

Investigar a partir de corpus de clases de espa1ol L2 con *Sketch Engine* y de manera manual: ejemplos de estudios sobre el l3xico

Nausica MARCOS MIGUEL
Universidad de Gante (B3lgica)
nausica.marcosmiguel@ugent.be
<https://orcid.org/0000-0003-0351-8174>

Resumen: Este art3culo muestra c3mo se pueden examinar los datos compilados en los Corpus de Clases de L2 (CCL2). En concreto, se centra en c3mo analizar, tanto manualmente como con la herramienta *Sketch Engine* (<https://www.sketchengine.eu>), la ense1anza de vocabulario en espa1ol a partir de este tipo de corpus. Para situar al lector, en la revisi3n de la literatura se presentan varios CCL1 y CCL2 disponibles a trav3s de alguna interfaz espec3fica o a trav3s de *Sketch Engine*. Adem3s, se describe c3mo investigadores con distintos enfoques han utilizado los CCL2 y con qu3 prop3sitos. Asimismo, puesto que los ejemplos que se dan en el art3culo se centran en investigaciones relacionadas con el l3xico en la L2, se incluye una secci3n muy somera sobre esta cuesti3n. A continuaci3n, al examinar las aplicaciones de *Sketch Engine*, se centra la atenci3n en las funcionalidades *LISTA DE PALABRAS*, *CONCORDANCIA* y *PALABRAS CLAVE*. Para ejemplificar la utilidad de *Sketch Engine* combinado con los an3lisis de tipo manual, se describen tres CCL2 creados por la autora del art3culo y se revisa la metodolog3a de cuatro investigaciones sobre la ense1anza del vocabulario y de la morfolog3a l3xica en las que se usaron estos tres CCL2. Tras describir c3mo se analizaron los datos en estos estudios, se proponen nuevas investigaciones posibles con *Sketch Engine* y se ofrecen algunas recomendaciones que pueden resultar 3tiles para quienes se adentren en el uso de esta herramienta aplicada al an3lisis de CCL2.

Palabras clave: corpus de aula; discurso del profesor; vocabulario; morfolog3a; ense1anza de L2

Catal3:

Investigar a partir de corpus de classes d'espanyol L2 amb *Sketch Engine* i de manera manual: exemples d'estudis sobre el l3xic

Resum: Aquest article mostra com es poden examinar les dades compilades en els Corpus de Clases de L2 (CCL2). En concret, se centra en com analitzar, tant manualment com amb l'eina *Sketch Engine* (<https://www.sketchengine.eu>), l'ensenyament de vocabulari en espanyol amb aquesta mena de corpus. Per a situar al lector, en la revisi3n de la literatura es presenten diversos CCL1 i CCL2 disponibles a trav3s d'alguna interf3cie espec3fica o a trav3s de *Sketch Engine*. A m3s, es descriu com investigadors amb diferents enfocaments han utilitzat els CCL2 i amb quins prop3sits. A m3s, com que els exemples que es veuen en l'article se centren en recerques relacionades amb el l3xic en L2, s'inclou una secci3n molt succinta sobre aquesta qüestió. A continuaci3n, en examinar les aplicacions de *Sketch Engine*, se centra l'atenci3n en les funcionalitats *LLISTA DE PARAULES*, *CONCORDANÇA* i *PARAULES CLAU*. Per a exemplificar la utilitat de *Sketch Engine* combinat amb les an3lisis de tipus manual, es descriuen tres CCL2 creats per l'autora de l'article i es revisa la metodolog3a de quatre recerques sobre l'ensenyament del vocabulari i de la morfologia l3xica en les quals es van usar aquests tres CCL2. Despr3s de descriure com es van analitzar les dades en aquests estudis, es proposen noves recerques possibles amb *Sketch Engine* i s'ofereixen algunes recomanacions que poden resultar 3tils per als qui s'endinsin en l'ús d'aquesta eina aplicada a l'an3lisis de CCL2.

Paraules clau: corpus d'aula; discurs del professor; vocabulari; morfologia; ensenyament de L2



English:**Analyzing Spanish L2 corpus with Sketch Engine and through manual searches: examples from vocabulary research**

Abstract: This article discusses how to examine L2 classroom corpora. Specifically, it focuses on how to analyze, manually and with Sketch Engine (<https://www.sketchengine.eu>), vocabulary instruction in these kinds of corpora. The literature review introduces available L1 and L2 classroom corpora that can be studied using their own online interface or Sketch Engine. To situate these corpora, the two main lines of research that use these kinds of corpora are mentioned and described. The literature review also briefly summarizes current research on instruction of vocabulary including derivational morphology as part of vocabulary instruction. The focus is on three Sketch Engine tools: Wordlist, Concordance, and Keywords. The article shows how to use these tools for analysis as well as how to complement these tools with manual analysis by examining four previous studies and the corpora created by the authors. Besides, the study explores future lines of research. This article aims to be helpful for new researchers who want to explore teaching and learning in classroom corpora and especially teaching and learning of L2 vocabulary.

Keywords: classroom corpora; teacher talk; vocabularies; morphology; L2 teaching

Introducción¹

Para investigar sobre la enseñanza y aprendizaje de segundas lenguas (L2)² hace falta examinar los múltiples contextos de aprendizaje-enseñanza que existen e, indiscutiblemente, un contexto fundamental de aprendizaje formal es el aula de L2. Los Corpus de Clases de L2 (CCL2) nos permiten examinar las prácticas de enseñanza, el *input* del aula y las interacciones entre alumnos y profesores.

Aunque los CCL2 también recogen muestras de estudiantes, no debemos confundirlos con corpus de aprendices que recogen muestras tanto escritas (p. ej. CEDEL2, <http://cedel2.learnercorpora.com/>, o COWS-L2H, <https://github.com/ucdaviscl/cowsl2h>) como orales (p. ej. SPLOCC, <http://www.splloc.soton.ac.uk/>, o CORELE, <http://cartago.llff.uam.es/corele/index.html>). Con estos corpus de aprendices se busca investigar el desarrollo lingüístico de los aprendices en un género específico (p. ej., entrevistas, narraciones, etc.). Aunque se puede investigar el desarrollo de la lengua de los estudiantes en el contexto de un CCL2, este no suele ser ni el objetivo principal ni el único objetivo a la hora de desarrollar un CCL2.

Comparado con los corpus de aprendices, encontramos pocos CCL2 disponibles públicamente. De ahí que los investigadores se vean en la necesidad de crear nuevos corpus. Este artículo mostrará cómo la herramienta *Sketch Engine* (<https://www.sketchengine.eu>) se puede utilizar para explorar el léxico en CCL2 pequeños. No obstante, dado que para utilizar esta herramienta primero es necesario crear un corpus, también revisaremos cómo se puede crear.

Sketch Engine es una plataforma disponible en línea que recoge varias herramientas de corpus tales como *LISTA DE PALABRAS*, *CONCORDANCIA* o *PALABRAS CLAVE*. Aparte de estas herramientas, *Sketch Engine* dispone de una variedad de corpus en múltiples lenguas que pueden servir como corpus de referencia. Uno especialmente relevante para español

¹ Querría agradecerles a los profesores y a sus alumnos por su colaboración. Sin su ayuda estos corpus no habrían sido posibles. También querría agradecer al equipo de CRELLT por el acceso a CEAPP, y a *Sketch Engine* por el permiso para usar las imágenes. Por último, me gustaría darles las gracias a los evaluadores y editores, por sus acertados comentarios en el proceso de revisión de este artículo.

² En este texto usaremos L2 como término general para incluir cualquier aprendizaje centrado en una lengua que no sea la dominante de los aprendices.



es *esTenTen 18* (ver Kilgarriff y Renau, 2013). En *Sketch Engine* también podemos subir nuestro propio corpus para analizarlo e incluso compartirlo con investigadores específicos o en abierto. El programa es de pago, pero ofrece la posibilidad de un periodo de prueba de un mes y también precios para académicos que son relativamente asequibles para estudiantes y profesores. En la página web se pueden consultar los precios, que dependerán tanto del país del que proviene el investigador como del tipo de suscripción que escoja (mensual, trimestral o anual). La página web también ofrece una variedad de materiales de libre acceso para aprender a usar las herramientas (p. ej. vídeos, descripciones detalladas, glosarios, etc.). En este artículo nos limitaremos a estas tres herramientas de *Sketch Engine* (*LISTA DE PALABRAS*, *CONCORDANCIA* y *PALABRAS CLAVE*) que pueden resultar útiles para estudiar la enseñanza de vocabulario en CCL2. De ahí esperamos que los lectores extrapolen la información a sus propias necesidades de investigación.

En concreto, hablaremos de los corpus P1, P2 y D1³, que creamos con el objetivo común de examinar la enseñanza de vocabulario en el aula (Marcos Miguel, 2015, 2016, 2017 y 2021). Con estos corpus, la decisión de usar *Sketch Engine* se supedita a los objetivos de investigación que se discutirán más adelante. Para situar primero por qué los investigadores examinamos CCL2, revisaremos estudios y metodologías que han utilizado tales corpus y con qué objetivos, en español L2 y también en inglés L2, pues esta sigue siendo la L2 más investigada. De este modo será más fácil entender cómo y cuándo *Sketch Engine* puede ayudar. En la última sección de esta revisión de la literatura ofreceremos una introducción a la situación actual de la enseñanza de vocabulario y morfología léxica en el aula de ELE para entender mejor los análisis que se harán con *Sketch Engine*.

1. Revisión de la literatura

1.1 Investigaciones previas con corpus de clases de lengua (CCL): interacciones e input

Las necesidades de cada investigación (o, incluso, la situación de enseñanza) influyen en la creación y el modo de analizar un CCL. Por ejemplo, el objetivo puede ser que los profesores de una L2 tengan acceso a estos corpus para disponer de ejemplos reales de lengua en el aula y reflexionar sobre las prácticas docentes de otros y las suyas propias (p. ej. en inglés L2 Jäkel, 2010; Nicaise, 2021; Walsh, 2011). También el objetivo del estudio puede ser describir las características del habla de estudiantes o profesores en su contexto habitual (p. ej. en inglés L2 Frigal, Lee, Polat y Robertson, 2017; Horst, 2010; Nicaise, 2021; p. ej. en español L2 Bayés y Montmany, 2012; González Argüello, 2011; López Bastidas, 2023). Este objetivo, analizar la lengua en el aula, no se limita a clases con un enfoque en la enseñanza de la lengua, sino que también incluye a las clases de Aprendizaje Integrado de Contenidos y Lenguas Extranjera (AICLE, también llamado CLIL en inglés) (ver Dalton-Puffer, 2007) o a aquellas clases en las que la L2 sea la lengua de instrucción, aunque no haya un enfoque específico en el aprendizaje de lenguas (ver Jablonkai, 2021). Por

³ Por el tipo de permisos que se obtuvieron cuando se recogieron, estos corpus no están disponibles públicamente. No obstante, si algún investigador está interesado, se podrían compartir las transcripciones para investigaciones específicas.



supuesto, encontramos estudios similares en aulas donde la lengua de instrucción es la lengua dominante de la mayoría de los alumnos, es decir, su L1 (p. ej. Biber, 2006; Kunioshi, Noguchi, Tojo y Hayashi, 2016; Lin, 2012), a la que también pueden asistir alumnos cuya lengua dominante sea otra (p. ej. Basturkmen y Shackleford, 2015; Lessard-Clouston, 2010).

Por su parte, los enfoques metodológicos para investigar un CCL vienen determinados por las orientaciones teóricas de las que se parta. Algunos estudios en los que se han explorado CCL2 siguen enfoques de análisis del discurso, de análisis de la conversación o de análisis de corpus (ver Jablonkai, 2021; González Argüello, 2001; Lee, 2020; Nicaise, 2021; Walsh, 2011). Otras perspectivas teóricas como la lingüística sistémica (p. ej. Dalton-Puffer, 2007) y la teoría sociocultural (ver Thoms, 2012) también justifican la necesidad de analizar un CCL. La mayoría de estos estudios se centran en análisis de interacciones o en análisis del *input*, es decir, en los elementos que componen el discurso del aula. A continuación, mencionamos algunos estudios significativos en los que se han explorado estas dos áreas, interacciones e *input*, desde distintas perspectivas.

Para analizar las interacciones en el aula desde la perspectiva del análisis de la conversación es necesario crear y anotar los CCL siguiendo las convenciones del área (p. ej. Batlle, 2021 para español L2; Hall y Looney, 2021; Seedhouse, 2004; Sert, 2015 para inglés L2). Por ejemplo, en el corpus de González Argüello (2001) se usan los signos de transcripciones propuestos por Jefferson y otros signos que la autora consideró necesarios. Además, siguiendo un formato de tabla, cada turno se escribe en una línea a la que le siguen una serie de columnas donde se anotan los factores examinados en el estudio (objetivo de la interacción, las formas gramaticales, el medio de expresión, fase y el tipo de modificación). Si el turno es más largo que una línea, se añaden líneas a la tabla, pero no se codifican de nuevo las columnas. Por tanto, los turnos se codifican en el propio corpus.⁴

Walsh (2011) también explica cómo se pueden combinar los estudios de análisis de la conversación o del discurso con metodologías de corpus. Aunque para algunos investigadores el análisis de corpus se considera un enfoque de investigación, para otros es una metodología (ver Parodi, 2008). Esta última es la perspectiva que Walsh (2011) defiende. También es bueno recordar que el término *análisis del discurso* es polisémico. Walsh explica cómo este término puede referirse a una escuela de investigación, pero también al *input* del aula. Usar el término *discurso* ayuda a pensar en este *input* como un elemento estructurado organizado por las personas que lo producen y su contexto y no simplemente un continuo de palabras sin contexto.

Las investigaciones sobre el *input* de la clase se han beneficiado de los avances en la lingüística de corpus (p. ej. Horst, 2010; Donzelli, 2007; Meara *et al.*, 1997; Nicaise, 2021). En estos estudios el aula es el contexto de comunicación y, cuando se habla de análisis del discurso, más que de un análisis de cómo los participantes interactúan, se trata del discurso del aula como contexto de aprendizaje. Al investigador le interesa conocer qué elementos

⁴ En la sección “Recomendaciones para futuros estudios” se mencionan brevemente otros sistemas de transcripción.



gramaticales y léxicos se usan para la comunicación, es decir, el *input*. La presencia de dichos elementos y el enfoque incidental o explícito en ellos es lo que facilitará su adquisición. En esta línea de investigación, más allá del uso del vocabulario (p. ej., Horst, 2010; Donzelli, 2007), también se han explorado los elementos gramaticales. Por ejemplo, en el caso de español L2 encontramos revisiones del tiempo y del aspecto verbal (p. ej. Daidone, 2019), de las formas de sujeto (p. ej. Long y Geeslin, 2020) o de los marcadores discursivos (p. ej. Amador Moreno, Chambers y O’Riordan, 2006). Los estudios cuantitativos y automatizados sobre el *input*, tanto léxicos como gramaticales, se pueden complementar con análisis manuales que, desde la perspectiva de la lingüística aplicada, ayudan a comprender los datos cuantitativos y a reflexionar sobre las preguntas de investigación (ver Egbert, Larsson y Biber, 2020).

1.2 La construcción de corpus de clases (CCL)

Para los estudios mencionados, la mayor parte de los investigadores han creado sus propios corpus. No obstante, también hay proyectos de más amplia escala llevados a cabo en una o más universidades y que están disponibles para otros usuarios. En contextos universitarios se han grabado clases de contenido para estudiantes anglófonos (p. ej. BASE, https://www.reading.ac.uk/acadepts/ll/base_corpus/; MICASE, <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>) y para estudiantes hispanohablantes (p. ej. Vázquez y Navas Méndez, 2005, <http://www.comunicacionacademica.com/clasesmagistrales.html>) que pueden usarse como corpus de referencia. También encontramos clases grabadas con y para estudiantes de inglés L2 (CEAPP, <https://ceapp.la.psu.edu/>; Reder *et al.*, 2003, https://labschool.pdx.edu/maelc_access.html). En el contexto escolar, existen también varios corpus con hablantes de alemán como lengua dominante (p. ej. Helmke *et al.*, 2008; Jäkel, 2010; Kurtz, 2013). En todos estos casos se han usado videocámaras o grabadoras de audio o ambas.

Una vez que se han recogido datos de audio o audiovisuales, estos necesitan ser transcritos para su posterior análisis.⁵ La manera de recoger los datos y transcribirlos también podrá verse influida, más allá de los intereses del investigador, por los permisos dados por los participantes o por el tipo de consideraciones éticas de los distintos contextos. Por ejemplo, en el corpus *Michigan Corpus of Academic English* (MICASE, <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>) hay intervenciones de estudiantes que no han sido transcritas pues no se contaba con el permiso del estudiante para incluirlas en el corpus. Por eso es importante saber qué se va a analizar y a quién hay que pedir permiso antes de empezar el proyecto.

Con todo, la manera de transcribir las clases vendrá determinada por el enfoque metodológico. Por ejemplo, en el volumen (Mathieu *et al.*, 2021) encontramos artículos con diversos enfoques al transcribir y analizar los datos. Según el enfoque, los gestos, las miradas y los materiales no son elementos que se puedan pasar por alto. Por eso, la

⁵ Para los investigadores que necesiten más apoyo en la creación del CCL2, algunas lecturas útiles son Adolphs y Knight (2010), Thompson (2010) o Lee (2020).



creación de corpus multimodales es importante para estudios de interacciones que versen sobre los estudiantes, los materiales y los profesores (ver Guerrettaz *et al.*, 2021). CEAPP (<https://ceapp.la.psu.edu/>), con su plataforma digital, es un excelente ejemplo de un corpus multimodal disponible para una variedad de análisis. Sin embargo, para examinar el *input* léxico, el acceso a las transcripciones puede ser más que suficiente, es decir, para un análisis de frecuencias léxicas no hace falta un corpus multimodal. Por ejemplo, las transcripciones que conforman el corpus *The British Academic Spoken English Corpus* (BASE, https://www.reading.ac.uk/acadepts/ll/base_corpus/) se encuentran disponibles dentro de la herramienta a *Sketch Engine*, lo que permite acceder a frecuencias de formas gramaticales y léxicas.

1.3. Enseñanza del vocabulario y de la morfología léxica en el aula de español L2

En esta sección revisaremos brevemente algunos puntos principales sobre la enseñanza de vocabulario en el aula. A grandes rasgos se considera que, si los estudiantes necesitan el vocabulario, lo usan y tienen información para contrastar el uso correcto del mismo, lo aprenderán más fácilmente (ver Nation y Webb, 2011). Por eso, el aprendizaje explícito debe ocupar un lugar destacado en el aula (ver Nation y Webb, 2011). Además, todas las oportunidades de aprendizaje incidental, es decir, el estar también expuestos al vocabulario en actividades e interacciones que no tratan del aprendizaje intencional del mismo, son fundamentales para los estudiantes. En resumen, los profesores deben planificar la enseñanza de vocabulario en el aula (Nation, 2001).

Los estudios de enfoque en la forma han visto que tanto los estudiantes como los profesores se enfocan en vocabulario en lo que normalmente llaman *episodios léxicos* o de *enfoque en la forma léxica* (p. ej. Alcón-Soler, 2009; Marcos Miguel, 2021; Plonsky y Lowen, 2013). Este enfoque en la forma normalmente se clasifica como *reactivo*, causado por un error, o *proactivo*, cuando se cree que un elemento puede causar dificultades. Tanto los estudiantes como los profesores pueden comenzar el episodio (ver Loewen, 2011, para un resumen completo). En general, los estudiantes tienden a aceptar más fácilmente las correcciones léxicas orales que las de otro tipo (ver Lyster, Saito y Sato, 2013). También los estudiantes preguntan para clarificar el significado de palabras nuevas mientras que los profesores lo hacen para asegurarse de que su significado está claro (p. ej., Thoms, 2012; Plonsky y Loewen, 2013).

Ya que el profesor no puede enseñar todo el vocabulario de manera explícita se espera que, por un lado, sea selectivo en los objetivos de su curso. Por otro lado, no debe solo pensar en enseñanza de forma y significado, p. ej. *casa* significa *house* en inglés, sino que también debe incluir elementos relacionados con su forma, significado y uso (ver Nation, 2001). Por ejemplo, dentro de la forma de una palabra un aspecto fundamental es su estructura morfológica, ej., *to housesit*, *cuidar la casa de alguien* tenemos el verbo *sit*, *sentarse*, y el sustantivo *house*, *casa*. Al prestar atención a la forma de una palabra, un estudiante descubre más fácilmente su significado e incluso la memoriza más rápido (Hu y Nassaji, 2012; Laufer, 2023).

De ahí que se anime a los profesores de ELE a desarrollar la conciencia morfológica de sus estudiantes con ejercicios planificados (Serrano-Dolader, 2019). También, cuando los profesores proporcionen explicaciones sobre una palabra, es recomendable que enfatizen su estructura morfológica. Otra buena práctica consiste en incluir los afijos como elementos de aprendizaje léxico, es decir, no solo se deben aprender palabras sino también sufijos como *-dad*, *-ncia*, *-dura*, *-dor*, *-nte* o *-ero*. Estos sufijos son productivos y con una relación relativamente clara entre su forma y su significado, lo que se espera que ayude a los estudiantes. Por supuesto, también se debe recordar a los estudiantes que la atención a la morfología no es la única estrategia de memorización o de reconocimiento que deben usar pues el significado de una palabra no siempre deviene de sus partes y una confianza excesiva en la morfología también conduce a errores.

A pesar de la importancia del aula en el aprendizaje de léxico, este contexto no está tan analizado como uno esperaría. Por ejemplo, Newton (2020) apunta que, en las aulas de inglés L2, faltan estudios para entender cómo los profesores seleccionan vocabulario, cómo desarrollan la competencia de sus alumnos para adquirirlo, cómo conectan el aprendizaje fuera y dentro del aula, cómo maximizan el aprendizaje en el aula y cómo usan la tecnología. Esta situación parece repetirse en español L2 pues, aunque se han estudiado los efectos de actividades específicas y se han dado recomendaciones sobre cómo enseñar vocabulario (ver Sánchez Rufat y Jiménez Calderón, 2015, 2018), apenas se ha examinado el *input* léxico en el aula.

Como se ha afirmado en la sección 1.1., necesitamos ahondar en el aula como un contexto en el que se produce *input* y cómo la presentación de este *input* facilita el aprendizaje o no de determinadas formas. Los estudios que se presentarán a continuación (ver Tabla 1) se recogen en esta línea de investigación al centrarse en el *input* léxico del profesor.

2. Ejemplos de estudios a partir de CCL2 creados por la autora

A continuación mostraremos cómo se han analizado tres CCL2 creados por la autora (P1, P2, D1) en artículos previos. En la Tabla 1 se resumen estos tres corpus y los estudios en los que se han utilizado. Si bien los estudios con P1 y P2 se realizaron de manera manual, consideramos que es útil contraponerlos al análisis con la herramienta *Sketch Engine* (D1). En la sección de análisis, aunque se mencionan los cuatro estudios, se van a comentar dos de ellos en profundidad (Marcos Miguel, 2017 y 2021). Además, se propondrá un nuevo análisis con el corpus P2 para complementar Marcos Miguel (2017).

Tabla 1

Resumen de corpus y estudios que han usado estos corpus

Nombre del corpus	Preguntas de investigación relacionadas con el CCL2 examinado	Número de clases grabadas	Número de profesores	Horas grabadas (aprox.)	Enfoque de la clase	Artículo
*P1	1. ¿Se incluye la enseñanza de morfología derivativa en el aula de español L2? 2. Si es así, ¿qué tipos de episodios de vocabulario aparecen relacionados con la morfología derivativa?	24	5	20 horas	En estudiantes principiantes e intermedios de L2 (de segundo a quinto semestre)	Marcos Miguel, (2017)
*P2	1. “¿Es posible que un profesor profundice en la enseñanza de ML [morfología léxica] aunque ese tema apenas aparezca en el libro de texto si se le anima?” 2. “¿Qué desafíos se le plantean al incorporarla?” (p. 147)	15	3 (presentes en P1)	12 horas y media	En L2 intermedios (cuarto semestre)	Marcos Miguel (2016)
*P2	1. ¿Qué tipo de adaptaciones del libro de texto realizan los profesores de español en sus aulas para la enseñanza de vocabulario?	15	3 (presentes en P1)	12 horas y media	En L2 intermedios (cuarto semestre)	Marcos Miguel (2015)
**D1	1. ¿Qué términos especializados se pueden aprender potencialmente en esta clase por su frecuencia en el discurso y en los materiales curriculares, p. ej. lecturas y <i>PowerPoints</i> preparados por el profesor? 2. ¿Cómo pueden estos materiales curriculares (lecturas y <i>PowerPoints</i>) contribuir a los episodios léxicos de enfoque en la forma para términos especializados?	10	1	8 horas	En contenido (literatura) con estudiantes de L2 intermedios y avanzados y hablantes de herencia (de tercer o cuarto año)	Marcos Miguel (2021)

*=Análisis manual

**=Análisis manual y con Sketch Engine

Todos estos estudios se centran en aulas de ELE en distintas universidades de los Estados Unidos. En todos estos casos era importante documentar la experiencia del profesor observado y de los alumnos con los que trabajaba. Por eso, aparte del análisis de corpus que se plantea aquí, también se recogieron datos adicionales como entrevistas a los profesores. En este artículo nos centraremos sólo en las transcripciones.



3. Análisis de CCL2 con *Sketch Engine* y de manera manual

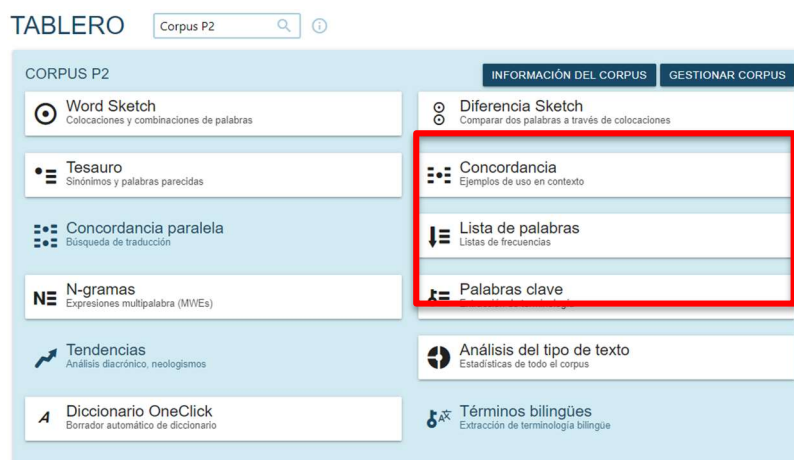
Esta sección se divide en dos partes. En la primera subsección (3.1) mostraremos cómo *Sketch Engine* puede complementar análisis manuales previos cuando se plantean nuevas preguntas de investigación. En la segunda subsección (3.2) nos centraremos en el artículo (Marcos Miguel, 2021), en el que, tras un primer análisis cuantitativo con *Sketch Engine*, se abordó un análisis manual.

Por otro lado, esta sección aunará la información metodológica con el análisis lingüístico. En el análisis lingüístico se ofrecerán breves reflexiones tanto del *input* léxico como de las interacciones y se conectará brevemente con la literatura sobre aprendizaje del léxico. Un análisis meramente descriptivo con herramientas de corpus no es suficiente para entender cómo tiene lugar la enseñanza en el aula (ver Egbert, Larsson y Bibber, 2020).

La Figura 1 muestra enmarcadas en rojo las herramientas de *Sketch Engine* en las que nos centraremos: *CONCORDANCIA*, *LISTA DE PALABRAS* y *PALABRAS CLAVE*. La interfaz incluye la opción de español que se ha preferido usar en este artículo. En las dos siguientes subsecciones, se incluirán capturas de pantalla de la interfaz para visualizar mejor cómo se desarrollaron los análisis.

Figura 1

Herramientas en el Tablero de *Sketch Engine*



Los investigadores que quieran replicar las explicaciones que aquí se ofrecen con su propio corpus deben primero cargarlo en la plataforma *Sketch Engine*. En la página web de la plataforma se encuentra información detallada de cómo hacerlo (<https://www.sketchengine.eu/guide/create-corpus-from-files>). En los análisis que se describen a continuación los archivos se subieron con formatos .txt, .doc o .docx y no se borraron las anotaciones que se habían hecho directamente en estos documentos. Esta forma de trabajar con los textos trae consigo ciertas consecuencias.

Primero, aunque un CCL2 es normalmente un corpus bilingüe o incluso multilingüe, al compilar el corpus (<https://www.sketchengine.eu/guide/compile-a-corpus/>), se debe seleccionar una lengua, en este caso “español” (ver Figura 6). La aplicación asumirá que todos los lemas generados son de esta lengua. Segundo, aunque es posible

trabajar con anotaciones y *Sketch Engine* sin que esto suponga ningún problema a la hora de lematizar los datos (<https://www.sketchengine.eu/guide/annotate-a-corpus/>), si las anotaciones no siguen este formato, se lematizarán también. A lo largo de los análisis siguientes se señalarán momentos en los que estas opciones influirán en los resultados. En la sección “recomendaciones para futuros estudios” se reflexionará de nuevo sobre estos dos aspectos.

3.1 Análisis manuales complementados con *Sketch Engine*

En el caso de los tres primeros estudios de la Tabla 1, los objetivos eran localizar 1) episodios de vocabulario donde se tratara la morfología derivativa (Marcos Miguel, 2017 y 2016) y 2) actividades de vocabulario que se trabajaran en clase (Marcos Miguel, 2015). Para alcanzar estos objetivos puede ser suficiente con revisar manualmente las transcripciones y los audios. De esa manera, a partir de las transcripciones, se crea un documento de texto donde se copian todos los ejemplos y también una hoja de cálculo donde se catalogan.

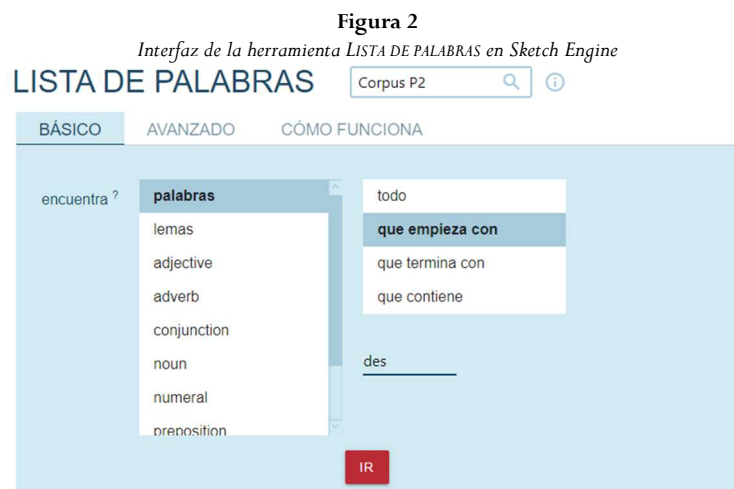
El Ejemplo 1 muestra uno de los episodios de vocabulario de P1. En este caso la profesora se enfocaba en la palabra *despegar* y en el prefijo *des-*. El nombre que se le da a la profesora, Sally, es un seudónimo.

Ejemplo 1
(Marcos Miguel, 2017, p. 51)

1	Sally: ¿Qué hizo el avión? ¿Sí?
2	Estudiante1: Aterrizo.
3	Sally: Bien, aterrizó [pizarra]. Ok, aterrizó, que es de tierra. Cuando llega de nuevo a la tierra, ¿Ok? ¿Cuál es el opuesto de aterrizar, Estudiante2?
4	Estudiante2: Despegar.
5	Sally: Despegar . Ok [pizarra]. Despegar . Es como... ¿saben qué quiere decir ‘pegar’? Como lo que hacen con cintas, o por ejemplo, es pegar, ¿no? Entonces, despegar es como uuffs [<i>Gesto de separar las manos</i>]. The plane literally unsticks itself from the ground. Ok, muy bien. Ok, ¿y número 6, qué hace? [...]

Si para ampliar este análisis hubiéramos querido explorar más usos del prefijo *des-* en las clases de Sally, *Sketch Engine* habría sido útil para buscar todas las formas con este prefijo. *Sketch Engine* permite recuperar todos los casos que comienzan con un principio de palabra específico con la herramienta *LISTA DE PALABRAS*. A continuación describiremos cómo es posible hacer este análisis y cómo sería beneficioso para ampliar este primer estudio (Marcos Miguel, 2017).

La Figura 2 muestra el interfaz básico de la herramienta *LISTA DE PALABRAS*. Aunque se pueden hacer búsquedas con más delimitaciones en la pestaña *AVANZADO*, en este caso la opción *BÁSICO* es suficiente. Para resolver dudas, la tercera pestaña, *CÓMO FUNCIONA*, incluye información sobre la herramienta y la terminología que usa. Por ejemplo, el vídeo informativo, en inglés con subtítulos en español, está disponible en *YouTube* (<https://www.youtube.com/watch?v=nqpCIICCEdw>). En resumen, la Figura 2 muestra cómo buscar la secuencia *des* al principio de cada palabra en todas las clases de Sally del corpus P2.



Aunque el análisis se ha llevado a cabo con la opción “palabras” (Figura 2), se podría también hacer con la opción “lemas.” No obstante, se ha preferido usar la opción “palabras” pues, como se ha mencionado al principio de la sección (3), la lematización de un corpus bilingüe no siempre es correcta. *Sketch Engine* lematiza como si todo el texto estuviera en la misma lengua. Tras realizar la búsqueda con *des* (Figura 2), llegamos a la pantalla de la Figura 3 con todas las formas que aparecen en el corpus. La herramienta también permite descargarse los datos en formato cvs, xlsx, xml e incluso como pdf. En la esquina superior derecha hemos enmarcado en rojo la opción para las descargas.



Una vez que hemos descargado el documento csv generado en la Figura 3, la Tabla 2 muestra las formas con *des*. En el momento de la grabación del corpus P2, Sally ya había hablado con la investigadora sobre la importancia de la enseñanza de la morfología derivativa en el aula. Como en la primera tanda de observaciones (recogida en Marcos Miguel, 2017) Sally sacó provecho del ejemplo con *des-* (Ejemplo 1), se esperaría algún enfoque similar en la metodología en esta segunda tanda de observaciones. Es decir, ahora buscaríamos saber si el enfoque en este prefijo es recurrente en sus clases.

Tabla 2

Formas que empiezan con *des-* en la clase de Sally (corpus P2) generadas con la herramienta LISTA DE PALABRAS en Sketch Engine

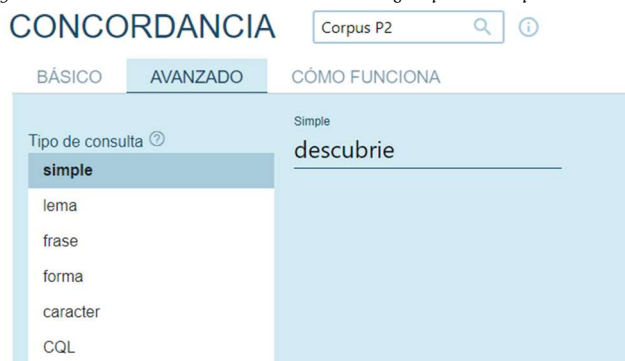
Item	Frequency
Desde	33
Después	10
Describe	4
Descargar	3
Destacar	3
Descubiertos	2
Describir	2
Descubrió	2
Describing	1
Descubrieron	1
Descansen	1
Descuido	1
Descubrie	1
Descriptiva	1
Descripción	1
Desaparece	1
Descobrir	1
Descrito	1
Descubierta	1
Desconocidos	1
Descubierto	1
Descubiertas	1
Desaparecieron	1
Des	1

De la Tabla 2 pasamos a una exploración manual. Por ejemplo, como no todas las formas que empiezan con *des-* son prefijos ni palabras en español, se eliminan para análisis posteriores (*desde* y *después*; *describing*). Este ejemplo con *describing* ilustra el bilingüismo del corpus. Este bilingüismo presenta una dificultad para las herramientas de corpus pues, hasta donde llega mi conocimiento, no existe ninguna herramienta que pueda reconocer dos lenguas a la vez. En este caso específico, un análisis manual ayuda a eliminar las palabras en inglés.

Además, en la Tabla 1, algunas de las formas no parecen palabras en español como *des* o *descubrie*. Una búsqueda de las concordancias con la herramienta *CONCORDANCIA* de *Sketch Engine* localiza las palabras exactas en contexto.

Figura 4

Interfaz de la herramienta CONCORDANCIA en Sketch Engine para la búsqueda de “descubrie”



En ambos casos, se trata de momentos donde la profesora no usó la palabra completa como se ve en Ejemplo 2, que se ha copiado y pegado del resultado de la búsqueda. Como la forma de transcribir que usamos en estos corpus es literal, es normal encontrar formas que no se corresponden con palabras tal y como se recogerían en un diccionario.

Ejemplo 2

Concordancia de “descubrie”

cuando tenemos la voz pasiva también tenemos por alguien. </s><s> O sea quién lo hizo, ¿no? </s><s> Pero también podemos usar esta estructura cuando estamos hablando por ejemplo de un momento importante. </s><s> No. </s><s> Entonces por ejemplo (silence) a ver, laaa, los europeas **descubrie**, descubrieron las Ámericas en mil, ah, 492. </s><s> Entonces lo importante es la fecha, no Cristóbal Colón, en esta, este caso, ¿no? </s><s>

En la Tabla 3 encontramos la revisión manual de la Tabla 1 con los lemas relevantes. Por ejemplo, no aparecen las palabras *desde* o *después*, ni las formas *descripción*, *describir* o *destacar*. Estas últimas no parecen reconocibles por un estudiante como prefijo en la actualidad y no cargan con los significados de *des-* como elemento contrario al verbo de la raíz. Es decir, los estudiantes descubren más fácilmente los significados de las palabras cuando se combinan raíces y sufijos que reconocen (p. ej., Marcos Miguel, 2018; Sánchez Gutiérrez, 2020). Cuando hablamos de consciencia morfológica, no nos referíamos a un conocimiento etimológico.⁶

Tabla 3

Prefijo *des-* en la clase de Sally (corpus P2) (recopilación manual)

Ítem y número de apariciones
descargar (3)
descubrir (9)
descansar (1)
desaparecer (2)
descuido (1)
desconocido (1)

⁶ Por supuesto, no sería inapropiado trabajar la etimología en la clase de ELE para facilitar la adquisición de vocabulario. Este es un campo que va más allá de este artículo. Para algunas ideas prácticas para profesores recomendamos Pierson (1989).

En la Tabla 3, se recogen cuatro verbos y dos sustantivos que podrían potencialmente haber sido objeto de enseñanza en las clases de Sally. No obstante, como sabemos que los profesores tienden a repetir las palabras que explican, una aparición de una vez en el corpus sugiere que no hay un enfoque explícito en la palabra. A continuación se buscan las concordancias de todos estos lemas.

La Figura 5 muestra la interfaz de la herramienta *CONCORDANCIA* para buscar los lemas de *descargar*, que tiene tres apariciones. Para examinar si se incluye o no enseñanza explícita de las otras palabras seleccionadas en la Tabla 3, cada palabra se buscará individualmente.

Figura 5

Interfaz de la herramienta *CONCORDANCIA* en *Sketch Engine* para buscar el lema "descargar" en la clase de Sally (P2)



La Tabla 4 recoge siete concordancias que hemos seleccionado de las generadas con la herramienta *CONCORDANCIA*. Las hemos copiado y pegado directamente de la aplicación digital al documento de texto. También es posible descargar las concordancias en formato txt, csv, xlsx, xml o pdf de la misma manera que se puede hacer con la herramienta *LISTA DE PALABRAS* (ver Figura 3).

Tabla 4

Ejemplos de palabras con el prefijo *des-* en contexto generadas con la herramienta *CONCORDANCIA* en *Sketch Engine* de la clase de Sally (P2)

Palabra	Concordancia
i. descargar	S. ah, (incomprehensible) de música #00:41:56-4# T. No se prohíbe descargar música, del internet. #00:41:58-2# S. mm #00:41:59-4# T. okay, muy bien, en su, durante su gobierno no se prohíbe descargar música del internet, okay, muy bien, ¿qué más? </s><s>
ii. descubrir	S. "cada vez que quería tomar decisiones importantes acerca de la política y la religión, el rey Arturo se reunía con los caballeros de la mesa redonda. </s><s> En parejas estudien las pistas para descubrir quién es quién." [S reading aloud] #00:25:45-3#
iii. descansar	Okay. </s><s> Buen fin de semana. </s> <s> Descansen y no festejen tanto.
iv. descuido	El primero dice, yo perdí el collar que me regaló mi mamá, ¿no? </s><s> Yo perdí el collar. </s><s> En ese caso no lo usamos e implica un poco de culpabilidad. </s><s> Un descuido o estaba enojada con ella. </s><s> Entonces, no lo cuidé bien o lo que sea. </s><s> ¿Okay? entonces en la estructura accidental, lo cambiamos para decir se me perdió el collar, ¿no? </s><s> El collar "it lost itself",

v. desaparecer	¿Cuál es un ejemplo de una "polémica" ese, en los debates políticos de hoy? </s><s> Sí. #00:08:42-5# S. ah, pienso en el vuelo que desaparece #00:08:46-0# T. aha #00:08:47-0# S. en, de Indonesia. #00:08:50-7# T. Aha ¿Y por qué es una polémica? #00:08:55-7# S. porque ahh CIA no da información a la, al país que tiene #00:09:06-4# T. okay, bien, porque no,
vi. desconocido	#00:47:51-3# S. con amigos con los (incomprehensible) es muy (incomprehensible) común pero con la gente general es un poco menos común, normal #00:48:07-7# T. es un poco más o ¿menos normal dices? #00:48:10-8# S. sí #00:48:10-8# T. con los desconocidos #00:48:13-4# S. sí #00:48:13-4# T. ah, okay, entonces hay que como meter el dedito en el agua primero #00:48:19-6# S. sí #00:48:19-6# T. para saber si podemos hablar de esas cosas o no. </s><s> Okay,

Con este análisis de concordancias, confirmamos que no hay ninguna enseñanza explícita de este prefijo. Sin embargo, en el ejemplo IV con *descuido*, la profesora usa también otra palabra de la misma familia, *cuidó*, que podría facilitar el reconocimiento incidental de los estudiantes del sustantivo *descuido*.

Además, podemos suponer que varias palabras ya son conocidas por los estudiantes. En el ejemplo I con *descargar* y en el ejemplo V con *desaparece*, las palabras vienen de los estudiantes con lo que un enfoque en la forma léxica explícito parece innecesario por parte del profesor. En el caso II de *descubrir* encontramos esta palabra con cierta frecuencia (un total de 10 veces entre formas verbales y adjetivas). Debido a este elevado número de repeticiones, podemos imaginar que la palabra es conocida por los estudiantes.

Otra palabra posiblemente conocida es *descansar* (III). Imaginamos que los estudiantes ya la conocen pues *cansado*, de experiencia anecdótica, es una de las palabras que los estudiantes suelen aprender en su primer semestre de español para responder a la pregunta *¿Cómo estás?* Al contemplar este dato adicional, que proviene del conocimiento curricular de los investigadores, no obstante, nos salimos de la transcripción. Con todo, la triangulación de datos suele ser una estrategia que se favorece en estudios con corpus (ver Egbert y Baker, 2019). Sin duda, esta triangulación nos ayudará, pero siempre es importante explicar bien el porqué de nuestras inferencias.

Finalmente, nos queda la palabra *desconocidos* (VI). *Conocer* es uno de los verbos que los estudiantes aprenden relativamente pronto pues se suele incluir en los libros estadounidenses de primer semestre. En este caso, la profesora usa la palabra *desconocidos* para matizar la expresión *la gente general* del alumno. Esta expresión no es muy idiomática y de ahí entendemos la reformulación de la profesora. Por tanto, vemos un enfoque en el léxico en este episodio, pero no con un enfoque destacable en la morfología.

En general, aunque este análisis muestra poco potencial de aprendizaje del prefijo *des-* en la clase, parece que las palabras que se usan con *des-* ya son conocidas por los estudiantes, lo que justifica la falta de reflexión sobre este afijo en el aula. Para afianzar este análisis sería apropiado consultar con la profesora y de ese modo confirmar estas hipótesis. A pesar de que en este artículo nos centramos en la importancia de las transcripciones de clases, también reivindicamos indagar en las creencias y conocimientos del profesor (ver Borg, 2006) para complementar los análisis de corpus, como se hace en Marcos Miguel (2017) y (2016). Las entrevistas a los profesores ayudan a triangular los datos.

Resumimos aquí los puntos principales de este análisis modelo y cómo puede servir para otros análisis de afijos. En Marcos Miguel (2017), donde se encuentra el ejemplo 1, el análisis fue manual. Se empezó con los datos y de ahí se pasó a la categorización. Es decir, se crearon categorías que mostraban cómo la morfología se trabajaba o no en la clase (p. ej. *planeada*, *no planeada*, etc.). En esta sección (3.1.), este primer análisis se ha complementado con *Sketch Engine* al buscar automáticamente todas las palabras que contenían el sufijo *des-* (ver Figura 2) en un nuevo CCL2. Algo fundamental para el análisis es que la profesora es la misma en los dos corpus (ver Tabla 1).

Sketch Engine ofrece varios puntos de partida para la investigación de palabras derivadas (o de otro tipo de palabras dado el caso). Para empezar, *Sketch Engine* permite que los investigadores se centren en los afijos en el discurso de la clase. Por un lado, proporciona la lista de palabras más frecuentes en la clase de la que podemos compilar una lista de afijos frecuentes que analizar posteriormente. Por otro lado, como se ha visto en el ejemplo con *des-*, partiendo de estudios previos sobre ciertos afijos, podemos profundizar en cómo se trabajan en la clase. En ambos casos, de la lista de frecuencias con todas las palabras o de la lista de frecuencias con un afijo específico, podemos pasar a análisis manuales. En la siguiente sección (3.2.) se sigue explorando cómo *Sketch Engine* ayuda a que los investigadores realicen subsecuentes análisis manuales. De nuevo, el enfoque depende de los objetivos de investigación.

3.2 Análisis con *Sketch Engine* complementados con análisis manuales (Marcos Miguel, 2021)

Como se ha indicado en la Tabla 1, en Marcos Miguel (2021) era necesario usar herramientas de corpus para la primera pregunta que se exploraba (1. ¿Qué términos especializados se pueden aprender potencialmente en esta clase por su frecuencia en el discurso y en los materiales curriculares, p. ej. lecturas y *PowerPoints* preparados por el profesor?). Para contestar a la segunda pregunta (2. ¿Cómo pueden estos materiales curriculares, lecturas y *PowerPoints*, contribuir a los episodios léxicos de enfoque en la forma para términos especializados?), se requerían análisis manuales de los corpus.

En este estudio se analizaron cinco corpus: 1) el calendario (aunque en este caso más bien podemos decir que es solo un documento, 2) los *PowerPoints* creados por el profesor, 3) las lecturas en inglés, 4) las lecturas en español y 5) las transcripciones de la clase. Este último es por tanto el CCL2 del que estamos hablando. La diversidad de corpus analizados se debe a que, aparte del discurso de las clases reflejado en las transcripciones, o sea, el CCL2, también es importante considerar otras fuentes que influirán en el aprendizaje de los estudiantes como los materiales de lectura. En este caso exploramos una clase de cultura impartida en la L2.

Para crear el CCL2, una vez que se transcribieron las clases, se subieron a *Sketch Engine* estas transcripciones como archivos de txt⁷ en la herramienta *NUEVO CORPUS*. La Figura 6 muestra cómo asignamos la lengua al corpus y le damos un nombre. La opción de corpus multilingüe es sólo para corpus paralelos, con lo que aquí debemos seleccionar

⁷ También es posible subir los materiales en otros formatos, pero por experiencias anteriores, recomendaría crear corpus en formato txt con Unicode-UTF8. Esto soluciona algunos posibles quebraderos de cabeza pues no todas las aplicaciones reconocen los caracteres del español.



Spanish como lengua principal del corpus. En los siguientes pasos añadiremos las transcripciones y compilaremos el corpus, lo que significa que *Sketch Engine* lo lematizará y lo etiquetará con la información de la clase gramatical a la que pertenecen los lemas. Esta lematización facilitará enormemente nuestro análisis posterior.

Figura 6

Interfaz de la herramienta CREAR CORPUS en Sketch Engine

Como la unidad de análisis del estudio eran los términos especializados que los estudiantes pudieran aprender, la primera tarea fue generar una lista de frecuencias del CCL2. En el artículo original, esta lista se creó con *VoyantTools* (Sinclair y Rockwell, 2020, <https://voyant-tools.org/>). Esta interfaz gratuita, *online* también permite crear listas de palabras y buscar concordancias, pero no lematiza los textos. Con *Sketch Engine* se genera la lista con la herramienta *LISTA DE PALABRAS* (ver Figura 2).

A partir de esta lista se seleccionaron los términos especializados que constituirían el posterior objeto de análisis. Los dos criterios de selección implicaron un análisis manual de la lista. El criterio principal se centró en la definición de *término especializado* en el estudio. Como la clase trataba de religión, artefactos culturales y comunidades religiosas de la Península Ibérica en la Edad Media, se consideraron *términos especializados* los que se centraban en estos temas. De este modo, palabras frecuentes como *iglesia* se clasificaron como especializadas. Es decir, cuando eran palabras polisémicas, se tuvo en cuenta su significado especializado en el aula. El ejemplo para ilustrar el significado especializado de *iglesia* era el siguiente: “aquí hablo de la iglesia en los términos teológicos, como no de edificios, sino . . . ni siquiera un lugar físico, sino la comunidad de la gente” (pp. 40-41). Un segundo criterio fue que las palabras

aparecieran al menos cuatro veces en el discurso para que se distribuyeran en más de un turno. De esta manera, se garantizaba su relevancia y su potencial de aprendizaje. En total, se localizaron 27 términos especializados⁸.

A continuación, con la herramienta *CONCORDANCIA* (ver Figura 5) se buscaron las 27 palabras en cada corpus y así se recopilaban sus frecuencias específicas (ver Marcos Miguel, 2021). De esta manera se comprobó si las palabras aparecían en los distintos materiales de la clase.

En un tercer análisis se usó la herramienta *PALABRAS CLAVE* (ver Figura 7) para calcular el valor de *keyword* de una palabra o de una expresión multipalabra (también llamadas *término* en *Sektch Engine*). Este valor *de keyword* se refiere a la frecuencia relativa entre un corpus de referencia, en este caso *esTenTen18*, y el corpus que se examina, aquí el corpus de transcripciones (ver Kilgarriff, 2009). Cuanto más alto sea el valor, más específica será la palabra al corpus que se examina. La Figura 7 muestra las opciones de búsqueda que vienen por defecto en *Sketch Engine*. Es decir, tanto si se busca con *básico* o con *avanzado*, sin cambiar los parámetros, se producirán los mismos resultados.



La Figura 8 muestra el resultado de la búsqueda con las opciones por defecto. Al clicar en el botón enmarcado en rojo en la esquina derecha se pueden descargar los valores de las 1.000 primeras palabras.

⁸ Hay distintos enfoques para definir los términos especializados de los que no hablamos aquí, pero recomendamos Dang (2020) para ampliar sobre este tema en el aula de L2.

Figura 8

Lista de palabras únicas - generada por la herramienta PALABRAS CLAVE en Sketch Engine del corpus D1

Lema	Lema	Lema	Lema	Lema
1 okey	11 s9	21 groups	31 s15	41 comment_start
2 overlap	12 s6	22 s3	32 s7	42 sneezes
3 incomprehensible	13 s2	23 corán	33 catecismo	43 allahu
4 laughing	14 morisco	24 s10	34 abraham	44 diso
5 mumbling	15 abindarráez	25 hadiz	35 mudéjar	45 s11
6 students	16 s1	26 writing	36 abencerraje	46 blah
7 silence	17 jarifa	27 professor	37 ismael	47 cosmología
8 aljamiado	18 s4	28 okay	38 alfaquí	48 islam
9 coránico	19 talking	29 potifar	39 hagar	49 arabic
10 s13	20 s5	30 s12	40 comment_end	50 fillo

En la lista (Figura 8), algunas palabras se refieren a anotaciones en el corpus (*S9=estudiante 9, comment_start*). Aunque a efectos del discurso de la clase estas palabras no son relevantes, sí que son características del corpus y no del corpus de referencia. Para trabajar mejor con los datos, nos descargamos el documento en formato csv. En la Figura 9 obtenemos tanto los valores de *keywordness (score)*, frecuencias generales y frecuencias relativas del corpus de referencia (*esTenTen18*) y del corpus de transcripciones.

Figura 9

Lista de valores de *keywordness* y frecuencias (generales y relativas) generados con la herramienta PALABRAS CLAVE en Sketch Engine del corpus D1

Item	Frequency (focus)	Frequency (reference)	Relative frequency (focus)	Relative frequency (reference)	Score
6 okey	272	6618	4221.17725	0.33777	3156.126
7 overlap	128	643	1986.4364	0.03282	1924.286
8 incomprehensible	75	89	1163.92761	0.00454	1159.66
9 laughing	68	1155	1055.29431	0.05895	997.493
10 mumbling	49	4	760.43268	0.0002	761.277
11 students	50	6582	775.95172	0.33593	581.579
12 silence	49	8597	760.43268	0.43878	529.222
13 aljamiado	35	760	543.1662	0.03879	523.847
14 coránico	43	6083	667.31848	0.31047	509.985
15 s13	33	545	512.12811	0.02782	499.241
16 s9	45	7936	698.35651	0.40504	497.748
17 s6	61	19625	946.66107	1.00163	473.445
18 s2	61	21797	946.66107	1.11248	448.6
19 morisco	73	30976	1132.88953	1.58097	439.328
20 abindarráez	28	149	434.53296	0.0076	432.246
21 s1	48	15480	744.91364	0.79007	416.694

Con esta lista se obtuvieron los valores de *keywordness* de las palabras seleccionadas previamente y se confirmó su idoneidad como términos especializados. Como ejemplo, en las primeras 15 palabras de la Figura 9, se localizan ya tres de las palabras especializadas (*aljamiado, coránico, morisco*) y un nombre propio (*Abindarráez*).

No obstante, aunque en Marcos Miguel (2021) se siguió esta metodología, también se podrían determinar los términos especializadas del corpus a partir de un análisis de *keywordness*. Con un CCL2 limpio, sin anotaciones o usando la opción “excluir palabras” (ver Figura 7), se extraerían automáticamente los términos especializados. También se podría mantener el criterio de frecuencia de al menos en cuatro ocasiones con los ajustes de visualización (ver Figura 7, “frecuencia mínima”).

En el estudio de Marcos Miguel (2021) se prefirió la revisión manual de la lista de frecuencias ya que ayudó al investigador a familiarizarse con el corpus. Esta es una ventaja de los análisis manuales que no se puede obviar. Por ejemplo, de las 27 palabras seleccionadas, cuatro no estaban entre las 1.000 más frecuentes. Dos de ellas eran nombres compuestos que aparecían en la lista generada en la pestaña *EXPRESIONES MULTIPALABRAS* (Figura 8) y las otras dos era palabras frecuentes pero polisémicas (*español*, *exposición*). Al haberse completado primero el análisis manual se sabía que el significado usado en el CCL2 difería del más frecuente. En este caso, *exposición* se refería al comienzo de un texto. Además, al tratar la clase de comunidades religiosas en la Península Ibérica, el término *español* no marcaba una comunidad religiosa específica, según explicó el profesor en clase. Por tanto, el significado especializado en clase era más restringido que el habitual.

Para el análisis de la enseñanza de vocabulario se realizó un análisis manual de los momentos de enfoque léxico en la forma (ver sección 1.3.). De los 83 ejemplos que se encontraron, 31 se referían a los 27 términos especializados que se habían examinado. Para ilustrar la enseñanza en la clase, en el artículo original se mostró en detalle la interacción con la palabra *cosmología*. De esta manera se añadió una perspectiva más cualitativa que dio importancia a los *PowerPoints* que se proyectaban en la clase como fuente de selección del léxico y a la co-construcción de la enseñanza del vocabulario tanto por parte del profesor como de los alumnos con la ayuda de los materiales de la clase (pizarra, lecturas, *PowerPoints*).

Por tanto, si en la sección anterior hemos ampliado un análisis manual con *Sketch Engine*, en esta sección hemos revisado cómo un análisis de corpus para el que se usaron las herramientas de *Sketch Engine* se complementó con un análisis cualitativo del discurso de la clase. En el análisis con *Sketch Engine*, aparte de las herramientas *LISTA DE PALABRAS* que ya habíamos descrito en la sección anterior (Figura 2 y 3), hemos revisado las herramientas *CREAR CORPUS* (Figura 6) y *PALABRAS CLAVE* (Figura 8). Además, hemos propuesto dos maneras para seleccionar términos especializados que pueden ser el objeto de enseñanza en clases de contenido impartidas en la L2 (primero automático y luego manual o solo automático). Como se ha indicado en la sección 3.1., esta metodología se podrá replicar y ajustar a nuevas preguntas de investigación.

Recomendaciones para futuros estudios

Esperamos que los análisis que se han mostrado en este artículo sirvan como inspiración para otros estudios con CCL2 relacionados o no con la enseñanza de vocabulario. A continuación, daremos algunas recomendaciones para futuros estudios en CCL2. Específicamente reflexionaremos sobre la naturaleza bilingüe de los CCL2 al igual que la importancia de la transcripción y la lematización para los análisis posteriores.⁹

En primer lugar, un desafío técnico que tenemos los investigadores de CCL2 se debe a la presencia de más de una lengua en el aula. Por ejemplo, si queremos calcular el número de palabras (lemas y formas, *types/tokens* en inglés), tenemos que estar seguros de que el texto está solo en una lengua. Además, cuando *Sketch Engine* compila el corpus y lo lematiza, si existe una palabra en otra lengua que es formalmente similar a una palabra en la lengua del corpus, lo lematizara así. Por ejemplo, *time* (*tiempo*) puede ser lematizado como *timar*, y *more* (*más*) como *morar*. Ante esta dificultad, recomendamos subir tres corpus a *Sketch Engine*: uno con el texto en español, otro con el texto en inglés y otro completo para examinar mejor las interacciones. Otra opción posible, pero que a la larga supone más trabajo, es examinar manualmente la lista de palabras generadas para asegurarnos de que la información es correcta. Más allá de las dificultades por las combinaciones de lenguas, algunas palabras no se lematizan bien. Por eso, recomendamos hacer búsquedas adicionales con palabras simples. Como ejemplos anecdóticos, en otro corpus que examinamos, la lematización de *res* (*animal*) pasaba a ser *re*, y la de *vaca*, *vacar* o *vaco*.

También la calidad de nuestra transcripción influirá en la lematización en *Sketch Engine*. Es fundamental que la transcripción se haya hecho de manera cuidadosa siguiendo unas convenciones específicas del campo o que el propio investigador haya establecido. En los corpus P1, P2 y D1 se optó por una transcripción ortográfica y se decidieron algunas anotaciones como por ejemplo (*silence*), (*blackboard*), (*overlap*), (*incomprehensible*) y (*laughing*). Si el investigador quiere usar transcripciones más sistemáticas, sugerimos el sistema CHAT (MacWhinney, 2023) que se ha implementado en una variedad de corpus orales (ver MacWhinney, 2000). Por ejemplo, CEAPP sigue este sistema que muestra de manera simultánea la grabación de la clase y la transcripción. Otro sistema de transcripciones que se utiliza frecuentemente es el *Jeffersonian transcript* que Park y Hepburn (2022) recomiendan.

Otra consideración para tener en cuenta con *Sketch Engine* y nuestra transcripción es cómo escribimos los nombres compuestos, tanto compuestos ortográficos como sintagmáticos. Si son importantes en nuestro análisis, podemos escribirlos juntos para que aparezcan como palabras individuales y así aparezcan en la lista de palabras. También podemos explorar la herramienta *N-GRAMS* que sirve para buscar compuestos sintagmáticos frecuentes. Dentro de la herramienta *PALABRAS CLAVE*, encontramos también una lista de *expresiones multipalabras* (ver Figura 7) que

⁹ Partimos de la base de que el lector de este artículo es un investigador con poca experiencia en el uso de corpus en general, de CCL2 y de *Sketch Engine*. Para usos más avanzados con *Sketch Engine* recomendamos los tutoriales de su página web y su servicio de consulta.



nos permitirá encontrar compuestos. Todas estas opciones las tiene que considerar como siempre el investigador según sus necesidades.

Por otro lado, en los artículos de los que hemos hablado, no se han analizado de manera sistemática los errores de los estudiantes pues ese no era el objetivo de los estudios. Aunque los CCL2 no son corpus de aprendices, sí que incluyen a aprendices como interlocutores fundamentales. Por tanto, el investigador tiene que decidir qué quiere hacer con los errores y si necesita algún sistema de etiquetación de errores (ver Díaz-Negrillo y Fernández-Domínguez, 2006; Lüdeling y Hirschman, 2015).

En futuros estudios, tanto si el enfoque es en la interacción, en características lingüísticas o en errores de estudiantes, según cómo se prepare el documento, aparecerán anotaciones en las herramientas como *LISTA DE PALABRAS*. Una opción relativamente sencilla es elegir la opción *avanzado* en la herramienta *LISTA DE PALABRAS* (Figura 2) y excluir palabras que se sepa que son anotaciones y no parte de la clase. Por ejemplo, en el corpus P2 se utilizan expresiones como (*silence*) o (*blackboard*), que se pueden quitar de esta manera. Tener cuidado con esta información es importante sobre todo cuando el investigador quiere mostrar el número total de palabras (lemas y formas) en el corpus. Por supuesto, podemos subir otro corpus en el que no aparezca ningún tipo de anotación. Otra opción es seguir el sistema de anotaciones de *Sketch Engine* para evitar que se produzcan lematizaciones incorrectas. A su vez, usar este tipo de anotaciones permitirá recuperarlas para una variedad de análisis (ver <https://www.sketchengine.eu/documentation/corpus-querying/>).

Por último, añadir metadatos, es decir, información adicional sobre el corpus, puede resultar fundamental tanto para el creador del CCL2 como para otros investigadores con los que se comparta. En los corpus de aprendices ya se insiste en que se estandaricen los metadatos (ver Paquot, 2023; Paquot *et al.*, 2023). Esta estandarización también sería una práctica muy recomendable para los CCL2. Entre los metadatos que podrían ser útiles para otros usuarios se encontraría información sobre el profesor (p. ej., L1, L2, años de experiencia, formación), el contexto de enseñanza (p. ej., tipo de centro, número de alumnos por clase, objetivos del curso), sobre el alumnado (p. ej., edad media, L1, L2, nivel de competencia lingüística medio, número) y sobre los materiales didácticos (p. ej., nombre del libro de texto).

En resumen, después de esta descripción de cómo usar algunas de las funciones de *Sketch Engine*, recomendamos a los investigadores que siempre piensen en análisis manuales complementados con *Sketch Engine* o en el uso de *Sketch Engine* complementado con análisis manuales. Este tipo de análisis nos ayudará a conocer mejor los datos, lo cual es fundamental no solo antes sino también después de usar cualquier tecnología.

Referencias

Adolphs, Svenja; Knight, Dawn (2010). Building a spoken corpus: what are the basics? En Anne O’Keeffe y Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 21-34). Routledge.



- Alcón-Soler, Eva (2009). Focus on form, learner uptake and subsequent lexical gains in learners' oral production. *International Review of Applied Linguistics in Language Teaching*, 47(3-4), 347-365. <https://doi.org/10.1515/iral.2009.015>
- Amador Moreno, Carolina. P.; Chambers, Angela; O'Riordan, Stéphanie (2006). Integrating a corpus of classroom discourse in language teacher education: the case of discourse markers. *ReCALL*, 18, 83-104. <https://doi.org/10.1017/S0958344006000619>
- Basturkmen, Helen; Shackelford, Nick (2015). How content lecturers help students with language: an observational study of language-related episodes in interaction in first year accounting classrooms. *English for Specific Purposes* 37, 87-97. <https://doi.org/10.1016/j.esp.2014.08.001>
- Batlle, Jaume (2021). "Muy bien" as a transition token in teacher-student interactions in the Spanish as a foreign language classroom. *System*, 97, 102438. <https://doi.org/10.1016/j.system.2020.102438>
- Bayés, Marc; Montmany, Begoña (2012). Instrucciones en el aula de ELE: Pasado, presente y futuro. En Beatriz Bleuca, Sara Borrel, Berta Corus y Fermín Sierra (Eds.), *Plurilingüismo y enseñanza de ELE en contextos multiculturales, Actas del XXXIII Congreso Internacional ASELE* (pp. 189-198). ASELE. https://cvc.cervantes.es/Ensenanza/biblioteca_ele/asele/pdf/23/23_0019.pdf
- Biber, Douglas (2006). *University language*. John Benjamins Publishing.
- Borg, Simon (2006). *Teacher cognition and language education: Research and practice*. Continuum.
- Dalton-Puffer, Christiane (2007). *Discourse in Content and Language Integrated Learning (CLIL) classrooms*. John Benjamins Publishing.
- Daidone, Danielle (2019). Preterite and imperfect in Spanish instructor oral input and Spanish language corpora. *Hispania*, 102(1), 45-58. <https://doi.org/10.1353/hpn.2019.0010>
- Dang, Thi Ngoc Yen (2020). The potential for learning specialized vocabulary of university lectures and seminars through watching discipline-related tv programs: Insights from medical corpora. *TESOL Quarterly*, 54(2), 436-459. <https://doi.org/10.1002/tesq.552>
- Díaz-Negrillo, Ana; Fernández-Domínguez, Jesús (2006). Error tagging systems for learner corpora. *Revista Española de Lingüística Aplicada*, 19, 83-102. <https://dialnet.unirioja.es/servlet/articulo?codigo=2198610>
- Donzelli, Giovanna (2007). Foreign language learners: words they hear and words they learn: A case study. *ELIA: Estudios de Lingüística Inglesa Aplicada*, 7, 103-125. <https://revistas.uned.es/index.php/ELIA/article/view/18092>
- Egbert, Jesse; Baker, Paul (Eds.) (2019). *Using corpus methods to triangulate linguistic analysis*. Routledge.
- Egbert, Jesse; Larsson, Tove; Biber, Douglas (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.
- Friginal, Eric; Lee, Joseph J.; Polat, Brittany; Roberson, Audrey (2017). *Exploring spoken English learner language using corpora: Learner talk*. Springer.
- González Argüello, María Vicenta (2011). La repetición en el discurso del profesor de ELE: formas y funciones. *ELUA: Estudios de Lingüística*, 25, 205-234. <https://doi.org/10.14198/ELUA2011.25.07>
- González Argüello, María Vicenta (2001). Modificaciones en el discurso del profesor de español como lengua extranjera. [Tesis de doctorado no publicada]. Universidad de Barcelona. <https://diposit.ub.edu/dspace/handle/2445/41411>



- Guerrettaz, Anne Marie; Engman, Mel M.; Matsumoto, Yumi (2021). Empirically defining language learning and teaching materials in use through sociomaterial perspectives. *The Modern Language Journal*, 105(S1), 3-20. <https://doi.org/10.1111/modl.12691>
- Gurzynski-Weiss, Laura; Geeslin, Kimberly L.; Long, Avizia Yim; Daidone, Danielle (2017). Linguistic variation in instructor provision of oral input. En Laura Gurzynski-Weiss (Ed.), *Expanding individual difference research in the interaction approach: Investigating learners, instructors, and other interlocutors* (pp. 225-254). John Benjamins Publishing.
- Hall, Joan Kelly; Looney, Stephen Daniel (2021). The role of self-talk in downgrading a teacher's certainty about grammar matters. *Tesol Quarterly*, 55(1), 185-218. <https://doi.org/10.1002/tesq.583>
- Helmke, Tuyet; Helmke, Andreas; Schrader, Friedrich-Wilhelm; Wagner, Wolfgang; Nold, Günter; Schröder, Konrad (2008). Die Videostudie des Englischunterrichts. En DESI-Konsortium (Eds.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (pp. 345-363). Beltz.
- Horst, Marlise (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language*, 22(1), 161-180. <https://doi.org/10.125/66646>
- Hu, Hsue-chao Marcella; Nassaji, Hossein (2012). Ease of inferencing, learner inferential strategies, and their relationship with the retention of word meanings inferred from context. *Canadian Modern Language Review*, 68(1), 54-77. <https://doi.org/10.3138/cmlr.68.1.054>
- Jablonkai, Reka R. (2021). Corpus linguistic methods in EMI research: A missed opportunity? En Jack K.H. Pun y Samantha M. Curle (Eds.), *Research methods in English Medium Instruction* (pp. 92-106). Routledge.
- Jäkel, Olaf (2010). *The Flensburg English Classroom Corpus (FLECC): Sammlung authentischer Unterrichtsgespräche aus dem aktuellen Englischunterricht auf verschiedenen Stufen an Grund-, Haupt-, Real- und Gesamtschulen Norddeutschlands*. Flensburg University Press.
- Kilgarriff, Adam (2009). Simple Maths for Keywords (CL2009). En Michaela Mahlberg, Victorina González-Díaz y Catherine Smith (Eds.), *Proceedings of the Corpus Linguistics Conference*. University of Liverpool. <http://ucrel.lancs.ac.uk/publications/cl2009/>
- Kilgarriff, Adam; Renau, Irene (2013). esTenTen, a Vast Web Corpus of Peninsular and American Spanish. *Procedia-Social and Behavioral Sciences*, 95, 12-19. <https://doi.org/10.1016/j.sbspro.2013.10.617>
- Kunioshi, Nilson; Noguchi, Judy; Tojo, Kazuko; Hayashi, Hiroko (2016). Supporting English-medium pedagogy through an online corpus of science and engineering lectures. *European Journal of Engineering Education*, 41(3), 293-303. <https://doi.org/10.1080/03043797.2015.1056104>
- Kurtz, Jürgen (2013). *The Dortmund Historical Corpus of Classroom English (DOHCCE)*. Flensburg University Press.
- Laufer, Batia (2023). Understanding L2-derived words in context: Is complete receptive morphological knowledge necessary? *Studies in Second Language Acquisition*, 1-14. <https://doi.org/10.1017/S0272263123000219>
- Lee, Joseph J. (2020). Spoken classroom discourse. En Eric Friginal y Jack A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 82-97). Routledge.
- Lin, Chia-Yen (2012). Modifiers in BASE and MICASE: A matter of academic cultures or lecturing styles? *English for Specific Purposes*, 31(2), 117-126. <https://doi.org/10.1016/j.esp.2011.08.003>
- Long, Avizia Yim; Geeslin, Kimberly L. (2020). Examining the role of instructor first language in classroom-based oral input. En Laura Gurzynski-Weiss (Ed.), *Cross-theoretical explorations of interlocutors and their individual differences* (pp. 127-157). John Benjamins Publishing.



- López Bastidas, Lani Guadalupe (2023). *Vocabulary use and classroom practices through teacher talk: a comparative and longitudinal study*. [Tesis Doctoral]. University of California, Davis. <https://www.proquest.com/docview/2866004924?pq-origsite=gscholar&fromopenview=true>
- Loewen, Shawn (2011). Focus on form. En Eli Heinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 576-592). Routledge.
- Lüdeling, Anke; Hirschmann, Hagen (2015). Error annotation systems. En Sylvianne Granger, Gättanelle Gilquin y Fanny Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 135-157). Cambridge University Press.
- Lyster, Roy; Saito, Kazuya; Sato, Masatoshi (2013). Oral corrective feedback in second language classrooms. *Language teaching*, 46(1), 1-40.
- MacWhinney, Brian (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Lawrence Erlbaum Associates.
- MacWhinney, Brian (2023). *Tools for analyzing talk part 1: The CHAT transcription format*. Carnegie Mellon University. <https://doi.org/10.21415/3mhn-0z89>
- Marcos Miguel, Nausica (2015). Textbook consumption in the classroom: Analyzing a classroom corpus. *Procedia-Social and Behavioral Sciences*, 198, 309-319. <https://doi.org/10.1016/j.sbspro.2015.07.449>
- Marcos Miguel, Nausica (2016). Voluntad y viabilidad en la enseñanza de morfología léxica en el aula de español como L2. En Santiago Alcoba, Cristina Buenafuente, y Gloria Clavería (Eds.), *Los lindes de la morfología* (Vol. 37, pp. 145-158). Universidade da Coruña, Servizo de Publicacións.
- Marcos Miguel, Nausica (2017). Instruction in derivational morphology in the Spanish L2 classroom: What do teachers believe and do? *Konin Language Studies*, 5(1), 37-60. <https://doi.org/10.30438/ksj.2017.5.1.2>
- Marcos Miguel, Nausica (2018). Analyzing the relationship and development of proficiency, derivational knowledge, and vocabulary size in Spanish L2 learners. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 31(1), 224-256. <https://doi.org/10.1075/resla.16027.mar>
- Marcos Miguel, Nausica (2021). Vocabulary instruction in an L2 culture course: The intersection of classroom discourse, materials, and instructor's choices in lexical focus-on-form episodes. *Classroom Discourse*, 25(1-2), 35-55. <https://doi.org/10.1080/19463014.2020.1867592>
- Mathieu, Corinne S.; Marcos Miguel, Nausica; Jakonen, Teppo. (2021). Introduction to the special issue Discourse at the intersection of language education and materiality. *Classroom Discourse*, 25(1-2), 1-14. <https://doi.org/10.1080/19463014.2020.1870151>
- Meara, Paul; Lightbown, Patsy; Halter, Randall H. (1997). Classrooms as lexical environments. *Language Teaching Research*, 1(1), 28-46. <https://doi.org/10.1177/136216889700100103>
- Nation, Paul I.S. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, Paul I.S.; Webb, Stuart (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- Newton, Jonathan (2020). Approaches to learning vocabulary inside the classroom. En Stuart Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 255-270). Routledge.
- Nicaise, Eric (2020). *Native and non-native teacher talk in the EFL classroom: A corpus-informed study*. Routledge.
- Paquot, Magali (2023). The Core Metadata Schema for L2 data: Collaborative efforts towards improved data findability, metadata quality and study comparability in L2 research. En "Corpus Linguistics and Applied Linguistics Research" charlas online del grupo de investigación *Lenguajes de especialidad, corpus lingüísticos y*



- lingüística inglesa aplicada a la ingeniería del conocimiento*, Universidad de Murcia.
<https://www.youtube.com/watch?v=jvnVT40qLcw>
- Paquot, Magali; Köning, Alexander; Stemle, Egon; Frey, Jennifer-Carmen (2023). Core Metadata Schema for Learner Corpora [Data set]. <https://doi.org/10.14428/DVN/4CDX3P>
- Park, Song Hee; Hepburn, Alexa (2022). The benefits of a Jeffersonian transcript. *Frontiers in Communication*, 7, 1-4.
<https://doi.org/10.3389/fcomm.2022.779434>
- Parodi, Giovanni (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 93-119. <http://dx.doi.org/10.4067/S0718-48832008000100006>
- Pierson, Herbert D. (1989). Using etymology in the classroom. *ELT journal*, 43(1), 57-63.
- Plonsky, Luke; Loewen, Shawn (2013). Focus on form and vocabulary acquisition in the Spanish L2 classroom. *Language, Interaction and Acquisition*, 4(1), 1-24. <https://doi.org/10.1075/lia.4.1.01plo>
- Reder, Stephen; Harris, Kathryn; Setzler, Kristen (2003). The multimedia adult ESL learner corpus. *TESOL Quarterly*, 37(3), 546-557. <https://doi.org/10.2307/3588405>
- Sánchez Gutiérrez, Claudia H. (2020). Morphological awareness in L2 Spanish: What is learned incidentally and what requires explicit teaching? *Revista Internacional de Lenguas Extranjeras= International Journal of Foreign Languages*, 13. <https://raco.cat/index.php/RILE/article/view/375592>
- Sánchez Rufat, Anna; Jiménez Calderón, Francisco (2018). Vocabulario (Vocabulary). En Javier Muñoz-Basols, Elisa Gironzetti y Manel Lacorte (Eds.), *The Routledge Handbook of Spanish Language Teaching* (pp. 229-242). Routledge.
- Sánchez Rufat, Anna; Jiménez Calderón, Francisco (2015). New perspectives on the acquisition and teaching of Spanish vocabulary/Nuevas perspectivas sobre la adquisición y la enseñanza del vocabulario del español, *Journal of Spanish Language Teaching*, 2(2), 99-111, <https://doi.org/10.1080/23247797.2015.1106123>
- Seedhouse, Paul (2004). *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Blackwell.
- Sert, Olcay (2015). *Social interaction and L2 classroom discourse*. Edinburgh University Press.
- Serrano-Dolader, David. (2019). *Formación de palabras y enseñanza del español LE/L2*. Routledge.
- Sinclair, Stéfan; Rockwell, Geoffrey (2020). Voyant-Tools. <http://voyant-tools.org/>
- Thoms, Joshua J. (2012). Classroom discourse in foreign language classrooms: A review of the literature. *Foreign Language Annals*, 45(s1), s8-s27. <https://doi.org/10.1111/j.1944-9720.2012.01177.x>
- Thompson, Paul (2010). Building a specialised audio-visual corpus. En Anne O’Keeffe y Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics*, 93-103.
- Trofimovich, Pavel; Collins, Laura; Cardoso, Waldir; White, Joanna; Horst, Marlise (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly*, 46(1), 176-186. <https://doi.org/10.1002/tesq.9>
- Vázquez, Graciela; Navas Méndez, Alejandra (2005). *ADIEU Plus. Clases Magistrales. (Derecho, Economía, Arqueología, Filología Española, Ciencias Sociales)*. Universidad de Potsdam.
<http://www.comunicacionacademica.com/clasesmagistrales.html>
- Walsh, Steve (2011). *Exploring classroom discourse: Language in action*. Taylor & Francis.



Corpus mencionados en inglés L1 y L2 (disponibles):

- *British Academic Spoken English Corpus* (BASE). https://www.reading.ac.uk/acadepts/ll/base_corpus/
- *Corpus of English for Academic and Professional Services* (CEAPP) (2014). Corpus of videos and accompanying transcripts from educational contexts. Unpublished raw data. <https://ceapp.la.psu.edu/>
- *Michigan Corpus of Academic Spoken English* (MICASE) (2019). <https://quod.lib.umich.edu/m/micase/>

