estos fallos cuando analiza los datos. Es uno de los motivos por los que proponemos este tipo de actividades para alumnos avanzados.

Antes de usar cualquier herramienta tecnológica hay que introducirla debidamente. Es el caso de los corpus digitales en general: en los estudios sobre el uso de los corpus que hemos citado en la sección 2.2.2. (Boulton 2010, 2012), se hacía una introducción breve (de 5-10 minutos) sobre su manejo. Sketch Engine es un sistema más complejo, por lo que en su caso la presentación previa debería ser más larga, pero no necesariamente exhaustiva. Consideramos que 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 Difference y N-grams. Aunque resulta difícil fijar un número mínimo 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.

En el siguiente subapartado (§5.2.) introduciremos brevemente las herramientas de Sketch Engine de las que haremos uso (Concordance, Word Sketch, Word Sketch Difference y N-grams), y en la sección 5.4. presentaremos las actividades agrupadas según la herramienta usada en cada caso.

5.2. Herramientas de Sketch Engine usadas en la propuesta

Las concordancias (*concordances*) son la función más básica de cualquier herramienta de consulta de corpus (ya la hemos mencionado en la sección 2.2.). Las concordancias son

[1] Por motivos de claridad referencial, usaremos la denominación inglesa de las principales herramientas de Sketch Engine, puesto que este sistema solo tiene interfaz en inglés.

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 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 presentamos en la sección 5.3.

Figura 1. Fragmento de concordancias para el lema *violencia* (Lexical Computing, 2003)

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*.

Además de listados de palabras, en Word Sketch se pueden ver representaciones gráficas, parecidas a mapas conceptuales, en las que diferentes funciones sintácticas están incluidas en zonas marcadas con colores diferentes y donde la frecuencia de coaparición con la palabra buscada está reflejada a través de la distancia con respecto a esta palabra. Este tipo de visualizaciones son muy útiles desde el punto de vista didáctico, porque facilitan una percepción más intuitiva de las restricciones de coaparición de diferentes palabras.

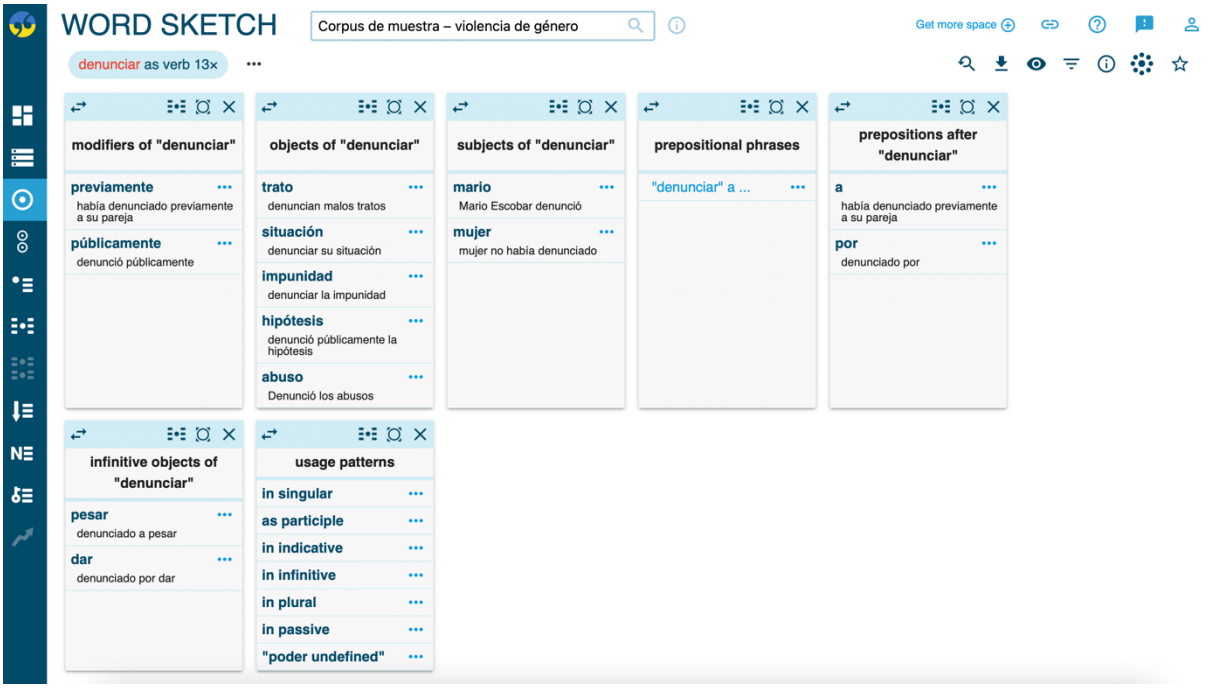


Figura 2. Resumen de Word Sketch para el lema *denunciar* (Lexical Computing, 2003)

La herramienta *Word Sketch Difference* 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 Difference para los sustantivos *cuerpo* y *cadáver*. La combinatoria de *cuerpo* está marcada en verde, la de *cadáver* en rojo, y en las franjas intermedias aparen los elementos contextuales que estas palabras tienen en común.

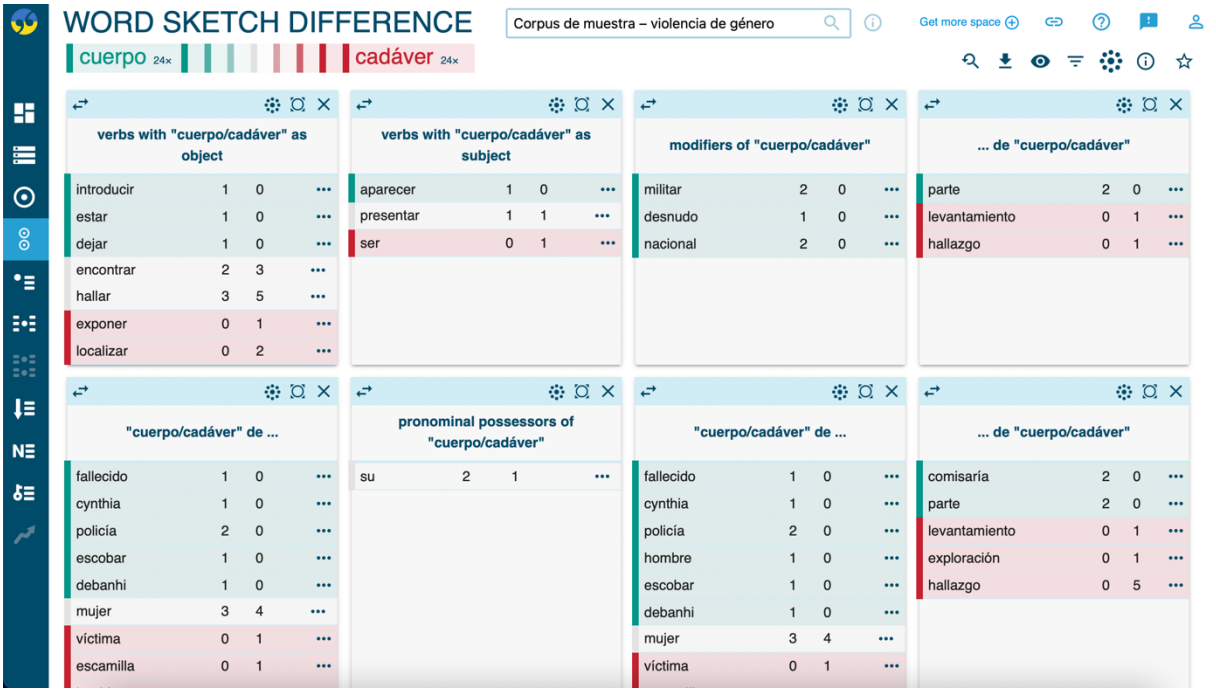
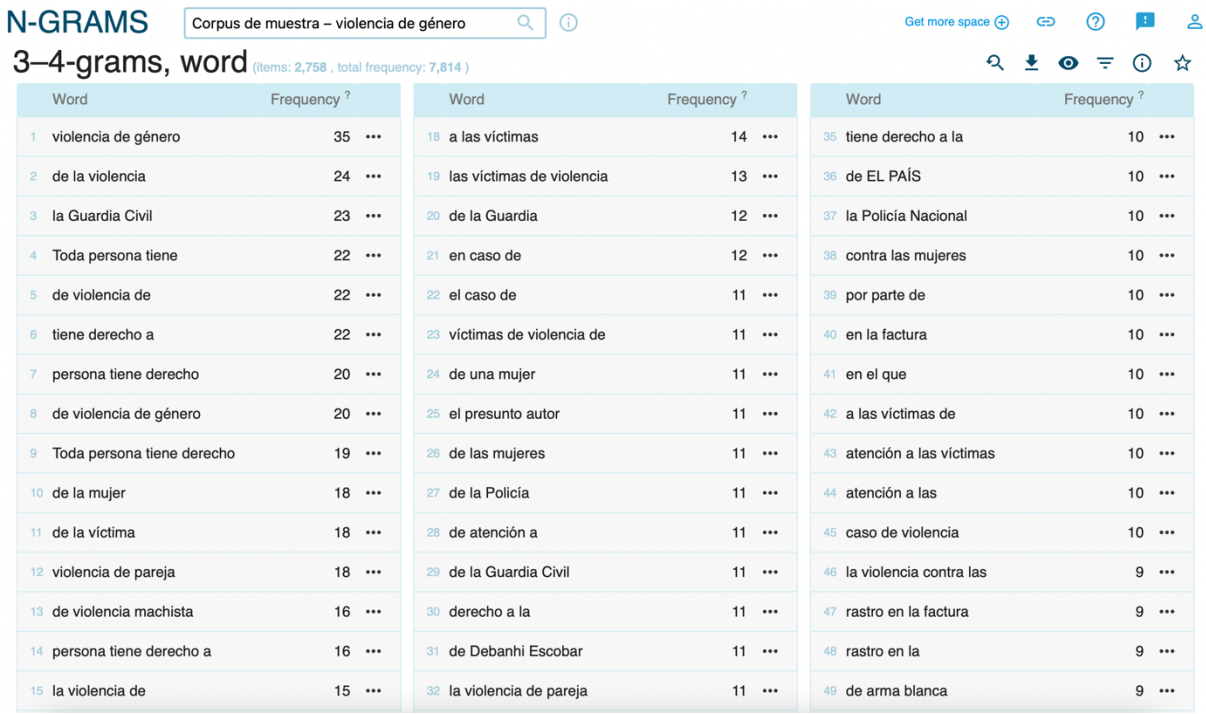


Figura 3. Comparación de la combinatoria de *cuerpo* y *cadáver* a través de Word Sketch Difference (Lexical Computing, 2003)

Por último, la herramienta N-grams 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 que se introducirá en el siguiente punto (§5.3.).


N-GRAMS Corpus de muestra – violencia de género

Get more space

3–4-grams, word (items: 2,758 , total frequency: 7,814)

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

Figura 4. Expresiones estables de 3 y 4 palabras detectadas por N-grams (Lexical Computing, 2003)

Todas estas herramientas proporcionan la frecuencia de aparición (absoluta y/o relativa) de los resultados y otros datos estadísticos, relativos, por ejemplo, a la fuerza de la relación contextual que existe entre dos palabras (*LogDice*).

5.3. Corpus de muestra

Para ilustrar la aplicación de corpus propios en la enseñanza de ELE (en la sección 5.4.), 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 de nivel avanzado: la violencia de género. En la creación del corpus, 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 25 000 palabras.

La creación de un corpus propio con textos sobre el tema tratado (como el corpus de muestra que brevemente hemos descrito aquí) 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 un corpus común 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.

El hecho de que la búsqueda en un corpus se realice por la forma gráfica de la palabra tiene repercusiones importantes en el resultado de la búsqueda, sobre todo cuando la palabra buscada es polisémica. Por ejemplo, en un corpus general del español europeo el verbo *abusar* aparecería con las dos principales acepciones que registra el DLE (RAE, 2021a): 'hacer un uso excesivo, injusto o indebido de algo o de alguien' –como en *abusar del poder*– y 'hacer objeto de trato deshonesto a una persona de menor experiencia, fuerza o poder' –como en *abusar sexualmente*–. Al enfrentarse a casos como este, el aprendiente va perfeccionando uno de los aspectos importantes de su competencia léxica, la *profundidad léxica*, que es uno de los indicadores del dominio de un idioma extranjero (recuérdese la sección 3.2.). En un corpus temáticamente acotado, en cambio, la polisemia estaría resuelta en la mayoría de los casos, lo que tiene otro tipo de ventajas: se facilita la comprensión de palabras desconocidas, se acelera la asimilación de su uso y mejora la *precisión léxica* (por la adquisición de significados específicos para un contexto concreto). La segunda acepción de *abusar* que hemos citado ('hacer objeto de trato deshonesto a una persona de menor experiencia, fuerza o poder') es la que se da de manera predominante en nuestro corpus de muestra.

5.4. 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 de la organización del léxico. Por este motivo, y para no saturar a los alumnos con información sobre el manejo de los corpus, conviene distribuir su uso entre temas o unidades didácticas diferentes. Para afianzar el aprendizaje de clase, conviene integrar este tipo de taras en el trabajo autónomo de los alumnos (por ejemplo, los deberes de casa) y realizar su seguimiento individual.

5.4.1. Concordancias

Las concordancias son una herramienta muy fácil de usar, que permite detectar los fragmentos textuales en los que se usa la palabra o expresión buscada. Basta con que el profesor las introduzca brevemente: explique en qué consisten y cómo se pueden definir los criterios de búsqueda en el corpus.

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. Para entender el significado de palabras desconocidas, el alumno debe aplicar la deducción y examinar cuidadosamente el contexto. Sabemos de los datos de la encuesta que los alumnos tienden a usar esta estrategia en niveles avanzados por sí mismos, con herramientas como el buscador de Google. Si se manejan correctamente, las concordancias proporcionan datos más fiables y de forma más rápida. Por ejemplo, si un alumno quiere ver con qué preposiciones se combina un verbo de régimen preposicional (p.ej., *depender*), en Google tiene introducir *depender* (o algo como "*depender* + 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, podríamos 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). De esta forma el usuario recupera todos usos preposicionales sin necesidad de revisar líneas de concordancias irrelevantes.

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. 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 el segundo paso. El volumen del *input* que se obtiene en las concordancias es 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 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 formato 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 en el corpus. 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.

5.4.2. Word Sketch

Como se ha explicado en la sección 5.2., la herramienta Word Sketch compila listados las palabras que coocurren 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 que tengan cierta frecuencia en el corpus propio y que pertenezcan a una de las principales categorías gramaticales léxicas (sustantivos, verbos y adjetivos), 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 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*.

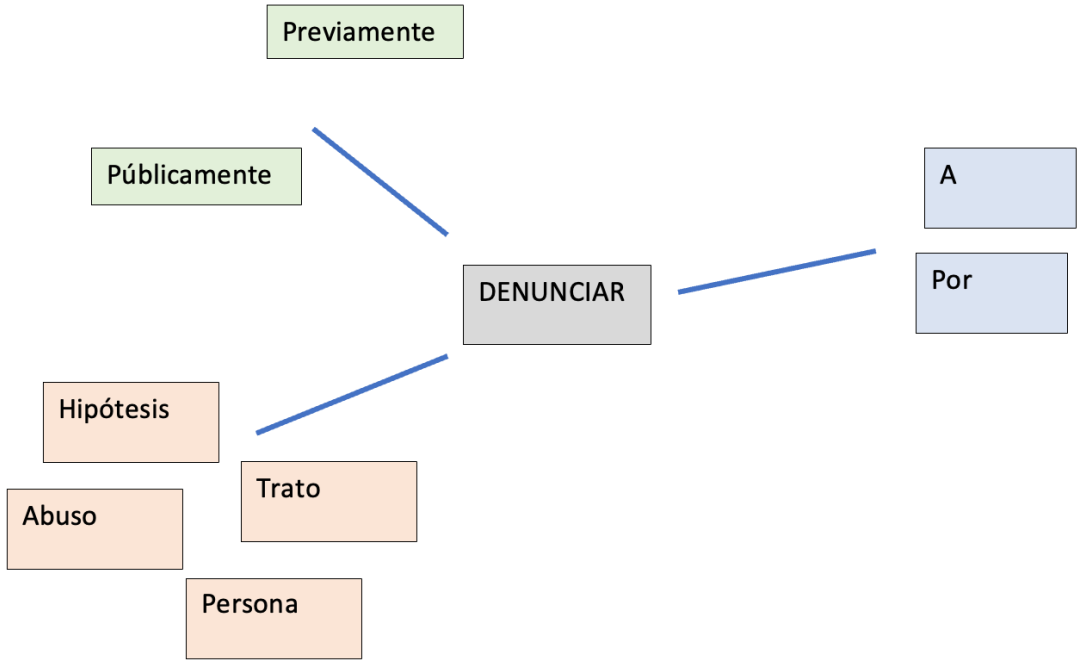


Figura 5. Mapa conceptual inicial basado en el comportamiento contextual del verbo *denunciar*

Siguiendo el formato de las visualizaciones gráficas de Word Sketch, que hemos mencionado en la sección 5.2., 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 beis, 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.

En el siguiente paso, 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 6 ilustramos cómo se puede hacer con *abuso* (objeto directo de *denunciar* del primer mapa conceptual). Además, si el corpus no es muy grande, el mapa se puede enriquecer con unidades léxicas que no se encuentran allí. Así se fomentaría en los alumnos la capacidad de buscar ejemplos propios.

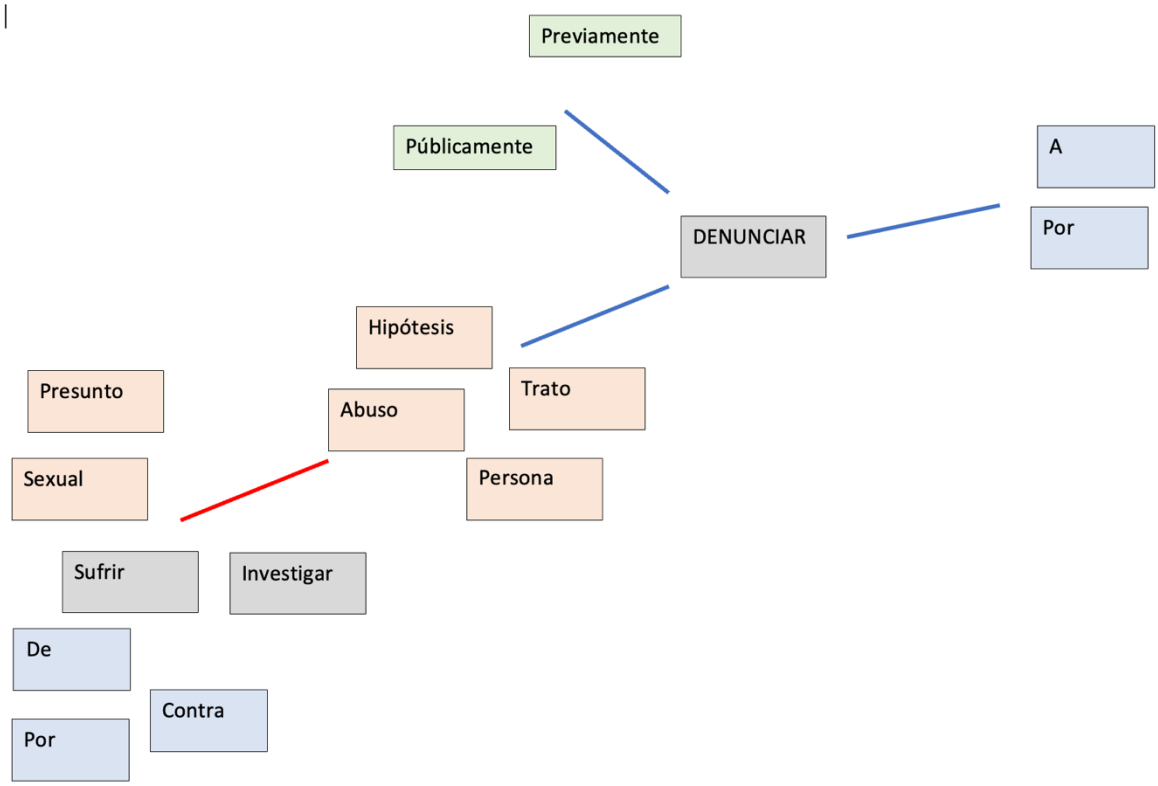


Figura 5. Mapa conceptual desarrollado basado en el comportamiento contextual del verbo *denunciar*

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.). Como vimos en punto 2.1., uno de los aspectos controvertidos del enfoque léxico original fue la disociación no justificada entre el léxico y 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 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 aparee 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 detenid*o al estado que es el resultado de dicho acto.

5.4.3. Word Sketch Difference

La función de la herramienta Word Sketch Difference es contrastar el comportamiento gramatical de dos palabras. Como vimos en el punto 3.2., 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 comodines por terminología específica, ejercicios de diferenciación semántica de términos cognados y de cuasisiononímos. 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 Difference. Para detectar (cuasi)sinónimos, el profesor puede hacer uso de la herramienta de frecuencia (Wordlist), que hemos introducido en el punto 5.4.2. 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 Difference y analizar las diferencias que existen en cada par. Los nombres *crimen* y *delito*, por ejemplo, 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 grande (p.ej., CORPES XXI o esTenTen18).

Además de la comprensión del léxico, en actividades de este tipo se trabaja la distinción, fundamental dentro del enfoque léxico, entre los parámetros de *selección libre* e *idiomaticidad* (véase el punto 2.1.). Desde la perspectiva de selección libre, no existe ninguna razón por la que no se pueda usar la combinación *delito machista*: se trata de una combinación semánticamente válida y gramaticalmente correcta. No obstante, no se da en el uso real; es cuando interviene el principio de la idiomaticidad. Tareas de este tipo ayudan a los alumnos a entender la naturaleza idiomática de la lengua e interiorizar las expresiones estables o institucionalizadas. Es un aspecto importante de la adquisición de LE en niveles avanzados: un estudiante avanzado suele expresarse correctamente desde el punto de vista gramatical, pero lo

que dice a veces suena raro porque usa combinaciones de palabras que un hablante nativo no usaría.

5.4.4. N-grams

La herramienta N-grams detecta expresiones estables de palabras de diferente nivel de fijación y las ordena según su frecuencia de aparición. Su uso facilita el aprendizaje de léxico formulaico. Puesto que N-grams muestra la frecuencia absoluta de cada expresión detectada, el estudiante puede distinguir entre combinación de voces frecuentemente usadas por los hablantes nativos por un lado y los hápax o usos ocasiones (por ejemplo, invenciones de hablantes específicos) por otro lado. La búsqueda en N-grams 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 grande (como esTenTen18) devolvería muchísimos resultados, únicamente clasificados por frecuencia. Por ello su utilidad didáctica sería limitada. Con un corpus general, 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 nuestro corpus de muestra) 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 en la sección 5.4.2., en las actividades 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 podría énfasis en las expresiones pluriverbales. En la primera actividad se hace una búsqueda no restringida de expresiones con N-grams en un 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 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-grams, 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-grams 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.

6. Conclusiones

Los corpus son una herramienta de potencial indudable para la lingüística y el uso en el proceso de enseñanza de lenguas extranjeras. No obstante, aunque se han dedicado ya algunos estudios a su explotación didáctica, su aplicación práctica queda, hasta el momento, escasa o inexistente. En nuestro trabajo, hemos intentado explorar un área menos conocida y estudiada de uso de los corpus, en cuanto a su temática y herramientas de uso. Por lo tanto, nos hemos propuesto crear un trabajo que investigue el uso de los corpus propios. Es más, nuestro trabajo examina su explotación a través de varias funciones de la herramienta Sketch Engine que constituyen diferentes maneras de enseñar y aprender el léxico, en línea con la teoría del enfoque léxico. Como hemos argumentado, hay un potencial en la explotación de este tipo de trabajo gracias a su acotamiento temático y a las posibilidades diversas de Sketch Engine. Los corpus propios permiten examinar el uso de unidades léxicas en un contexto específico y ayudan a cultivar la precisión y profundidad léxicas que son necesarias para el dominio de una lengua extranjera. La diversa naturaleza de herramientas que ofrece la plataforma Sketch Engine, por su parte, permite una explotación diversa e interesante de los corpus a través de observación de ejemplos de uso, examinación de relaciones gramaticales entre las unidades léxicas y las expresiones pluriverbales que caracterizan la expresión formulaica humana.

Bibliografía

Álvarez Cavanillas, J. L. (2017). Un enfoque léxico en los manuales ELE. En Francisco Herrera (Ed.), *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 57-69.

Bardel, C. (2016). The lexicon of advanced L2 learners. En Kenneth Hyltenstam (Ed.) *Advanced proficiency and exceptional ability in second languages.* New York: De Gruyter Mounton, 73-109.

Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: the case of the Euralex French Tests. *Applied Linguistics*, 21(4), 490–516.

Bogaards, P. (2001). Lexical Units and the Learning of Foreign Language Vocabulary. *Studies in Second Language Acquisition*, 23(3), 321-343.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534-572.

Boulton, A. (2012). Hands-on / hands-off: Alternative approaches to data-driven learning. En Thomas, J.,y Boulton, A. (Eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press. 2012, 153-169.

Broader/Continued Web-Scale Provision of Parallel Corpora for European Languages. (2019). *ParaCrawl Corpus release v9*. https://www.paracrawl.eu [consulta 9/04/2022]

Buyse, K. (2017). Los corpus como herramientas de aprendizaje de léxico. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras.* Barcelona: Difusión, 121-140.

Cabot, M. (2017). Aprender léxico, una cuestión de estilo y mucha estrategia. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 83-93.

Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, 3(1), 1-14.

Cobb, T., y Boulton, A. (2015). Classroom applications of corpus analysis. En Biber, D., y Reppen, R. (Eds.) Cambridge Handbook of Corpus Linguistics. Cambridge: Cambridge University Press, 478-497.

Chamorro, D. (2017). Dar forma al enfoque léxico: directrices para clase y tipología de actividades. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 71-81.

Cruz Piñol, M. (2012). *Lingüística de corpus y enseñanza del español como 2-L*. Madrid: Editorial Arco Libros.

Curiosity Media. (2016). Spanish Learning Made Easy. *Spanishdict.com*, https://www.spanishdict.com [consulta 10/06/2022]

DeepL SE. (2009). Linguee. *Linguee.com, https://www.linguee.com* [consulta 8/06/2022]

Ferrando, V. (2017). El papel de las colocaciones en la enseñanza y el aprendizaje del español. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 95-104.

Goulden, R., Nation, P., y Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341–363.

Herrera, F. (Ed.). (2017). *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión.

Higueras, M. (2009). Aprender y enseñar léxico, *MarcoELE*, 9, 111-26.

Instituto Cervantes. (2006). Plan curricular del Instituto Cervantes. *Cervantes.es*, https://www.cervantes.es/lengua_y_ensenanza/aprender_espanol/plan_curricular_instituto_cervantes.htm. [consulta 15/05/2022]

Kilgarriff, A., y Renau, I. (2013). esTenTen, a Vast Web Corpus of Peninsular and American Spanish. *Procedia - Social and Behavioral Sciences*, 95, 12-19.

Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D. (2004). The Sketch Engine. En *Proceedings of the Eleventh EURALEX International Congress.* 105-15.

Koehn, P. (2005). Europarl: a parallel corpus for statistical machine translation. *Proc MT summit 5*, 79–86.

Lee, D., y 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.

Lehrberger, J., y Bourbeau, L. (1988). *Machine Translation: Linguistic characteristics of MT systems and general methodology of evaluation.* Amsterdam: John Benjamins Publishing.

Lexical Computing Ltd. (2003). *Sketch Engine*. https://www.sketchengine.eu [consulta 20/04/2022]

Marín Peris, E. (2017). Textos y palabras en un aprendizaje de ELE orientado a la acción. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 25-33.

Milton, J. (2009). Measuring second language vocabulary acquisition. Clevedon: Multilingual Matters.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63(1), 59–82.

Nation, P. (2001). *Teaching & Learning Vocabulary*. Cambridge: Cambridge University Press.

Pérez Serrano, M. (2017). *La enseñanza-aprendizaje del vocabulario en ELE desde los enfoques léxicos*. Madrid: Editorial Arco.Libros-La Muralla.

Pustejovsky, J., y Batiukova, O. (2019). *The Lexicon*. Cambridge: Cambridge University Press.

Real Academia Española. (2008). CREA escrito. *Rae.es*, https://www.rae.es/banco-de-datos/crea/crea-escrito [consulta 05/05/2002]

Real Academia Española. (2021a). Diccionario de la lengua española, 23.ª ed. [versión 23.5 en línea]. *Rae.es.* https://dle.rae.es/ [consulta 7/06/2022]

Real Academia Española (2021b). CORPES XXI [versión 0.94 en línea]. *Rae.es, https://www.rae.es/banco-de-datos/corpes-xxi* [consulta 5/05/2022]

Real Academia Española (2014). Corpus diacrónico del español [en línea]. *Rae.es*, https://www.rae.es/banco-de-datos/corde [consulta 5/05/2022]

Reverso SAS. (1998). Reverso traducción. Reverso.net. https://www.reverso.net/traducción-texto [consulta 10/06/2022]

Ruffat, A., y Jiménez Calderón, F. (2017). Aplicaciones de enfoques léxicos a la enseñanza comunicativa. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 47-55.

Serradilla Castaño, A. (2014). La fraseología en el aula de ELE: nuevos enfoques y propuestas didácticas. En Jacinto Gonzáles Cobas et. al (Eds.) *¿Qué necesitamos en el aula de ELE?: reflexiones en torno a la teoría y la práctica.* redELE, 73-98.

Tarrés, I. (2017). Categorización y combinatoria: algunas preguntas sobre el funcionamiento del sistema léxico. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 35-45.

Tribble, C., y Wingate, U. (2013). From text to corpus: A genre-based approach to academic literacy instruction. *ScienceDirect*, 41, 307-21.

Troitiño, S. (2017). Implicaciones de crear materiales desde una perspectiva léxica. En Francisco Herrera (Ed.) *Enseñar léxico en el aula de español: el poder de las palabras*. Barcelona: Difusión, 143-161.

Wray, A. (2002). *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

Zhao, Y., y Shi, J. (2015). Self-compiled On-line Parallel Corpus in Translation Teaching. *Conference on Education and Teaching in Colleges and Universities*, Atlantis Press, 68-71.

Anexo 1

Cuestionario sobre la adquisición del léxico en ELE

PARTE 1

1. Tú eres:

☐ Hombre
☐ Mujer
☐ Prefiero no decirlo

2. Tu edad es:

☐ Menos de 18
☐ 18 – 20
☐ 21-25
☐ Más de 25

3. ¿Cuál es tu nacionalidad?

_____

4. ¿Cuál es tu lengua materna?

_____

5. ¿Qué lenguas extranjeras dominas, además del español?

_____

PARTE 2

6. ¿Cómo has aprendido el español? Puedes elegir más de una opción.

☐ En el colegio y el instituto (antes de la universidad)
☐ En la universidad
☐ En una escuela de idiomas
☐ De manera autónoma

7. ¿Cómo estás estudiando el español ahora?

☐ Doy clases en español.
☐ Estoy estudiando de manera autónoma (con algún apoyo, como Duolingo o un manual).
☐ No estoy estudiando el español ahora.

8. ¿Cuántos años llevas estudiando el español?

- [ ] 0 – 1
- [ ] 2
- [ ] 3 – 5
- [ ] 6 – 7
- [ ] Más de 8

PARTE 3

9. **¿Qué es lo primero que haces cuando te encuentras con una palabra española que no conoces?**

- [ ] Busco la traducción a mi lengua materna
- [ ] Busco la definición en español
- [ ] Intento deducir su significado a partir del contexto

PARTE 4

¿Cuáles de las siguientes herramientas usas y para qué?

10. **Diccionarios monolingües (por ejemplo, el Diccionario de la lengua española en dle.rae.es)**

- [ ] Uso este recurso
- [ ] No uso este recurso

Si tu respuesta anterior ha sido afirmativa, por favor específica para qué lo usas (puedes elegir más de una opción)

- [ ] Para buscar el significado de palabras que no conozco
- [ ] Cuando conozco el significado de la palabra, pero no sé cómo se usa (por ejemplo, para un verbo, si se usa con o sin preposición)
- [ ] Para ver ejemplos de uso

11. Diccionarios bilingües (por ejemplo, a través de WordReference o Google translate)

- [ ] Uso este recurso
- [ ] No uso este recurso

Si tu respuesta anterior ha sido afirmativa, por favor específica para qué lo usas (puedes elegir más de una opción)

- [ ] Para buscar el significado de palabras que no conozco
- [ ] Cuando conozco el significado de la palabra, pero no sé cómo se usa (por ejemplo, para un verbo, si se usa con o sin preposición)
- [ ] Para ver ejemplos de uso
- [ ] Para buscar traducciones

12. Traductores de corpus paralelos (por ejemplo, Linguee)

- [ ] Uso este recurso
- [ ] No uso este recurso

Si tu respuesta anterior ha sido afirmativa, por favor especifica para qué lo usas (puedes elegir más de una opción)

- [ ] Para buscar el significado de palabras que no conozco
- [ ] Cuando conozco el significado de la palabra, pero no sé cómo se usa (por ejemplo, para un verbo, si se usa con o sin preposición)
- [ ] Para ver ejemplos de uso
- [ ] Para buscar traducciones

13. Corpus (p. ej. Corpes XXI, CREA)

- [ ] Uso este recurso
- [ ] No uso este recurso

Si tu respuesta anterior ha sido afirmativa, por favor especifica para qué lo usas (puedes elegir más de una opción)

☐ Para buscar el significado de palabras que no conozco

☐ Cuando conozco el significado de la palabra, pero no sé cómo se usa (por ejemplo, para un verbo, si se usa con o sin preposición)

☐ Para ver ejemplos de uso

☐ Para buscar traducciones

14. ¿Qué otros recursos, aparte de los mencionados, usas cuando tienes dudas sobre el significado y el uso de palabras españolas?

_____

15. **Marca en la escala la utilidad de los siguientes recursos: 1 (poco útil) – 5 (muy útil) – 0 (no se aplica)**

	1	2	3	4	5	0
Diccionario de la Lengua Española (RAE)	☐	☐	☐	☐	☐	☐
WordReference	☐	☐	☐	☐	☐	☐
Linguee	☐	☐	☐	☐	☐	☐
Google translate	☐	☐	☐	☐	☐	☐
Otro recurso (de la pregunta anterior)						
_____	☐	☐	☐	☐	☐	☐