

# Recursos en línea para la consulta del léxico en la ESO: corpus textuales y diccionarios

Blanca Luján Molina

(MESOB) Especialidad de Lengua y Literatura



MÁSTERES  
DE LA UAM  
2021-2022

Facultad de Formación de Profesorado



**MÁSTER EN FORMACIÓN DEL PROFESORADO DE  
EDUCACIÓN SECUNDARIA OBLIGATORIA Y BACHILLERATO**

*Recursos en línea para la consulta del léxico en la ESO: corpus  
textuales y diccionarios*

Autora: Blanca Luján Molina

Director: Santiago Urbano Sánchez Jiménez

TRABAJO DE FIN DE MÁSTER

Curso académico: 2021/2022



## **AGRADECIMIENTOS**

A mi tutor, Santiago, por su acompañamiento y orientación durante el proceso de elaboración de este trabajo.

# ÍNDICE

<b>Resumen</b> .....	1
<b>1. Introducción</b> .....	2
1.1. Presentación y justificación de la propuesta.....	2
1.2. Currículo.....	3
1.3. Estructura del trabajo.....	6
<b>2. Fundamentación teórica: los corpus textuales y los diccionarios</b> .....	7
2.1. Los corpus textuales .....	7
2.1.1. <i>Definición y propiedades</i> .....	7
2.1.2. <i>Tipología de corpus</i> .....	10
2.1.3. Corpus del Español del Siglo XXI.....	12
2.1.4. <i>Instrucciones de uso del CORPES</i> .....	14
2.2. Los diccionarios.....	19
2.2.1. <i>Definición y propiedades</i> .....	19
2.2.2. <i>Tipología de diccionarios</i> .....	20
2.2.3. Diccionario de la lengua española.....	22
2.3. Conclusiones.....	28
<b>3. Revisión de propuestas de uso didáctico de los corpus</b> .....	29
<b>4. Propuesta didáctica</b> .....	32
4.1. Presentación y enfoque.....	32
4.2. Objetivos.....	33
4.3. Contenidos .....	34
4.4. Metodología.....	35
4.5. Recursos .....	37
4.6. Secuenciación .....	38

4.6.1.	<i>Primera sesión: El DLE, un vastísimo edificio (I)</i> .....	38
4.6.2.	<i>Segunda sesión: El DLE, un vastísimo edificio (II)</i> .....	40
4.6.3.	<i>Tercera sesión: El CORPES: palabras interconectadas</i> .....	41
4.6.4.	<i>Cuarta sesión: Consultas básicas en el CORPES</i> .....	42
4.6.5.	<i>Quinta y sexta sesión: En la piel de un lexicógrafo</i> .....	43
4.6.6.	<i>Séptima sesión: Reflexión sobre la propuesta</i> .....	44
4.7.	Evaluación .....	44
4.8.	Atención a la diversidad .....	45
<b>5.</b>	<b>Aplicación parcial de la propuesta didáctica</b> .....	<b>46</b>
<b>6.</b>	<b>Conclusiones</b> .....	<b>47</b>
<b>7.</b>	<b>Bibliografía</b> .....	<b>50</b>
<b>8.</b>	<b>Anexos</b> .....	<b>53</b>
8.1.	Anexo I. Material de apoyo a la explicación de los corpus .....	53
8.2.	Anexo II. Material de apoyo a la explicación del diccionario .....	57
8.3.	Anexo III. Materiales de la propuesta didáctica .....	58
8.4.	Anexo IV. Rúbrica de evaluación de la entrada de diccionario .....	62
8.5.	Anexo V. Resultados de la aplicación parcial de la propuesta .....	64

## ÍNDICE DE TABLAS

Tabla 1.	<i>Bloque 3: Conocimiento de la lengua. Primer ciclo de la ESO</i> .....	5
Tabla 2.	<i>Bloque 3: Conocimiento de la lengua. Segundo ciclo de la ESO</i> .....	5
Tabla 3.	<i>Bloque 3: Conocimiento de la lengua. 1º de Bachillerato</i> .....	6
Tabla 4.	<i>Estándares de aprendizaje evaluables relacionados con la propuesta</i> .....	33
Tabla 5	<i>Bloques de contenidos relacionados con la propuesta</i> .....	34

## **Resumen**

En el presente Trabajo de Fin de Máster (TFM) se pretende exponer la pertinencia de incorporar un recurso como los corpus textuales al aula de secundaria, especialmente, atendiendo a su empleo como herramienta de consulta lingüística en el ámbito del léxico. En este ámbito, complementados con un diccionario general, los corpus textuales permiten al alumnado ampliar su conocimiento acerca de las piezas léxicas del español, puesto que presentan las palabras en combinación con otras conformando estructuras sintácticas y enmarcadas en un determinado contexto de uso. Tras abordar el análisis de los corpus lingüísticos y los diccionarios como herramientas de consulta –con especial atención a los proporcionados por la RAE: el *CORPES* y el *DLE*–, resaltando sus beneficios y limitaciones, y proseguir con un apartado que revisa otras propuestas de aprovechamiento didáctico de los corpus textuales, se sugiere una propuesta didáctica dirigida al uso combinado de estas dos herramientas en el nivel de 4º de la ESO. Tras la presentación de esta propuesta, se hace una consideración de aquellas actividades que se aplicaron en el aula durante el periodo de prácticas de la titulación. Finalmente, en las conclusiones, se recogen las consideraciones más relevantes de este trabajo prestando atención a los resultados derivados de la aplicación parcial de la propuesta didáctica.

**Palabras clave:** herramientas lingüísticas, corpus textuales, diccionarios, léxico, enseñanza secundaria.

## 1. Introducción

### 1.1. Presentación y justificación de la propuesta

El manejo de herramientas de consulta lingüística es un conocimiento instrumental esencial para la asignatura de Lengua Castellana y Literatura. Estas herramientas lingüísticas, como los diccionarios o los corpus textuales, ayudan al estudiante en los procesos de codificación y decodificación de mensajes –así como en los procesos de manifestación de intenciones comunicativas y de interpretación de textos– en cualquier situación comunicativa a través de la información que proporcionan sobre la lengua y su uso. La relevancia del conocimiento y manejo de estas herramientas queda patente en el currículo, donde se alude a la utilidad de las fuentes de consulta para “resolver dudas con respecto al manejo de la lengua”, “enriquecer el propio vocabulario” y “progresar en el aprendizaje autónomo” (Decreto 48/2015, 2015, p. 85 y p. 93)<sup>1</sup>.

Si contrastamos los contenidos relacionados con el manejo de estas herramientas propuestos en el *Decreto 48/2015* con respecto al *Real Decreto 1105/2014* –contenidos que solo ofrecen una mayor concreción en 1º y 3º de la ESO–, observamos que la alusión a las “otras fuentes de consulta”, aparte de los diccionarios generales, se atiende exclusivamente a diccionarios o glosarios especializados (Decreto 48/2015, 2015, p. 74 y p. 80). De esta manera, se descuida el empleo de otras herramientas de gran utilidad como los corpus textuales, que, en comparación con los diccionarios, proporcionan una información más refinada sobre la frecuencia y dispersión de uso de las palabras según los distintos parámetros de variación lingüística (diatópica, diacrónica, diastrática o diafásica) (Rojo, 2021, p. 164). Además, aportan ejemplos reales y contextualizados del uso de estas en situaciones comunicativas concretas y en combinación con otras palabras.

La información añadida de los corpus con respecto al diccionario –una herramienta también imprescindible para el aprendizaje del vocabulario– tiene importantes implicaciones en lo que respecta al conocimiento del léxico y, en general, en lo que refiere al dominio de la lengua. En este sentido, los ejemplos reales del corpus permiten conocer el alcance del significado de las palabras en un contexto o acto

---

<sup>1</sup> En la situación actual de transición legislativa en materia de educación y a falta de la publicación del nuevo decreto curricular en el momento de redacción de este trabajo, nos basamos aquí en las directrices recogidas en el *Decreto 48/2015* y el *Real Decreto 1105/2014*.

comunicativo concreto; lo cual mejora el aprendizaje y el manejo de esas piezas léxicas (Martín, 2013, p. 4). Por otro lado, la información sobre las restricciones o dispersión de uso de las palabras puede permitir la producción de textos más adecuados al contexto o la descodificación –e interpretación– más oportuna considerando el contexto comunicativo; es decir, puede contribuir al desarrollo de la competencia comunicativa del alumnado. Por último, la representación de las relaciones con otras palabras, tanto sintácticas o sintagmáticas (los complementos verbales, régimen preposicional, etc.) como semánticas o paradigmáticas (hiperónimos, sinónimos, etc.), aparte de ajustarse a la realidad lingüística –puesto que las palabras nunca aparecen aisladas (Battaner, 2019, pp. 17-20)–, es lo que permite conocer las piezas léxicas en un sentido amplio y lo que potencia el desarrollo de la competencia léxica (Martín, 2013, p.1).

Adoptando una metáfora que sugiere la lectura de un pasaje de *Rayuela* de Julio Cortázar (2013, p. 259) –y simplificando, por ahora, la cuestión– los diccionarios, en ocasiones, dan la sensación de ser en realidad un “cementerio” de palabras o “la necrópolis de las palabras”. En ellos se almacenan palabras en nichos, aisladas y sin vida, pues están desconectadas de las situaciones concretas en que se usan y, en ocasiones, también desligadas del uso real que hacen los hablantes de ellas. Los corpus textuales, por el contrario, recogen discursos donde “viven las palabras”. En este caso, las palabras aparecen junto a otras, rodeadas de una atmósfera discursiva y contextual que favorece un conocimiento integral del léxico, ya que una palabra “no se aprende de forma aislada sino en su contexto y en la estructura lingüística de la que forma parte” (Ávila, 2010, p. 173). En buena medida, ha sido esta reflexión la que ha motivado la redacción de este trabajo acerca de la pertinencia de incorporar este recurso (los corpus textuales), junto al diccionario, en las aulas de secundaria.

## 1.2. Currículo

El empleo habitual de un recurso como los corpus lingüísticos de carácter general (combinados, por supuesto, con la consulta de los diccionarios) habría de ser un ejercicio más frecuente en las aulas, puesto que constituyen una buena herramienta para trabajar contenidos de los diversos bloques del currículo, con especial atención a los incluidos en los bloques de lengua, aunque sin descartar el de literatura. En efecto, dado que los corpus textuales de ámbito general –con el fin de representar una gran variedad de muestras de

la lengua– incluyen fragmentos de textos literarios adscritos a distintos géneros y a diferentes autores, permiten estudiar el estilo lingüístico de autores cuyos textos estén recogidos en el corpus o estudiar los rasgos lingüísticos de los distintos géneros literarios<sup>2</sup>.

Por otro lado, en lo que respecta al “Bloque 1: Comunicación oral”, se podría recurrir a un corpus para extraer textos orales y analizarlos atendiendo a los parámetros contextuales que determinan los usos lingüísticos (adscripción social de los hablantes, ámbito de comunicación, etc.) –siempre que el corpus esté codificado y su diseño contemple la inclusión de estos parámetros contextuales o metadatos, como veremos–. Con respecto al “Bloque 2: Comunicación escrita”, el corpus puede servir, por ejemplo, como banco de textos clasificados por tipología textual (narración, diálogo, texto periodístico, etc.), lo que permite analizar y comparar sus rasgos lingüísticos prototípicos. Las opciones se multiplican para el “Bloque 3: Conocimiento de la lengua”, pues la consulta de estos puede servir para el estudio, observación y análisis de las distintas categorías gramaticales “en relación con la intención comunicativa del texto en el que aparecen” (Real Decreto 1105/2014, 2014, p. 369); el estudio de las variedades lingüísticas (diatópicas, diastráticas, diacrónicas...), o el estudio de la frecuencia y ámbito de uso de las palabras.

A pesar de todas las posibilidades de uso de un corpus lingüístico general en las aulas para abarcar los contenidos curriculares de Lengua Castellana y Literatura, en este trabajo hemos preferido centrarnos en el manejo de los corpus como herramientas de consulta lingüística en el nivel léxico. Partimos de la idea de que esta herramienta puede servir para ampliar la información proporcionada por los diccionarios, dotando de “vida” a las palabras que se almacenan en entradas o “nichos” lexicográficos para favorecer que nuestro alumnado alcance un conocimiento del léxico más completo. Esta cuestión tiene cabida en el currículo tanto del primer ciclo de la ESO como del segundo. En concreto, como presentamos en la Tabla 1 y la Tabla 2, se relacionaría directamente con los siguientes aspectos del *Real Decreto 1105/2014* (Real Decreto 1105/2014, 2014, pp. 363 y 369):

---

<sup>2</sup> En concreto el *Corpus del Español del Siglo XXI* –en adelante *CORPES*–, en el que centraremos la propuesta, contiene fragmentos de relatos de Gabriel García Márquez, de novelas de Camilo José Cela o de obras de teatro de Juan Mayorga. Este trabajo se ha realizado siguiendo la versión beta 0.94 del *CORPES*.

**Tabla 1***Bloque 3: Conocimiento de la lengua. Primer ciclo de la ESO*

Contenidos	Criterios de evaluación	Estándares de aprendizaje evaluables
Manejo de diccionarios y <b>otras fuentes de consulta</b> en papel y formato digital sobre el uso de la lengua.	6. Usar de forma efectiva los diccionarios y <b>otras fuentes de consulta</b> , tanto en papel como en formato digital para resolver dudas en relación al manejo de la lengua y para enriquecer el propio vocabulario.	6.1. Utiliza <b>fuentes variadas de consulta</b> en formatos diversos para resolver sus dudas sobre el uso de la lengua y para ampliar su vocabulario.

*Nota:* Se destacan en negrita los aspectos más directamente relacionados con esta propuesta didáctica

**Tabla 2***Bloque 3: Conocimiento de la lengua. Segundo ciclo de la ESO*

Contenidos	Criterios de evaluación	Estándares de aprendizaje evaluables
Manejo de diccionarios y <b>otras fuentes de consulta</b> en papel y formato digital sobre la normativa y el uso no normativo de las palabras e interpretación de las informaciones lingüísticas que proporcionan los diccionarios de la Lengua: gramaticales, semánticas, registro y uso	5. Usar correcta y eficazmente los diccionarios y <b>otras fuentes de consulta</b> , tanto en papel como en formato digital para resolver dudas sobre el uso correcto de la lengua y para progresar en el aprendizaje autónomo.	5.1. Utiliza los diccionarios y <b>otras fuentes de consulta</b> en papel y formato digital resolviendo eficazmente sus dudas sobre el uso correcto de la lengua y progresando en el aprendizaje autónomo.

*Nota:* Se destacan en negrita los aspectos más directamente relacionados con esta propuesta didáctica

Asimismo, estos criterios de evaluación y estándares de aprendizaje evaluables quedan recogidos para 1º de Bachillerato en el *Real Decreto 1105/2014* (Real Decreto 1105/2014, 2014, p. 373), como presentamos en la Tabla 3:

**Tabla 3**

*Bloque 3: Conocimiento de la lengua. 1º de Bachillerato*

Criterios de evaluación	Estándares de aprendizaje evaluables
6. Conocer y manejar <b>fuentes de información impresa o digital</b> para resolver dudas sobre el uso correcto de la lengua y avanzar en el aprendizaje autónomo.	6.1. Conoce y consulta <b>fuentes de información impresa o digital</b> para resolver dudas sobre el uso correcto de la lengua y para avanzar en el aprendizaje autónomo.

*Nota:* Se destacan en negrita los aspectos más directamente relacionados con esta propuesta didáctica

Como puede comprobarse, el uso de los corpus textuales tendría cabida en el currículo de todos los cursos de la ESO y en 1º de Bachillerato; no obstante, se ha decidido orientar la propuesta a 4º de la ESO por tres motivos. El primero es que, tras haber superado el primer ciclo, el alumnado contará con el grado de madurez lingüística y bagaje académico suficientes para acometer y aprovechar al máximo la tarea final que se propondrá. El segundo motivo es que la combinación del diccionario con el corpus permite contrastar “el uso normativo y no normativo” (o descriptivo) de la lengua e interpretar así de un modo más riguroso las informaciones “gramaticales, semánticas, de registro y uso” proporcionadas en los diccionarios; cuestiones recogidas literalmente en el ya citado bloque de contenidos del *Real Decreto 1105/2014* (2014, p. 369) de ese curso. Por último, la incorporación guiada del manejo de esta herramienta en el segundo ciclo de la ESO puede servir para afianzar los contenidos ya asimilados en el primer ciclo y servir de transición al uso autónomo de estos recursos que se pretende en Bachillerato.

### 1.3. Estructura del trabajo

Una vez expuestas las bases, se abordarán los apartados de que consta este trabajo. A partir de una aproximación teórica al carácter de los corpus textuales y los diccionarios –ejemplificados en el *Diccionario de la lengua española* (en adelante, *DLE*) y el *CORPES*– fruto de una revisión bibliográfica, se pretende diseñar una propuesta didáctica dirigida al manejo e integración de ambos recursos en el marco de la asignatura de Lengua Castellana y Literatura para el nivel de 4º de la ESO. Antes de desarrollar esta propuesta,

se hará un breve repaso a otras aproximaciones didácticas que toman como referencia el uso de los corpus para la enseñanza de la lengua. Nuestra propuesta didáctica englobará los siguientes apartados: presentación y enfoque, objetivos, contenidos, metodología, recursos, secuenciación, evaluación y atención a la diversidad. Tras la presentación de la propuesta, se dará cuenta de aquellos aspectos que se pudieron aplicar en el aula. Concluiremos este Trabajo de Fin de Máster con unas reflexiones derivadas de esta aplicación y del conjunto del trabajo.

## **2. Fundamentación teórica: los corpus textuales y los diccionarios**

### 2.1. Los corpus textuales

#### *2.1.1. Definición y propiedades*

La definición de *corpus* se concreta de maneras muy dispares según sea el objetivo con el que haya sido diseñado y el uso que se haga de esta herramienta desde distintas disciplinas. Con respecto al ámbito de los estudios lingüísticos, o Lingüística de Corpus, consideramos especialmente acertada la siguiente definición:

Un corpus es un conjunto de (fragmentos de) textos, orales o escritos, producidos en condiciones naturales, seleccionados de modo que resulten conjuntamente representativos de una lengua o una variedad lingüística, en su totalidad o en alguno(s) de sus componentes, que se almacenan en formato electrónico y se codifican con la intención de que puedan ser analizados científicamente. (Rojo, 2014, p. 371 citado en Rojo, 2021, p.1)

Con el fin de aclarar y desarrollar el alcance de esta definición, se profundizará a continuación en algunas de las propiedades de los corpus señaladas. En primer lugar, uno de los rasgos que más interés suscita para este TFM es que los textos de un corpus han sido producidos en situaciones comunicativas auténticas con una finalidad concreta y no han sido “concebidos para ilustrar un determinado fenómeno lingüístico” (Rojo, 2021, p. 1). De esta manera, los corpus nos permiten aproximarnos a la lengua real que usan los hablantes de un idioma. Esto último no es baladí, pues en muchas ocasiones los libros de texto proporcionan ejemplos que resultan artificiales y que no se corresponden con el conocimiento que tienen los hablantes sobre la lengua (Tolchinsky, 2014, pp. 9-10).

Asimismo, la observación del comportamiento de determinadas palabras a partir de corpus permite contrastar la información que proporcionan los diccionarios sobre estas; cuestión en la que profundizaremos adelante.

En segundo lugar, cabe señalar que todos los corpus han de incluir un conjunto representativo de textos de una lengua o variedad lingüística, ya sea a lo largo de su historia o en un momento concreto (Rojo, 2021, p. 1). Para ello, una vez fijadas las variedades que se pretenden representar en un corpus (v.g. el español de América), la selección de textos para la muestra ha de incluir, de manera equilibrada, textos de distinta tipología: orales y escritos, de ficción (novelas, obras de teatro, etc.) y de no ficción (noticias, trabajos académicos...), de distintos países (Paraguay, Colombia...) y periodos temporales (Rojo, 2021, p. 2). Cabe resaltar, al respecto, que los términos *representativo* o *representar* no son azarosos, pues la realidad lingüística es demasiado compleja como para poder ser recogida en su totalidad en un corpus, por lo que hemos de aspirar solo a que garantice una cierta representatividad (Parodi, 2010, p. 24).

En tercer lugar, aunque no todas las definiciones que se ofrecen de *corpus* aludan a su formato electrónico, este soporte es en la actualidad un requisito indispensable por el aprovechamiento que permite de este recurso (Rojo, 2021, p. 2; Tolchinsky, 2014, p. 11). Así, el almacenamiento de textos en formato digital permite realizar búsquedas específicas a través de las aplicaciones de consulta, obtener una recuperación selectiva de la información, ordenar y clasificar los datos textuales según unos determinados parámetros o hacer cálculos estadísticos (Tolchinsky, 2014, p. 11). Por ejemplo, podríamos buscar y contabilizar el número de textos periodísticos que incluyen la palabra *guagua* (el *CORPES*, por ejemplo, arroja un total de 389 casos) y ordenarlos por fecha. Todo ello sería infinitamente más laborioso si esa información se hubiera de obtener a partir de fuentes analógicas.

La codificación de un corpus, cuestión que permite la recuperación selectiva de la información, implica integrar la información intratextual (párrafos, tipografías) y extratextual (país, año de producción, tipo de texto) básica de cada texto que incluye (Rojo, 2021, p. 69). Para ello, el texto se somete a un minucioso proceso de marcación o etiquetado que requiere que cada texto que se incluya esté bien documentado (país, fecha,

autor, etc.) y que esté correctamente digitalizado. Cumplidos estos dos requisitos, se hace uso de aplicaciones de programación o sistemas de codificación (en XML, HTML) que permitan asignar etiquetas o valores a distintos fragmentos del texto (Anexo I, imagen 1). Normalmente, ese etiquetado se realiza en dos niveles: en el nivel de cabecera (extratextual), para asignar valores a las etiquetas de tipo “autor”, “tipo de texto” –que constituyen los metadatos del texto–, y en el nivel de texto (intratextual), donde se marca la división en párrafos y las marcas de formato (negrita, cursiva, etc.) (RAE, 2013, pp. 9-11).

En relación con la tarea de codificación de los corpus, algunos corpus han pasado por un proceso de anotación lingüística y de lematización, lo cual amplía las posibilidades del análisis lingüístico (Parodi, 2010, p. 39; Rojo, 2021, p. 2). La anotación consiste en incorporar información lingüística de cada forma presente en el corpus (gramaticales, léxicas, semánticas, etc.) (Rojo, 2021, p. 2). Esto supone, por ejemplo, que el corpus permita diferenciar entre *bajo* como adjetivo (*es un chico bajo*), como sustantivo (*vive en un bajo*) o como verbo (*bajo a las tres*). La lematización, por su parte, implica vincular cada forma a un lema, que se considera el prototípico de un paradigma (Rojo, 2021, p. 30). Esto hace posible, por ejemplo, que el corpus identifique que *voy* y *yendo* son dos formas verbales de un mismo paradigma representado por el lema *ir* o que *casas* es la forma plural del lema *casa*.

Por último, antes de cerrar este apartado, conviene destacar que una propiedad de los corpus que no se explicita en la definición de corpus de la que hemos partido es que su configuración parte de un diseño predeterminado (Parodi, 2010, p. 25; Rojo, 2016, p. 285). Un corpus no es una simple agrupación aleatoria de textos, sino que, al construirse, se establecen unas condiciones con respecto a la época, la tipología y la procedencia de los textos que lo componen; de esta manera, se favorece el análisis de los fenómenos lingüísticos no solo de manera aislada, sino en relación con los parámetros contextuales contemplados en su codificación (tipo de texto, filiación geográfica, etc.) (Rojo, 2016, p. 285). Por ejemplo, esto permite consultar no solo la frecuencia de una variante no normativa como *estuvistes*, sino en qué tipo de textos o países es más frecuente. Si acudimos al *CORPES* para resolver esta cuestión, descubrimos que tres de los cuatro casos de *estuvistes* pertenecen a textos orales producidos en España y el cuarto se registra en un texto teatral (que recrea, por tanto, la oralidad) producido en México.

### 2.1.2. Tipología de corpus

En el apartado anterior se ha proporcionado una idea general de qué es un corpus desde la perspectiva de los estudios lingüísticos, aunque la tipología de los corpus es muy variada según el diseño preconfigurado y sus objetivos particulares. Con el objetivo de introducir posteriormente la particularidad del corpus que se ha escogido como punto de partida para este TFM, se hará un breve recorrido por las principales categorías de corpus existentes a partir de varios parámetros.

Uno de los principales parámetros de configuración de un corpus es el abanico de variedades lingüísticas que incluye. Desde un punto de vista geográfico, o diatópico, se distingue entre corpus dialectales, que se centran en la representación de una variedad lingüística (v.g. el *COSER: Corpus Oral y Sonoro del Español Rural*, que reúne grabaciones y transcripciones de entrevistas a hablantes naturales de distintos enclaves rurales de España) y corpus generales o de referencia, que tratan de representar una lengua en todo su dominio lingüístico (como el *CREA: Corpus de Referencia del Español Actual*). Desde un punto de vista temporal, o diacrónico, se distingue entre corpus diacrónicos y corpus sincrónicos. Los corpus diacrónicos son aquellos que incluyen textos de un lapso temporal superior a los 25 años, como el *CORDE: Corpus Diacrónico del Español*, que alberga documentos desde los orígenes del español hasta 1975. Los corpus sincrónicos recogen textos dentro de un espacio temporal inferior a los 25 años, como el *Corpus del Español*, de Mark Davies (Rojo, 2021, p. 24).

De otra parte, los corpus se pueden clasificar, según el medio en que se presentan sus textos, en corpus orales (p. ej. el *COLA*, un corpus oral del lenguaje de adolescentes) o escritos (como el *Corpus de Sonetos del Siglo de Oro*). Aparte de las diferencias lingüísticas en ambos tipos de texto (oralidad frente a escritura), esta distinción implica marcadas diferencias en el desarrollo y uso del corpus (Rojo, 2021, p. 24). Así, el acopio y la codificación de los textos orales es más costosa, puesto que requiere proporcionar las características personales de cada participante de la comunicación (edad, adscripción sociocultural, etc.) y encontrar un equilibrio entre transcribir exactamente lo pronunciado (con sus vacilaciones, frases inacabadas, etc.) y aproximarse al sistema ortográfico de la lengua para facilitar la posterior recuperación de la información (Rojo, 2021, p. 22 y p.

65). Por estas razones, estos corpus suelen ser de tamaño más reducido (Parodi y Burdiles, 2019, p. 600) o, en caso de combinarse ambos tipos de texto en un corpus, el porcentaje de textos orales es menor que el de escritos (Rojo, 2021, p. 24).

En tercer lugar, los corpus pueden diferenciarse según su temática o ámbito de comunicación. Así, los corpus pueden tener un carácter general –corpus generales o de referencia– o pueden estar restringidos a un uso específico de la lengua –corpus especializados– (Torruella, 2017, p. 44). Un ejemplo de corpus especializado sería el *Corpus Científico-Técnico* de la RAE, dirigido al estudio y representación del lenguaje científico. Los corpus generales, por su parte, no se restringen a ningún dominio específico, por lo que permiten el estudio de una lengua en todas sus variedades (Rojo, 2021, p. 25; Torruella, 2017, p. 41), como el *CREA*.

Según si su diseño parte de un número de textos predefinido para su inventario o no, se distingue entre corpus abiertos o cerrados. Los corpus abiertos presentan un carácter dinámico, pues incorporan periódicamente nuevos textos y no se ciñen a un límite preestablecido, como el *Corpus del español* de Mark Davies (Torruella, 2017, p. 48). Los corpus cerrados, por el contrario, cuentan con un límite máximo de textos definido en la fase de diseño y cuando este límite se alcanza se da por finalizada la creación del corpus (Torruella, 2017, p. 48), como podría ser el *Corpus estadístico del euskera del siglo XX*. Los corpus abiertos reportan el beneficio añadido de posibilitar el estudio de los fenómenos lingüísticos más actuales (por ejemplo, las palabras de reciente incorporación como *limerencia* o *creepy*), mientras que los corpus cerrados destacan por aportar una mayor estabilidad en los resultados de las búsquedas (Rojo, 2021, p. 25).

Por último, interesa incidir en la diferenciación entre los corpus que han sido sometidos a algún proceso de codificación o anotación y los corpus simples. Los primeros incluyen un grado de información (extratextual o lingüística) sobre los textos más refinado y es esta condición lo que permite el estudio riguroso de la lengua que representa (Parodi, 2010, pp. 39-41; Rojo, 2021, p. 2). Este tipo de corpus permite la realización de búsquedas más específicas y análisis lingüísticos más profundos; por ejemplo, como señalamos en § 2.1.1., la dispersión de uso por área lingüística o tipología textual de una variante verbal no normativa como *estuvistes*. Los corpus simples, que no están

codificados ni anotados, solo permitirían la búsqueda de determinadas formas o secuencias lingüísticas.

### 2.1.3. Corpus del Español del Siglo XXI

La construcción del corpus escogido para esta propuesta, el *CORPES*, se encargó a la Real Academia Española tras el XIII Congreso de la Asociación de Academias de la Lengua Española en 2007 (RAE, 2013, p. 6). Con este proyecto, la RAE pretendía mejorar y ampliar sus bancos de datos anteriores, como el *CREA*, finalizado en 2013 y que, por aquel entonces, no estaba ni anotado ni lematizado (Rojo, 2016, pp. 290-291). Así, en el diseño de la primera fase del *CORPES*, se aprobó que este habría de incluir 25 millones de formas lematizadas y anotadas morfosintácticamente para cada año, desde 2001 a 2012, equilibrando textos de diversas tipologías y de todos los países hispanohablantes (RAE, 2013, p. 6). Siguiendo la clasificación propuesta en el apartado anterior, el *CORPES* es un corpus de referencia, sincrónico, que incluye textos orales y escritos, semiabierto, codificado, lematizado y anotado morfosintácticamente. A continuación, se dará cuenta de cada una de estas características.

En primer lugar, el *CORPES* es un corpus de referencia, ya que pretende representar todos los dominios del español en los ámbitos diatópico (geográfico) y diafásico (de situación comunicativa). En cuanto a la distribución geográfica, el 70 % se corresponde con textos producidos en América y el 30 % con textos producidos en España; porcentajes que luego se distribuyen por áreas lingüísticas (en el caso del español americano, área rioplatense, chilena...) y países (RAE, 2013, p. 7). En cuanto a las tipologías de los textos, este corpus atiende a una gran diversidad textual combinando varios parámetros situacionales, como el medio (lengua oral o escrita), soporte (libro, publicación periódica, etc.), bloque temático (ficción o no ficción), área temática (salud; actualidad, ocio y vida cotidiana...) o género discursivo (novela, guion, conversaciones, entrevistas...) (RAE, 2013, pp. 7-9).

En segundo lugar, se trata de un corpus que recoge tanto textos escritos como orales. No obstante, como se ha mencionado, dado el coste que supone la inclusión de los textos orales, el 90 % de los textos corresponde a la lengua escrita y solo el 10 % a la lengua oral (RAE, 2013, p. 7). En cuanto a las dificultades asociadas a la transcripción, en el

*CORPES* se opta por adoptar como pauta de transcripción el sistema ortográfico del español para facilitar la posterior recuperación de la información, aunque se señalan algunas informaciones de naturaleza paralingüística como “risas”, “silencio”, “vacilación”, etc. (RAE, 2018, p. 1). Además, ofrece algunos audios con o sin alineación a su transcripción o la posibilidad de descargar el audio o vídeo del que se extrae el texto oral para analizar, aparte de los rasgos lingüísticos de la comunicación oral, los aspectos paralingüísticos (entonaciones, etc.) y no lingüísticos (gestos) propios de esta.

En tercer lugar, se trata de un corpus semiabierto, pues combina las características de los corpus abiertos y cerrados. Es un corpus cerrado en cuanto que, de acuerdo con su diseño, se ciñe a un número límite de formas para cada año y una distribución fija de textos por país, año, tipología, etc.; sin embargo, se asemeja más a los corpus abiertos en tanto que va incorporando nuevos textos en sus sucesivas versiones (Rojo, 2021, p. 73). De esta manera, pese a mantener un diseño fijado y predefinido, permite el estudio de los fenómenos más actuales de la lengua española. A propósito de su dimensión temporal, se justifica su caracterización como corpus sincrónico, puesto que registra textos producidos desde 2001 hasta 2021, un rango de años inferior a 25.

En cuanto a su codificación, el *CORPES* incluye diversos metadatos o datos extratextuales para cada uno de los textos que lo componen. Estos datos son autor, título, país, área geográfica, año, soporte, medio, bloque de ficción o no ficción, género textual, área temática (Anexo I, imagen 4) y, exclusivamente para los textos orales, las características sociolingüísticas de los hablantes (género, edad, adscripción sociocultural, procedencia) (Rojo, 2021, p. 98). Asimismo, en el nivel intratextual, también se codifican en este corpus la división en párrafos y los formatos tipográficos habituales (cursivas, negritas, subrayados...) con el fin de que el usuario pueda consultar una versión digital lo más cercana posible al texto original (RAE, 2013, p. 10).

Por último, el *CORPES* cuenta con una lematización y anotación morfosintáctica. En efecto, este corpus, tras un proceso de análisis lingüístico automático, permite identificar cada forma ortográfica con la palabra prototípica del paradigma formal al que pertenece (lema) y explicita sus valores morfosintácticos: su categoría gramatical y los valores de las subcategorías que le corresponden (v.g. persona, número, tiempo y modo para verbos o género y número para sustantivos) (RAE, 2021a, p. 1). Este proceso de lematización y

anotación automática arroja un porcentaje de acierto del 97-98 % (RAE, 2021a, p. 1), lo cual, para perfeccionarse, requeriría de una revisión o etiquetado manual. Un aspecto que no incluye, a diferencia de otros corpus, es la anotación semántica: los lemas no aparecen vinculados a sus posibles acepciones en un diccionario, por eso, si queremos investigar los significados de una palabra, ha de realizarse una investigación cualitativa, analizando cada uno de los casos (Rojo, 2021, p. 17).

#### 2.1.4. Instrucciones de uso del CORPES

A diferencia de otros más antiguos, el *CORPES* es un corpus abierto a la consulta externa (<https://bit.ly/3Nb4zWW>) a través de internet desde cualquier parte del mundo y a partir de los navegadores de uso habitual (*Google Chrome, Mozilla Firefox, etc.*) (RAE, s.f.-a, p. 1). Esto lo convierte en un recurso muy accesible en las aulas de secundaria. Además, sus posibilidades de uso son muy amplias, como veremos a continuación siguiendo el *Manual de consulta en línea* (<https://bit.ly/3qpiQ9F>) de este<sup>3</sup>. Para facilitar la comprensión del funcionamiento de este recurso se atenderá, por separado, a tres tipos de búsqueda o ventanas: Concordancias, Estadística y Coapariciones.

##### *Concordancias*

En la ventana inicial de consulta (o Concordancias) se distinguen tres zonas: superior, intermedia e inferior. A grandes rasgos, en la zona superior aparecen los criterios de consulta o los parámetros de codificación o anotación ya citados (lema, clase de palabra, tipo de texto...), que permiten al usuario afinar su búsqueda (Anexo I, imagen 2). En la zona intermedia aparecen los ejemplos recuperados para la consulta del usuario, en los que la forma buscada aparece en una posición central (con un hipervínculo), rodeada de un mínimo contexto (Anexo I, imagen 3). Por último, en la zona inferior, tras pulsar el hipervínculo de la forma buscada, se abre una ventana que proporciona una ampliación del contexto de esta (aunque limitado por los derechos de propiedad de los textos), así como los datos bibliográficos y de codificación del texto incluidos en el *CORPES* (autor, medio, soporte, etc.) (Anexo I, imagen 4) (RAE, s.f.-a, pp. 3-5).

---

<sup>3</sup> Para acceder a los documentos alojados en la pestaña “Ayuda” de la web del *CORPES* se ha de pulsar dos veces sobre el enlace: una primera vez para ser dirigidos a la aplicación y, desde ahí, pulsar de nuevo en el enlace para acceder al documento.

Si recorremos la zona superior de esta ventana inicial de izquierda a derecha, interesa destacar, en primer lugar, la diferencia entre la búsqueda de una palabra desde la casilla “lema” y la casilla “forma”. Escribiendo en la casilla “lema” la forma paradigmática de una palabra (infinitivo en verbos, masculino singular en adjetivos de doble terminación, etc.), la aplicación devuelve los resultados de todas las formas o variantes asociadas al lema (si buscamos *ir*, obtendremos también *voy*, *fuimos*, etc.). En cambio, para obtener una variante en concreto (*fuimos*), ha de escribirse esta en la casilla “forma”. En ambas casillas se pueden buscar secuencias de hasta cinco palabras. También se pueden combinar estas dos casillas para afinar la búsqueda; por ejemplo, se puede escribir *ir* en “lema” y *fuimos* en “forma” para evitar la devolución de los casos de *fuimos* del verbo *ser* (RAE, s.f.-a, pp. 5-6).

Junto a las casillas de lema y forma, aparece a la derecha la casilla “clase de palabra”, que permite una búsqueda a partir de criterios lingüísticos aprovechando la anotación morfosintáctica de que dispone el corpus. Tras introducir cualquier valor en esta casilla, se despliega automáticamente una ventana con distintas listas de opciones para todos los valores morfosintácticos que le son de aplicación a esa categoría gramatical. Por ejemplo, si introducimos un verbo como *ir*, se abrirá una ventana en la que podremos precisar el modo, el tiempo, la persona y el número (Anexo I, imagen 5). Esto permite resolver ambigüedades léxicas (*bajo* como adjetivo, sustantivo o verbo) y también la consulta de ejemplos de categorías gramaticales sin emplear un lema o una forma concreta (RAE, s.f.-a, p. 7 y p. 14) como “interjecciones” (Anexo I, imagen 6).

Una posición a la derecha de la casilla “clase de palabra” aparece una opción denominada “grafía original”. Seleccionando esta casilla la aplicación devuelve los resultados de la palabra tal y como se ha escrito, lo que permite hacer búsquedas diferenciando entre la presencia o ausencia de tildes y la alternancia de mayúsculas o minúsculas (RAE, s.f.-a, p. 6). Por ejemplo, permite contabilizar los casos de dos variantes ortográficas como *período* (27.850) y *periodo* (23.566), o los casos de *Estado* como nombre de institución (105.651) frente a *estado* como participio verbal del tipo *he estado* (105.057), en caso de que estuvieran escritos de acuerdo con la norma en los textos reales (Anexo I, imagen 7).

A continuación de la “grafía original” encontramos otro recurso de gran utilidad, la opción “proximidad”. Se trata de un submenú que permite consultar varias formas que estén situadas en proximidad en un texto (a menos de diez palabras de distancia), para lo cual se puede precisar la posición (izquierda, derecha) y la distancia –exacta o en un intervalo– de los lemas secundarios con respecto al lema principal de búsqueda (RAE, s.f.-a, p. 8). Esto resulta especialmente útil para analizar unidades pluriverbales del español con grados de fijación distintos. Por ejemplo, se podrían buscar ejemplos de la construcción *dar un paseo* estableciendo un intervalo de 3-4 posiciones a la derecha de *paseo* frente a *dar*, para constatar que su grado de fijación no es completo, pues permite la intercalación de otros elementos. Por último, desde este submenú se pueden acometer búsquedas de proximidad estableciendo como lema o forma secundaria una categoría de palabras y no una forma concreta. Esto es interesante para analizar la productividad de construcciones sintácticas del tipo *marcarse un + sustantivo*<sup>4</sup> (Anexo I, imagen 8). Estas búsquedas, con sus matizaciones, no podrían ser efectuadas escribiendo en forma o lema simplemente “dar un paseo” o “marcarse un”.

Por último, junto a la opción “proximidad” encontramos otro submenú llamado “subcorpus”, el cual entronca con todos los aspectos de la codificación mencionados en los apartados anteriores. Este submenú permite la recuperación selectiva de la información del usuario a partir de la clasificación de los textos que establece el *CORPES* con los criterios ya citados: país, año o periodo, autor, medio, área temática, etc. A través de esta casilla, el usuario puede buscar, por ejemplo, los casos del verbo *amilanar* en la prensa rioplatense desde 2010 a la actualidad (Anexo I, imagen 9).

En cuanto a la zona intermedia de esta ventana inicial, cabe señalar dos aspectos. El primero es que situando el cursor del ratón encima de cualquiera de las palabras que aparecen en los ejemplos, la aplicación abre una ventana que proporciona su información morfosintáctica: su categoría y subcategorías gramaticales y el lema al que se vinculan (Anexo I, imagen 10) (RAE, s.f.-a, p. 4). El segundo aspecto es que en la esquina superior derecha de la tabla se abre una opción para ordenar los resultados según varios criterios: año, país, categoría gramatical, etc. (Anexo I, imagen 11) (RAE, s.f.-a, p. 4). La aplicación permite realizar una ordenación con la combinación de dos de estos criterios y, en caso

---

<sup>4</sup> Portolés, J. (2017) dedica un interesante artículo al estudio de la construcción ‘marcarse un + nombre propio’ en la revista *Círculo de Lingüística Aplicada a la Comunicación* ( <https://bit.ly/3MPZv7N>).

de no seleccionar ninguna opción, los ejemplos se ordenan, por defecto, en orden cronológico (RAE, s.f.-a, p. 10).

La aplicación ofrece, también en esta zona intermedia, la opción de realizar búsquedas complejas mediante operadores lógicos (Y, O, NO). Por ejemplo, si quisiéramos estudiar el significado de un verbo como *legitimar*, también presente en su variante *legitimizar*, podríamos buscar todos los casos del lema *legitimar* sumando (O) los casos de *legitimar* y observar los resultados arrojados para ambos al mismo tiempo. No obstante, se considera aquí que esta opción excede la complejidad y finalidad para la cual se pretende llevar este recurso al aula. Por ese motivo, se invita a consultar el *Manual de consulta en línea* (RAE, s.f.-a, pp. 11-13) para comprender su funcionamiento y utilidad. Del mismo modo, el uso de los comodines “?” y “\*” en las casillas de lema y forma, que permiten búsquedas más abstractas y complejas (escribiendo *\*ción* en la casilla de forma, obtendríamos todas las palabras que acaban con ese sufijo), supera los objetivos trazados para este TFM, por lo que se omite su tratamiento. En cualquier caso, en niveles superiores, y siempre con la orientación del docente, se podrían plantear actividades haciendo uso de estas opciones.

### *Estadística*

En la primera línea de opciones de la aplicación, cuatro casillas a la derecha de la opción “Concordancias”, se encuentra el acceso a la ventana “Estadística”. En la parte superior de esta ventana, tras cada consulta, se le proporcionan al usuario tres valores: la frecuencia absoluta o total de casos de la búsqueda en el corpus, el número de documentos en los que aparece y la frecuencia normalizada o casos por cada millón de palabras (RAE, s.f.-a, pp. 15-16). Además de estos datos, se ofrecen gráficos de sectores y tablas de frecuencia por zona geográfica (andina, España), periodo temporal, área temática (salud, política) y tipología textual (crónica, blog, etc.) (Anexo I, imagen 12) (RAE, s.f.-a, p. 16).

### *Coapariciones*

Una casilla a la derecha de “Concordancias” en la primera línea de opciones de la aplicación de consulta, se encuentra el acceso a la ventana “Coapariciones”. “Las coapariciones son las palabras que se combinan con el lema buscado con una frecuencia

mayor a la que sería esperable por azar” (RAE, s.f.-a, p. 17). De esta manera, esta ventana de consulta permite una aproximación al estudio del comportamiento combinatorio de las palabras, lo cual podría ser útil para estudiar las asociaciones semánticas de la palabra buscada con otras (hiperónimos, sinónimos, campos semánticos...) y las combinaciones habituales de esta: sus colocaciones<sup>5</sup>, las expresiones o unidades pluriverbales<sup>6</sup> en las que participa y sus relaciones sintácticas habituales (las preposiciones que se unen a un verbo o sustantivo, los tipos de complemento que acepta una determinada palabra: sintagmas preposicionales, nominales...) (Anexo I, imagen 13).

No obstante, hemos de apuntar que la asociación estadística en la que se basa el *CORPES* para identificar las coapariciones está claramente condicionada por los textos que se hayan incluido y a menudo devuelve interrelaciones que, desde una perspectiva lexicográfica o lingüística, no podríamos considerar correctas. Esto es especialmente notable al intentar detectar las unidades pluriverbales en las que participa una palabra, pues esto responde a criterios lingüísticos muy bien definidos y difícilmente traducibles a un sistema matemático; por este motivo, sería preferible acudir a un diccionario para resolver esta cuestión. Sin embargo, la aplicación sí puede darnos pistas sobre las asociaciones semánticas –y también frecuentes– de una palabra (la palabra *churro* se relaciona con *buñuelo* y *chocolate* en la aplicación), su comportamiento sintáctico (en el caso de los verbos, la preposición con la que se construyen, como *pensar en*; si va acompañado o no de un complemento directo, etc.) o sus posibles colocaciones (adjetivos y sustantivos vinculados, como *pensamiento rumiativo*; un verbo y un adverbio, como *vivir austeramente*), aunque todas estas cuestiones, a partir de los datos, habrían de analizarse de manera cualitativa.

Del mismo modo que ocurre en la ventana de las concordancias, este sistema proporciona la opción de establecer filtros sobre los resultados a partir de algunos parámetros de codificación o anotación del *CORPES* (clase de palabra del lema, país, área

---

<sup>5</sup> Una colocación es una combinación restringida de palabras formada por una base (elegida por su significado) y un “colocativo” (palabra determinada por la base, con la cual guarda una estrecha relación semántica), cuyo significado corresponde a la suma de sus partes. Por ejemplo, dos verbos con el significado de ‘incumplir’ participan en sendas colocaciones: *violar* (colocativo) *la ley* (base) o *romper* (colocativo) *una promesa* (base) (Alonso-Ramos, 2017, p. 175).

<sup>6</sup> Una unidad pluriverbal (o forma compleja) es una combinación bastante estable de dos o más palabras, que funciona como una unidad sintáctica (se inscribe toda ella en una misma categoría gramatical y tiene un cierto grado de fijación) y cuyo significado no se deduce de la suma de significados de sus partes (v.g. *lavarse las manos* en el sentido de ‘desentenderse’).

temática). Además, permite ordenar los resultados según la clase de palabra de los lemas con los que se combina la palabra pivote, la frecuencia de aparición de la combinación o la medida de asociación de las palabras combinadas según distintos índices estadísticos, como el *Mutual Information (MI)*, que mide la potencia de la asociación entre dos palabras, o el condicionamiento que ejerce una sobre otra (RAE, s.f.-a, p. 17).

## 2.2. Los diccionarios

### 2.2.1. *Definición y propiedades*

Los diccionarios, del mismo modo que los corpus, destacan por su heterogeneidad de fines, formas, temáticas, etc., lo cual dificulta establecer una definición única de estos. En nuestro caso, asumiremos la definición de *diccionario* propuesta por el *DLE* (RAE, 2014a, definición 1): ‘repertorio en forma de libro o en soporte electrónico en el que se recogen, según un orden determinado, las palabras o expresiones de una o más lenguas, o de una materia concreta, acompañadas de su definición, equivalencia o explicación’. A esta definición se puede añadir que un diccionario es una obra de consulta, con finalidad práctica y didáctica, que permite resolver cuestiones relacionadas con el léxico de una lengua (Campos y Pérez, 2003, p. 57). Una cuestión que permite conocer en profundidad el carácter de un diccionario es atender a los conceptos de macroestructura y microestructura que lo organizan. Resulta necesario detenernos ahora en estos dos conceptos.

La macroestructura de un diccionario, a grandes rasgos, haría referencia a la sucesión de entradas o lemas que recoge. En la práctica, se identifica con el conjunto de esos lemas, llamado también nomenclatura o lemario. La entrada de una palabra es el apartado que le corresponde a cada pieza léxica en un diccionario, donde se proporcionan distintas informaciones sobre esta. El lema, por su parte, designa la palabra de la que se proporcionan definiciones, informaciones gramaticales, de uso, etc., la cual aparece de forma canónica, prototípica, englobando todas sus variantes. Por último, el artículo lexicográfico hace referencia al conjunto formado por el lema y las informaciones que se proporcionan de este (cuerpo del artículo) (Castillo, 2003, p. 57) (Anexo II, imagen 1).

La microestructura, por otro lado, refiere a la organización interna de los elementos que componen cada artículo lexicográfico. Según sea el propósito con que ha sido

diseñado el diccionario, el artículo puede estar compuesto por elementos diversos. Así, se puede incluir para cada lema información sobre su ortografía o pronunciación, su etimología, su categoría y subcategoría gramatical, sus irregularidades morfológicas, las combinaciones de palabras en las que participa, sus sinónimos o sus antónimos, sus características sintácticas (restricciones de combinación, régimen preposicional), las restricciones de uso (diatópicas, diafásicas, diastráticas, diatécnicas o diacrónicas) y las definiciones o explicaciones de sus distintas acepciones junto a algunos ejemplos (Garriga, 2003, p. 105).

### 2.2.2. Tipología de diccionarios

Si bien en el subapartado anterior se facilita una aproximación general al concepto de diccionario, como mencionábamos, su concreción deriva en resultados muy distintos según el propósito con el que haya sido creado, el público al que va dirigido etc. Son muchas las tipologías de diccionarios propuestas; tanto es así que se ha llegado a plantear una tipología de tipologías de diccionarios (Campos y Pérez, 2003, p. 62). Sin pretender ofrecer más que una aproximación a esta cuestión, en este trabajo nos basaremos en la tipología de diccionarios del español propuesta por Haensch y Omeñaca (2004), aunque reducida solo a aquellos criterios de clasificación que sean de interés para explicar posteriormente las particularidades del *DLE* en el marco de este TFM.

Uno de los criterios básicos que no podemos obviar es el del número de lenguas cuyo léxico registre el diccionario. Según este criterio se distingue entre diccionarios monolingües –como el *Diccionario de uso del español* (1966-1967), *DUE*, de María Moliner– o plurilingües (aquellos que incluyen varias lenguas, como podría ser un diccionario escolar de español-inglés). Otro criterio fundamental para la clasificación de diccionarios alude a la ordenación de los materiales. Los diccionarios que se ordenan a partir del significante (o que parten del significante al significado), generalmente por orden alfabético, son los diccionarios semasiológicos (Haensch y Omeñaca, 2004, p. 64). En este tipo de diccionarios, como el *DLE*, buscaríamos una palabra concreta y obtendríamos su significado. Los diccionarios que se ordenan a partir de significados o campos léxicos, que agrupan las palabras por su afinidad semántica o temáticas, son onomasiológicos; es decir, parten del significado al significante (Haensch y Omeñaca, 2004, p. 64). En estos repertorios lexicográficos, como el *Diccionario ideológico de la*

*lengua española* (1942, 1959) de Julio Casares, buscaríamos, por ejemplo, el campo semántico de los sentimientos y, a partir de este, accederíamos a la palabra concreta que necesitábamos: *admiración*, *devoción*, etc.

Por otro lado, un criterio de clasificación muy relevante para este trabajo es el que distingue entre diccionarios normativos y diccionarios descriptivos. Los diccionarios normativos pretenden mostrar un modelo ideal del léxico de una lengua, fijan la ortografía de las formas, las acepciones admitidas y son estrictos con la inclusión de vulgarismos, neologismos, extranjerismos, etc. (Campos y Pérez, 2003, p. 69). Por el contrario, los diccionarios descriptivos pretenden representar la realidad lingüística tal y como se presenta en un momento determinado, sin rechazar los coloquialismos, neologismos, extranjerismos, etc. siempre que respondan a un uso extendido entre los hablantes (Haensch y Omeñaca, 2004, p. 62).

El ejemplo paradigmático de los diccionarios normativos sería el *Diccionario de autoridades*, publicado por la RAE entre 1726 y 1739 con el fin de estabilizar la lengua española y “depurarla” de los malos usos del idioma (Ruhstaller, 2003, pp. 238-239). Con tal fin, aparte de las definiciones de las palabras, aportaba citas reales de autores que se consideraban modelos o autoridades en el uso de la lengua (RAE, s.f.-b). Aun así, este diccionario admitió la entrada de varios tecnicismos y vocablos populares o dialectales (Haensch y Omeñaca, 2004, p. 203). Un ejemplo de los descriptivos sería el *Diccionario del español actual* (1999, 2011), *DEA*, de Manuel Seco. Este último se distingue por construir su lecionario a partir de un corpus de autores españoles y por acompañar de ejemplos reales todas las acepciones que propone<sup>7</sup>. Se considera que este repertorio marcó un hito en la lexicografía española, ya que, a diferencia de otros diccionarios generales, permitía conocer la realidad lingüística del español en todas sus variedades: jergas, coloquialismos, dialectalismos, neologismos... (Haensch y Omeñaca, 2004, p. 62).

El cuarto criterio que emplearemos, por su afinidad con el tema de este TFM, es el del sistema lingüístico (o conjunto de materiales) que toman como base los diccionarios para su confección. Según este criterio, en primer lugar, encontramos los diccionarios

---

<sup>7</sup> La primera acepción de *mutualizar* en el *DEA* es: “tr 1. dar carácter de mutualidad [a algo (cd)] || Act 7.7.66, 5: El ministro ... dijo recientemente que no se trataba de nacionalizar, sino de mutualizar el “Seguro de Accidentes de Trabajo” (Seco, Andrés y Ramos, 1999, definición 1).

basados en corpus textuales, como el *DEA*. Por otro lado, encontramos los diccionarios que aprovechan los materiales de otros diccionarios, añadiendo, modificando o suprimiendo entradas (nuevas acepciones, palabras en desuso, cambios de registro de las palabras, etc.) (Haensch y Omeñaca, 2004, p. 63), como sucede con el *DLE* con respecto al *Diccionario de Autoridades*, pero también con numerosos diccionarios no académicos con respecto al *DLE* (Ruhstaller, 2003, p. 237).

En último lugar, según el carácter del léxico que registren, distinguimos entre diccionarios generales o parciales. Los diccionarios generales registran un léxico muy variado; tratan de representar el léxico estándar o común de una lengua, usado por la mayor parte de los hablantes, e incluyen una selección representativa de palabras pertenecientes a algunos subconjuntos léxicos –tecnicismos, regionalismos, etc.– y a distintos niveles lingüísticos –coloquial, culto, etc.– (Haensch y Omeñaca, 2005, p. 188). Los diccionarios parciales, por su parte, pretenden representar un subconjunto específico del léxico de un idioma, subconjunto que suele establecerse atendiendo a alguno de los distintos ejes de variación de las lenguas (Campos y Pérez, 2003, p. 66): diatópica, como el *Diccionario de americanismos* (2010) de la ASALE<sup>8</sup>; diacrónica, como el *Diccionario medieval español* (1986) de Martín Alonso; diastrática y/o diafásica (jergas, tabúes, etc.), como el *Diccionario de expresiones malsonantes del español* (1974) de Jaime Martín o diatécnica, como el *Diccionario del español jurídico* (2016) de la RAE.

### 2.2.3. Diccionario de la lengua española

El *Diccionario de la lengua española*, confeccionado por la Real Academia Española, cuenta hoy con veintitrés ediciones. La primera edición de este repertorio lexicográfico se publicó en 1780 bajo el nombre de *Diccionario de la lengua castellana reducido a un tomo para su más fácil uso*, repertorio con el que se pretendía ofrecer una nueva versión del primer diccionario de la institución, el *Diccionario de autoridades* (Ruhstaller, 2003, p. 250). Con esta nueva versión, ya reducida, puesto que prescindía de las citas de los autores, aunque mantenía casi íntegra su nomenclatura, se inició la serie de diccionarios

---

<sup>8</sup> La Asociación de Academias de la Lengua Española (ASALE), fundada en 1951, reúne a las veintiuna academias de los correspondientes países de habla hispana y tiene como fin “trabajar en pro de la unidad, integridad y crecimiento del idioma común mediante el intercambio de información, colaboración y envío de materiales para el *DLE*” (Ruhstaller, 2003, p. 238).

comunes que ha llegado hasta la actualidad (RAE, s.f.-b). Pese a las profundas renovaciones y remodelaciones, el carácter del *Diccionario de autoridades* sigue permeando los diccionarios actuales y justificando muchos de sus rasgos (Ruhstaller, 2003, p. 237).

La edición de la que nos serviremos en esta propuesta, enmarcada en la 23ª edición de este repertorio publicada en formato físico en 2014, es la edición en línea o versión electrónica 23.5, que incluye las actualizaciones aprobadas por la RAE y la ASALE en 2021, tal y como se indica en el subapartado “Actualización 2021” de la página web de este diccionario (RAE, 2021b). Siguiendo la clasificación manejada en el apartado anterior, podríamos decir que se trata de un diccionario monolingüe, semasiológico, normativo-descriptivo, elaborado fundamentalmente a partir de otro diccionario y general.

Si bien el *Diccionario académico* se ha distanciado de su inicial carácter normativo, parece mantenerse la idea de que el léxico que registra es el considerado modélico o ideal de la lengua (Campos y Pérez, 2003, p. 69). El motivo de esta consideración vendría dado por el origen de este diccionario, vinculado al *Diccionario de Autoridades*, y al carácter de la institución de la RAE, que, bajo el lema de “Limpia, fija y da esplendor”, surgió con el fin de preservar el grado de esplendor del español alcanzado en los Siglos de Oro, evitando los cambios o descuidos en la lengua (Ruhstaller, 2003, p. 238).

La Academia se ha distanciado en buena medida de ese carácter normativo inicial (Ruhstaller, 2003, p. 238). Así, su objetivo primordial en la actualidad es el de “velar por que los cambios que experimente la lengua española en su constante adaptación a las necesidades de sus hablantes no quiebren la esencial unidad que mantiene en todo el ámbito hispánico” (RAE, s.f.-c). En esta defensa, se aprecia un objetivo normativo laxo y panhispánico: se asume que la lengua cambia de manera inevitable según las necesidades de los hablantes, pero que estos cambios han de combinarse con una mínima regulación y fijeza en la lengua con la finalidad última de conservar una cierta unidad del español en todo su dominio y conservar el carácter propio de la lengua y su corrección. Dado este carácter híbrido, y considerando que en las últimas ediciones se han admitido numerosas palabras marcadas en lo diastrático o diafásico (léxico vulgar, coloquial,

jergal) o extranjerismos, Haensch y Omeñaca (2004, p. 62) señalan que tiene un carácter mixto normativo-descriptivo, “pero con claro predominio del criterio normativo”.

En cuanto al sistema lingüístico que toma como base para su confección, como decíamos, el *DLE* sigue siendo muy dependiente del *Diccionario de Autoridades* (1726-1739). La nomenclatura del *Diccionario de Autoridades* se ha ido modificando a lo largo de los años en las sucesivas ediciones del *DLE*, pero nunca ha llegado a practicarse una remodelación total (Ruhstaller, 2003, p. 237). Por ello, decimos que el *DLE* toma como base otro diccionario. Sin embargo, en los últimos años, como se expondrá más adelante, la Academia ha comenzado a hacer uso de los corpus (*CREA*, *CORDE* y *CORPES*) para la construcción de su leuario (Haensch y Omeñaca, 2004, p. 215), aunque no sea suficiente como para considerarlo un diccionario basado en corpus.

Al respecto de su carácter general, la mayor parte del vocabulario registrado en el *DLE* pertenece al llamado nivel estándar de la lengua: un léxico no marcado en lo cronológico, diatópico, diatécnico, etc. El léxico correspondiente al nivel culto o literario también está muy representado. Por el contrario, como veremos en seguida, la representatividad de voces marcadas ha sido puesta en tela de juicio en numerosas ocasiones. Cabe decir en este punto que todos los diccionarios generales se enfrentan a las mismas dificultades para encontrar un equilibrio entre la representación de los subconjuntos manteniendo su carácter general, dado que no existe un criterio único que determine cómo hacerlo (Haensch y Omeñaca, 2004, pp. 188-189).

#### *Macroestructura del DLE*

Con respecto a la macroestructura del *DLE*, organizada en orden alfabético en su formato físico, se ha de analizar el carácter de su nomenclatura, o la inclusión o no de determinados lemas. En cuanto a las palabras con marca cronológica, el *DLE* ha sido muy criticado por el mantenimiento de palabras anticuadas y en desuso; no siempre marcadas como tal (Haensch y Omeñaca, 2004, p. 211), aunque cabe considerar que la presencia de estas palabras se debe a que este repertorio pretende también facilitar al usuario la descodificación de textos compuestos desde el siglo XVI. Al mismo tiempo, se ha cuestionado su excesiva cautela ante la inclusión de neologismos y extranjerismos, pues, dada la velocidad con la que se crean estas palabras y el estrecho contacto entre lenguas

actual, son muchas las ausencias detectadas (Ruhstaller, 2003, p. 254). Dado que el propósito principal de la Academia es mantener una cierta corrección y unidad en el idioma, solo admite la entrada de aquellas palabras con un uso más o menos consolidado y recomienda su adaptación ortográfica para no transgredir las normas ortográficas del español (Ruhstaller, 2003, p. 254).

La representación de palabras con marca diatópica, aunque es considerable, se ha descrito como azarosa e insuficiente. Además, se ha resaltado un desequilibrio en la representación de las áreas dialectales –con un marcado protagonismo del español peninsular– y algunos errores de marcación geográfica (v.g. mostrar como léxico común algunas palabras de uso exclusivo en España, como *albornoz*, o proponer la marca *América* para palabras de uso solo en algunos países americanos) (Haensch y Omeñaca, 2004, pp. 201- 211; Ruhstaller, 2003, p. 253). No obstante, esta es una cuestión en la que la Academia, colaborando con la ASALE y gracias a la elaboración *Diccionario de americanismos*, ha hecho grandes avances (RAE, 2014a, p. 10).

El léxico con marcas de nivel de uso (diafásicas o diastráticas), de nuevo, se ha criticado por sus ausencias, sin embargo, hemos de considerar que estas ausencias tienen una justificación considerando el carácter general del repertorio y teniendo en cuenta que, dado que estas palabras se renuevan constantemente, podrían multiplicar las entradas de vida efímera. Además, el léxico con marca “coloquial” aumenta de manera considerable en cada edición del *DLE*, aunque el léxico “malsonante”, en cambio, ha sido de los menos representados históricamente en el *DLE* (Ruhstaller, 2003, p. 254), pese a su frecuencia en los intercambios comunicativos cotidianos y espontáneos. Por último, la inclusión del léxico con marcas diatópicas, o perteneciente a un campo específico, se limita a aquellas palabras de uso extendido en la actualidad (RAE, 2014b, p. 5)

### *La microestructura del DLE*

Como señala Garriga (2003, p. 126), para conocer en detalle la microestructura del *DLE*, conviene consultar el análisis realizado por Ignacio Ahumada Lara (1989) el cual “sigue siendo válido, a pesar de que, por el momento en que se realizó, tome como referencia la 20ª edición (1984)” del repertorio. Con el fin de actualizar el análisis, ha de

sumarse a este estudio la información proporcionada por la propia Academia acerca de la vigesimotercera edición del repertorio.

El *DLE* puede incluir en sus artículos lexicográficos informaciones de distinto tipo, aunque su estructura básica y constante está compuesta por el lema, la indicación de su categoría y subcategoría gramatical y su definición (Ahumada, 1989, p. 57). El resto de las informaciones se consideran complementarias y pueden aparecer o no en las entradas. A continuación, presentamos las informaciones que, por lo general, puede presentar el diccionario académico en su 23.<sup>a</sup> edición siguiendo su disposición habitual. Como apoyo visual, proporcionamos en el Anexo II (imagen 2) un esquema de algunos de estos elementos a partir de la entrada simplificada de *saco*, que consideramos muy completa.

En primer lugar, englobadas en el llamado “paréntesis informativo” que sigue al lema, se pueden señalar las variantes secundarias del lema (v.g. señala en la entrada de *período* que hay otra variante extendida como *periodo*), la etimología, la información ortográfica (p. ej. señala que la palabra *virgen* se escribe en mayúscula para la tercera acepción) y la información morfológica. En esta última se incluye información acerca de flexiones irregulares en verbos, señalando su modelo de conjugación (indica en la entrada de *contener* que se conjuga como *tener*), la existencia de femeninos irregulares en sustantivos (apunta que, en algunos casos, *juez* se usa en femenino), o la relación entre un adjetivo y su superlativo irregular (v.g. entre *pobre* y *paupérrimo*) (RAE, 2014b, pp. 9-10; RAE 2021b).

A continuación, en la entrada lexicográfica se proporciona la acepción o serie ordenada de las acepciones de los lemas, que –como veremos– no sigue una ordenación estrictamente de frecuencia (Garriga, 2003, p. 108). En cada acepción, antes de la correspondiente definición, se proporciona, en primer lugar, la información gramatical de la palabra con el significado en cuestión, su categoría o subcategoría de palabra; por ejemplo: *adj. dem.* (adjetivo demostrativo), *tr.* (verbo transitivo), etc. A continuación, se informa de sus marcas de uso en caso de no ser de empleo general en la lengua: marcas de vigencia (*desus.* “desusado”, *p. us.* “poco usado”), diatópicas (*Ar.* “Argentina”, *Am.* “América”...), de especialidad (*Med.* “Medicina”), de nivel de uso (*coloq.* “coloquial”, *jerg.* “jergal”, etc.), o pragmáticas (*desp.* “despectivo”, *irón.* “irónico”) (Ahumada, 1989, pp. 65-72; RAE, 2014b, pp. 11-12).

Tras la definición de cada acepción, cuando esta se mantiene esencialmente invariable, se incluye una indicación de los posibles cambios gramaticales de la palabra en esa acepción (*U. t. c. s.* “usado también como sustantivo”) o algún apunte sobre sus transiciones semánticas (*U. t. en sent. fig.* “usado también en sentido figurado”) (Ahumada, 1989, p. 72 y pp. 79-80). Solo de manera excepcional, tras lo anterior, se incluye algún ejemplo de uso inventado, pese a ser de gran utilidad para la comprensión del significado de la palabra y para conocer su comportamiento sintáctico (Garriga, 2003, p. 122). De hecho, la información sobre la combinatoria o sintaxis de una palabra (colocaciones, régimen preposicional, selección de complementos) no suele aparecer de manera formalizada en el diccionario; aparece implícita en los ejemplos o en la propia definición, precedida de fórmulas como “con sustantivos”, “dicho de”, “se aplica a” (Castillo, 2003, p. 91). Por ejemplo, en la entrada de *pensar*, los ejemplos indican la preposición que rige el verbo en versalitas (*pensar EN algo*), y en la entrada de *hojiblanco/a*, la fórmula “Dicho de una aceituna...” (RAE, 2021b) que encabeza la entrada nos proporciona pistas de su comportamiento combinatorio.

Por último, tras la relación de las acepciones, se registran las formas complejas o unidades pluriverbales en las que participa el lema, de las cuales se proporciona una información similar a la de cualquier acepción y en el mismo orden descrito (categoría, marcas de uso, definición, etc.). Estas se agrupan en dos bloques: el primero, si la entrada corresponde a un sustantivo, está conformado por las combinaciones estables del lema con algún elemento que actúe como un modificador con respecto a él (*SACO terrero*); el segundo está compuesto por las expresiones y locuciones en las que participa el lema ( *echar en SACO roto*). Las distintas subentradas de cada bloque se ordenan por orden alfabético (RAE, 2014a, p. 52).

Sin pretender un análisis minucioso de cada una de estas cuestiones, por coherencia con las propuestas didácticas que se sugieren en este TFM, señalaremos tres limitaciones de este repertorio lexicográfico. En primer lugar, la insuficiente información gramatical sobre los lemas. A diferencia de los diccionarios de uso, como el *DEA* o el *DUE*, apenas proporciona información sobre el uso contextual de las palabras; no suele mencionarse el régimen preposicional, los complementos que aceptan los lemas o sus colocaciones (Haensch y Omeñaca, 2004, p. 191). En segundo lugar, la falta de ejemplos que ilustren

el uso de las palabras, que, cuando se proporcionan, son inventados, no documentados. Por último, cabe señalar que en la ordenación de las acepciones se intentan combinar distintos criterios (categoría gramatical, frecuencia, marcas de uso, etc.), por lo que no responde a una ordenación estrictamente de frecuencia (Garriga, 2003, p. 108). De esta manera, es posible que el usuario encuentre como primera acepción una definición que no coincida con su expectativa (Rojo, 2021, p. 194), como es el caso de la palabra *enervar*; según el *DLE* (RAE, 2014a, definición 1): ‘Debilitar, quitar las fuerzas’.

### 2.3. Conclusiones

En los últimos años, desde la década de los 90 del siglo pasado, se han subsanado algunas de las carencias del *DLE* citadas, tanto en la macroestructura como en la microestructura (Haensch y Omeñaca, 2004, p. 214). Se han incorporado miles de entradas, se han suprimido las entradas de algunas voces en desuso y se han enmendado otras tantas (errores de marcación, acepciones, etc.) (RAE, 2014a, p. 11); se han registrado más coloquialismos y extranjerismos de uso común, lo cual demuestra un cierto aperturismo y un claro distanciamiento del inicial criterio normativo, y se han multiplicado los americanismos recogidos (Haensch y Omeñaca, 2004, p. 214). Tales reformas han sido posibles, entre otros motivos, gracias a la mayor colaboración con la ASALE para conocer la realidad del español de América (RAE, 2014a, p. 10) y al uso de corpus construidos por el Instituto de Lexicografía de la RAE (*CREA*, *CORDE*, *CORPES*), con los que se estudia la vigencia, distribución y frecuencia de uso de las voces con el fin de marcarlas debidamente, suprimirlas o incorporarlas al diccionario (Haensch y Omeñaca, 2004, p. 215). En esta última cuestión reside, precisamente, el objetivo principal de la propuesta didáctica que plantearemos.

Como se ha indicado, el *CORPES* refleja la lengua tal y como la usan los hablantes, con sus incorrecciones o desviaciones normativas, innovaciones, etc. Además, nos proporciona pistas sobre la frecuencia (un criterio, por lo general, desatendido en diccionarios y fundamental para determinar si han de incluirse o no en el leuario) y distribución de uso de las palabras vinculada a los distintos parámetros de codificación (país, tipo de texto, fecha, etc.) (Rojo, 2021, p. 164). Por otro lado, este corpus nos proporciona ejemplos reales de uso de cada palabra en un contexto, tanto social como lingüístico. Así pues, permite conocer la manifestación lingüística, con frecuencia, de un

modo más cercano a la realidad que el que proporciona el *DLE*. A pesar de esto, hay que tener en cuenta que los corpus presentan también un límite de representación, y es posible que la búsqueda de una forma no devuelva resultados (o los condicione, como ocurre con las coapariciones), aunque tenga un uso relativamente extendido en la realidad (Rojo, 2021, p. 165).

Asimismo, no debemos olvidar la importancia de la labor de regularización del *DLE* para preservar la unidad de la lengua y su sistema ortográfico y garantizar así la intercomprensión de los hablantes. Tampoco hemos de pasar por alto la labor filológica que hay detrás de las entradas: la información etimológica de las voces; la cuidada categorización gramatical de las palabras, a partir del estudio individual de estas, no de mecanismos computacionales; la cautela al incluir nuevas voces para no multiplicar las entradas de vida efímera; el rigor al proponer las unidades pluriverbales en las que participan los lemas o la certeza de sus definiciones, la cual, a través de los corpus, requeriría, aparte de una destreza para definir, como señalaba Rojo (2021, p. 164), el análisis individual de un gran número de ejemplos.

En resumen, la rapidez y comodidad de obtención de datos lingüísticos que nos proporciona el corpus, con su representación ajustada a la realidad, contextualizada (en la estructura sintáctica y en el uso), y su información de frecuencia y distribución de uso, “no puede hacernos olvidar la necesidad de revisar los datos obtenidos desde una perspectiva filológica” (Rojo, 2021, p. 168). Se trata, en suma, de aprovechar los datos que nos proporciona de manera fácil y exhaustiva esta herramienta y analizarlos de manera crítica, a partir de criterios lingüísticos o lexicográficos, para perfeccionar nuestro estudio. Todo ello nos lleva a proponer el empleo de ambos recursos de la misma institución como complementarios y necesarios, sin pretender la preponderancia de uno sobre otro, para obtener una visión más completa de la manifestación lingüística y, en concreto, del léxico del español.

### **3. Revisión de propuestas de uso didáctico de los corpus**

La extensión del uso de los corpus textuales desde la investigación al ámbito de la enseñanza de lenguas ha sido una cuestión contemplada desde los inicios de la disciplina de la Lingüística de Corpus (Rojo, 2021, p. 22). Según recogen Parodi y

Burdiles (2019, p. 597), las propuestas de uso didáctico de los corpus toman como base el planteamiento de Johns (1986), que proponía que los estudiantes del inglés como L2 se sirviesen de los corpus a modo de *input* lingüístico para acceder a un uso real y contextualizado de la lengua que querían aprender. Desde finales del XX, parece haberse multiplicado el interés por los corpus en relación con la teoría y práctica de la enseñanza de lenguas, así como por la creación de estos recursos en línea para la explotación didáctica, aunque sobre todo con respecto a la enseñanza del inglés (Parodi y Burdiles, 2019, p. 597). Del mismo modo, como señala Tolchinsky (2014, p. 13), en los contextos en los que se ha difundido el uso pedagógico de los corpus, como EE. UU., Australia o Gran Bretaña, se propone que la formación del profesorado de lenguas contemple el uso de corpus con el fin de que el alumnado se acabe beneficiando de ellos, puesto que una de las limitaciones de este enfoque es el desconocimiento del profesorado de estas herramientas (Martínez, 2018, p. 17).

Las aplicaciones didácticas de estos recursos –simplificando la cuestión– pueden abordarse atendiendo a dos parámetros. El primero es si el uso que se hace de estos es directo (en la clase de lengua) o indirecto (para la investigación o la confección de diccionarios, materiales de aula o programas de enseñanza ajustados a la lengua de uso real) (Tolchinsky, 2014, p. 13). El segundo hace referencia a si se han aplicado para los procesos de enseñanza-aprendizaje de la lengua materna o de lenguas extranjeras. Aunque nuestro objetivo es recoger las propuestas de uso directo de corpus en la enseñanza del español como lengua materna, dada la escasez de propuestas al respecto, incluso para la enseñanza del español como L2<sup>9</sup> (Parodi y Burdiles, 2019, p. 601), recogeremos las orientaciones metodológicas y actividades propuestas en general para el uso directo de corpus en la enseñanza de lenguas.

A grandes rasgos, las propuestas contemplan el uso de los corpus como fuentes de textos orales y escritos cuya observación permite aprender cuestiones vinculadas a todos los niveles de análisis lingüístico: fónico, léxico, gramatical y discursivo (Parodi y Burdiles, 2019, p. 602). En el nivel fónico, por ejemplo, se plantea el uso de corpus orales para analizar los rasgos de la comunicación oral como los solapamientos o la pronunciación relajada (Martín, 2013, p. 8); estos corpus, a su vez, permiten observar las

---

<sup>9</sup> Algunas de estas, citadas en Parodi y Burdiles (2019), son: Drange (2008) Mendikoetxea (2013), Bailini (2014), Albelda (2010) y López-García (2013).

normas sociales de interacción comunicativa (Parodi y Burdiles, 2019, p. 602). En el nivel léxico, se propone el uso de corpus para analizar el uso contextual de las palabras, su variación en función de la situación comunicativa y sus relaciones con otros elementos –colocaciones, relaciones semánticas, etc.– (Martín, 2013; Parodi y Burdiles, 2019, p. 604). En el nivel gramatical, se plantea el uso de estas herramientas para analizar, comprobar o inducir normas gramaticales a partir de los ejemplos o contrastar la norma gramatical y el uso real de los hablantes (Parodi y Burdiles, 2019, p. 604) o estudiar el comportamiento sintáctico de las piezas léxicas (Bordón, 2019, p. 220; Martín, 2013, p. 8). En cuanto al nivel discursivo, dada la codificación de los corpus, esta herramienta se emplea para analizar y reflexionar sobre las variaciones (diatópicas, diafásicas, etc.) de la lengua en función de la situación comunicativa (Parodi y Burdiles, 2019, p. 604; Rojo, 2021, p. 285).

Una de las orientaciones metodológicas de uso directo del corpus en la enseñanza de lenguas es la que propone que el profesor y el alumnado, en conjunto, mediante la exploración de los datos del corpus y su discusión, procuren establecer teorías explicativas sobre estos (Tolchinsky, 2014, p. 14). Otro enfoque distinto es el que aboga por que sea el estudiante el que asuma el rol investigador, formule hipótesis y trate de comprobarlas mediante las búsquedas en el corpus, de modo que aprenda de un modo autónomo, mientras que el docente actúa como un mediador del proceso de aprendizaje (Martínez, 2018, p. 15; Parodi y Burdiles, 2019, p. 602). En última línea se sitúa el trabajo de Martín (2013), que propone que los alumnos usen de manera autónoma los corpus para tratar de resolver cuestiones planteadas por el docente tales como cuáles son las unidades pluriverbales en las que participa una determinada palabra, o la variación de la frecuencia de uso de las voces. A su vez, Tolchinsky (2014, p. 14) propone que los alumnos, de manera autónoma, traten de resolver la hipótesis de, por ejemplo, en qué tipo de texto encontrarán palabras de mayor longitud, mayor densidad léxica, etc.

No obstante, considerando que los corpus surgen del ámbito de la investigación, el diseño y la terminología de estos puede dificultar el uso autónomo por parte de los alumnos; por ese motivo, es necesario que el docente, como punto de partida para el uso de estos recursos en el aula, establezca un proceso de enseñanza en tres partes: explicar qué son los corpus, enseñar a “explotarlos” y, por último, hacer que los alumnos los utilicen para resolver tareas o investigaciones a partir de los datos que le proporcionan

(Parodi y Burdiles, 2019, p. 602). En otras palabras, se trata de que el docente aborde “un proceso de andamiaje” para que los alumnos aprendan progresivamente a utilizarlos de manera autónoma (Martínez, 2018, p. 12). Esta orientación metodológica es precisamente la que seguiremos en la propuesta didáctica que se desarrolla a continuación, la cual, enmarcada en las actividades relacionadas con el nivel léxico, va dirigida al manejo e integración de este recurso junto a los diccionarios para propiciar que los alumnos de 4º de la ESO alcancen un conocimiento más amplio del léxico español.

#### **4. Propuesta didáctica**

##### **4.1. Presentación y enfoque**

Las conclusiones derivadas de la fundamentación teórica presentada en los apartados precedentes nos permiten diseñar una propuesta didáctica dirigida a 4º de la ESO que abogue por el uso integrado de diccionarios y corpus textuales, partiendo de dos recursos de acceso gratuito alojados en la página web de la Real Academia Española ([www.rae.es](http://www.rae.es)), como institución de referencia de la lengua española: el *DLE* y el *CORPES*. Para ello, comenzaremos repasando y fortaleciendo los conocimientos del alumnado sobre el *DLE* para, sobre estos, incorporar el conocimiento y uso del *CORPES*. Tras algunos ensayos del manejo de la aplicación de consulta del *CORPES*, se pedirá a los alumnos que, por equipos, adoptando el rol de lexicógrafos de la RAE, compongan una entrada al estilo del *DLE* partiendo de la información obtenida mediante búsquedas en el corpus. Se cerrará esta secuencia didáctica con la exposición de las distintas entradas y una reflexión sobre el proceso de enseñanza-aprendizaje.

Esta secuenciación esbozada resulta pertinente si consideramos, como se propone, entre otros, en el marco de la teoría constructivista del aprendizaje, que para que los estudiantes puedan ampliar sus esquemas de conocimiento, es necesario que los nuevos contenidos (el uso del *CORPES*) se integren en las estructuras de conocimiento ya existentes (el conocimiento y uso de otra herramienta de consulta, como el *DLE*) (Ortiz, 2019, p. 101). Por otro lado, la tarea final propuesta implica “la resolución colaborativa de problemas” ajustados a situaciones existentes en la vida real y refuerza la “autonomía, la reflexión y la responsabilidad”, lo cual nos aproxima a algunos de los principios pedagógicos para la Educación Secundaria Obligatoria recogidos en la *Ley Orgánica*

3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación (Ley Orgánica 3/2020, 2020, p. 26).

#### 4.2. Objetivos

Los objetivos-estándares para la asignatura de Lengua Castellana y Literatura recogidos en el *Real Decreto 1105/2014* directamente relacionados con esta propuesta, presentados en la Tabla 4, son los siguientes (Real Decreto 1105/2014, 2014, p. 367 y p. 369):

**Tabla 4**

*Estándares de aprendizaje evaluables relacionados con la propuesta*

Bloque de la materia	Estándares de aprendizaje evaluables
Bloque 2. Comunicación escrita: leer y escribir	4.1. Conocer y manejar habitualmente diccionarios impresos o en versión digital, diccionarios de dudas e irregularidades de la lengua, etc.
Bloque 3. Conocimiento de la lengua	5.1. Utilizar los diccionarios y otras fuentes de consulta en papel y formato digital resolviendo eficazmente sus dudas sobre el uso correcto de la lengua y progresando en el aprendizaje autónomo.

*Nota:* Reformulamos los estándares de aprendizaje expresándolos con un verbo en infinitivo.

Los objetivos formulados específicamente para esta propuesta didáctica, que concretan y amplían los anteriores, son los siguientes:

- Describir el *Diccionario de la lengua española* como herramienta de consulta lingüística atendiendo a las particularidades de su macroestructura y microestructura.
- Identificar e interpretar correctamente las distintas informaciones (gramaticales, semánticas, registro y uso) que conforman la microestructura del *DLE* a partir de una selección de entradas.
- Usar el *CORPES* para resolver correctamente consultas básicas acerca del uso de determinadas palabras en los textos alojados en este corpus.

- Examinar y explicar las diferencias en la información proporcionada por el *DLE* y el *CORPES* acerca de determinadas palabras.
- Construir una entrada de diccionario siguiendo el modelo de la microestructura del *DLE* e incorporando la información obtenida a partir de búsquedas en el *CORPES*.
- Valorar la importancia de integrar la información proporcionada por los diccionarios con la información proporcionada por los corpus sobre el uso real y contextualizado de las palabras para alcanzar una comprensión más amplia del léxico español.

#### 4.3. Contenidos

En estrecha relación con los objetivos presentados en el apartado anterior, presentamos en la Tabla 5 los bloques de contenidos recogidos en el *Real Decreto 1105/2014* (Real Decreto 1105/2014, 2014, p. 367 y p. 369) directamente vinculados a nuestra propuesta:

**Tabla 5**

*Bloques de contenidos relacionados con la propuesta*

Bloque de la materia	Contenidos
Bloque 2. Comunicación escrita: leer y escribir	Utilización progresivamente autónoma de los diccionarios, de las bibliotecas y de las Tecnologías de la Información y la Comunicación como fuente de obtención de información
Bloque 3. Conocimiento de la lengua	Manejo de diccionarios y otras fuentes de consulta en papel y formato digital sobre la normativa y el uso no normativo de las palabras e interpretación de las informaciones lingüísticas que proporcionan los diccionarios de la Lengua: gramaticales, semánticas, registro y uso

Los contenidos que planteamos específicamente para esta propuesta, concretados y ampliados a partir de los anteriores y vinculados a las competencias clave –según el *Real Decreto 1105/2014*– que se desarrollarían, son los siguientes:

- Conocimiento del *Diccionario de la lengua española* atendiendo a los aspectos generales de su macroestructura (lemario) y microestructura (información semántica, gramatical, etimológica, ortográfica, de registro y de uso): CL (Competencia en comunicación lingüística)
- Conocimiento y uso del *CORPES* atendiendo a las casillas habilitadas para las consultas básicas en la ventana de “Concordancias” y las ventanas de “Estadística” y “Coapariciones”: CL, CD (Competencia digital), CMCT (Competencia matemática y competencias básicas en ciencia y tecnología).
- Uso autónomo del *DLE* y otras fuentes de consulta lingüística (*Observatorio de palabras de la RAE, Diccionario Etimológico Castellano en línea, Fundéu*): CL, CD, CPAA (Competencia para aprender a aprender).

#### 4.4. Metodología

En esta propuesta didáctica, tratando de adecuarnos a la naturaleza de los diferentes contenidos o actividades de aprendizaje-enseñanza implicados, alternaremos dos metodologías: el modelo de enseñanza directa con toda la clase y el modelo de aprendizaje mediante grupos de trabajo, ambos descritos por Moral (2019, pp. 151-172) y cuyos principios desarrollaremos a continuación. A partir del primero se les proporcionarán a los estudiantes unas nociones teóricas básicas acerca del *CORPES* y el *DLE* y a partir del segundo modelo se espera que los estudiantes elaboren una entrada de diccionario partiendo de los datos obtenidos con el corpus. La alternancia de las dos metodologías implica la combinación de distintos tipos de agrupamiento durante las sesiones: el individual y el de pequeños grupos (3-4 alumnos).

El modelo de aprendizaje mediante grupos de trabajo, apoyándose en la teoría social constructivista, plantea que el conocimiento se adquiere mediante la interacción y consenso con otros –tanto profesor como alumnos (Ortiz, 2015, p. 99)–; lo que exige la implicación activa del alumnado en su proceso de aprendizaje (Moral, 2019, p. 162 y p. 165). Uno de los supuestos de esta metodología es que el conocimiento del grupo es

superior al de sus componentes de manera individual, por lo que se aplica para la resolución de problemas más complejos, abiertos, y que requieran “un pensamiento crítico con apoyo en evidencias y argumentos” (Moral, 2019, p. 163 y p. 165). En nuestro caso, se trata de la elaboración de una entrada de diccionario *ex novo* tomando decisiones acerca de la inclusión de ciertas informaciones a partir de la exploración de datos del *CORPES*. En Mujis y Reynolds (2005) y Slavin (2017), citados por Moral (2019, p. 64), se definen cuatro fases en este modelo de enseñanza: 1) conformación de los grupos de trabajo, 2) explicación la tarea que se ha de acometer, 3) realización de la tarea mediante grupos de trabajo (sesiones 5 y 6 en nuestra propuesta) y 4) exposición final de las tareas de todos los grupos (sesión 7).

El modelo de enseñanza directa con toda la clase, por su parte, se define como un modelo basado en la organización y presentación de contenidos o habilidades por parte del profesor hacia el alumnado (Moral, 2019, p. 151). Esta metodología suele aplicarse a los contenidos y habilidades bien estructurados que se aprenden “paso a paso” o que los estudiantes utilizan para la resolución de problemas (Moral, 2019, p. 152). En nuestro caso, los contenidos de la secuencia didáctica se prestan a una organización escalonada en fases bien entrelazadas sobre las que los alumnos habrán de progresar de manera gradual (primero el *DLE*, luego las consultas básicas en *CORPES*, etc.) para, finalmente, construir una entrada de diccionario.

Este modelo de enseñanza-aprendizaje integra algunos principios de las teorías sobre organización del conocimiento y de la teoría del aprendizaje conductista; dos de los cuales se han tomado especialmente en consideración a la hora de diseñar esta secuencia didáctica: “organizar y dividir el material en pequeñas etapas para facilitar la construcción de la estructura de conocimiento” y “hacer practicar al alumno en cada paso e ir incrementando la complejidad” (Rosenshine y Stevens, 1989 citado en Moral, 2019, p. 152). En nuestro caso, además, en las sesiones desarrolladas conforme a esta metodología, retomando lo mencionado en cuanto al modelo de aprendizaje constructivista, insistiremos en el diálogo con los alumnos como forma de construir de manera conjunta el conocimiento (Ortiz, 2015, p. 100).

Las fases propias del modelo son tres: presentación, práctica guiada y práctica independiente. La fase de presentación (sesiones 1, 2 y 3) debe contemplar una etapa

inicial de activación de los conocimientos previos del alumnado en relación con lo que se estudiará, continuar con la exposición de los objetivos de la secuencia y finalizar con la presentación de los nuevos contenidos o habilidades (Moral, 2019, pp. 154-156). En la fase de la práctica guiada (sesión 4), los alumnos aplican las nuevas habilidades de un modo controlado; solventando errores de comprensión de lo presentado de manera inmediata para que los estudiantes adquieran la destreza necesaria para acometer tareas con mayor autonomía en la siguiente etapa (Moral, 2019, p. 157).

En la última etapa, o práctica independiente (sesiones 5 y 6), una vez que los estudiantes han asimilado las nuevas habilidades, se pretende facilitar que los alumnos transfieran los nuevos contenidos a otros contextos (Moral, 2019, p. 158); esta fase, en nuestro caso, coincide con el trabajo colaborativo. Por lo general, en las sesiones de clase ajustadas a este modelo se dedica un espacio a la revisión y control de las tareas del día anterior para fomentar la práctica, hacer correcciones, asegurar la realización del trabajo y ofrecer un *feedback* ajustado (repetición colaborativa, corrección de tareas...) (Moral, 2019, pp. 153-159).

#### 4.5. Recursos

Los materiales o recursos didácticos que se emplearán serán variados, aunque adquieren especial protagonismo las TIC, por lo que habrá que solicitar con antelación, si es necesario, la sala de informática del centro en el que se aplique y asegurarse de que todo el alumnado tiene acceso a ordenadores en tiempo no lectivo. Los principales recursos de esta secuencia serán el *DLE* en versión digital y el *CORPES*, aplicación de consulta *online*. Junto a estas dos herramientas, se usará un formulario de *Google* para la realización de una evaluación y dos imágenes interactivas realizadas con *Genially* (un esquema de la microestructura del *DLE* y un manual de uso del *CORPES*). Aparte de los recursos digitales, se les proporcionará a los estudiantes el siguiente material impreso: una adaptación de las entradas lexicográficas de *saco* y *amarrar* del *DLE* (Anexo III.A), una propuesta de ejercicios para el aula de informática (Anexo III.B) y una guía para elaborar la entrada del diccionario (Anexo III.C).

## 4.6. Secuenciación

La propuesta didáctica se desarrollará en siete sesiones de aula, que habrán de completarse con cierto trabajo autónomo del alumnado. Algunas de estas sesiones tendrán lugar en el aula ordinaria y otras habrán de realizarse en el aula de informática del centro. Como decíamos, todo ello requiere asegurarse de que todos los estudiantes tienen acceso a ordenadores en horario no lectivo (bibliotecas, préstamos de material informático, ordenador en casa, etc.) y solicitar la sala de ordenadores del centro con antelación.

### 4.6.1. Primera sesión: El *DLE*, un vastísimo edificio (I)

Esta primera sesión, al igual que la segunda, irá dirigida a reforzar los conocimientos del alumnado acerca del *DLE* para asentar sobre estos el conocimiento y uso del *CORPES*. Como forma de activar sus conocimientos previos sobre el tema, empezaremos la sesión leyendo dos fragmentos que ofrecen una concepción subjetiva de ‘diccionario’: un fragmento del cuento *La conjuración de las palabras* de Benito Pérez Galdós y un fragmento de *Rayuela* de Julio Cortázar; ambos modificados para el aula.

#### *La conjuración de las palabras*<sup>10</sup>

Érase un gran edificio llamado ... Por dentro era un laberinto tan maravilloso que ni el mismo de Creta se le igualara... [En sus crujías] se hallaban innumerables celdas ocupadas por los ochocientos o novecientos mil seres que en aquel vastísimo recinto tenían su habitación. Estos seres se llamaban palabras.

#### *Rayuela*<sup>11</sup>

Fue a buscar el [...], lo abrió al azar y preparó para Manú el siguiente juego en el cementerio. «Hartos del cliente y de sus cleonasmos, le sacaron el clíbano y el clípeo y le hicieron tragar una clica. Luego le aplicaron un clistel clínico en la cloaca, [...]»

---

<sup>10</sup> (Pérez Galdós, 1868 citado en Seco, 13 de marzo de 2014, p. 1)

<sup>11</sup> (Cortázar, 2013, p. 259)

—Joder — dijo admirativamente Oliveira. Pensó que también joder podía servir como punto de arranque, pero lo decepcionó descubrir que no figuraba en el cementerio [...] «Es realmente la necrópolis», pensó.

Tras la lectura común de estos textos, dejaremos algunos minutos para que los alumnos escriban qué palabra creen que funciona como sinónimo de edificio, laberinto, cementerio y necrópolis (el *Diccionario académico*) y por qué lo creen. A continuación, mediante un diálogo abierto con todos los alumnos, abordaremos una explicación grupal de la metáfora del diccionario en los textos: por qué la han identificado, cómo conecta su visión previa del diccionario con esas metáforas y qué les sugieren (connotaciones positivas, negativas y por qué). Asimismo, se les invitará a reflexionar acerca de la importancia de los diccionarios y sobre el tipo de vocabulario que registran (coloquialismos, extranjerismos, etc.).

Hecha esta primera aproximación al tema, el docente esbozará brevemente cómo será la secuencia didáctica que se desarrollará durante los siguientes días y cuál es el objetivo: conocer el *Diccionario de la lengua española* y toda la información que proporciona sobre las palabras que se alojan en él y conocer y usar otra herramienta, como el *CORPES*, en la que las palabras (seres) que la habitan no se sitúan en celdas o nichos, sino en atmósferas discursivas.

Concluidas estas dos etapas, la etapa de activación de conocimientos previos y la de la exposición de los objetivos, empezaremos con la presentación de los nuevos contenidos: la descripción del *DLE*, atendiendo a las cuestiones básicas de su macroestructura y microestructura. En cuanto a lo primero, se indicará que, al tratarse de un diccionario general, recoge, sobre todo, el léxico no marcado, por lo que no figuran en este repertorio todos los dialectalismos, coloquialismos, etc. Asimismo, se expondrá que, dado su carácter más normativo que descriptivo, con el fin de cuidar la corrección y unidad en la lengua y garantizar una mínima intercomprensión entre los hablantes, también son infrecuentes los extranjerismos o las formas no normativas, a pesar de que su uso extendido haga notable la ausencia. Para demostrar estas cuestiones, se buscarán en el *DLE* en línea algunas palabras o acepciones marcadas como *rayada* (coloquial

juvenil) y neologismos como *crush* para constatar su ausencia, y otras como *feedback*, *chat*, o *flipar* (coloquial), que sí están recogidas.

#### 4.6.2. Segunda sesión: El *DLE*, un vastísimo edificio (II)

Esta segunda sesión irá dirigida a abordar, de un modo general y no exhaustivo, la microestructura del *DLE*. Atendiendo al enfoque metodológico adoptado, comenzaremos la sesión con un repaso colaborativo de las cuestiones abordadas el día anterior acerca de este repertorio y su leuario, preguntando, por ejemplo, qué tipo de palabras aparecen mejor y peor representadas, a qué se debe esto considerando la tipología del diccionario, etc.

Para exponer la microestructura del *DLE* se tomará como base una imagen diseñada mediante la herramienta de creación de contenidos interactivos *Genially*, en la que figura un esquema reducido de lo que sería la microestructura del *DLE* manteniendo el formato tipográfico de la versión en línea (<https://bit.ly/3DWBudP>). En este esquema aparecen representadas las siguientes informaciones: el lema, las variantes secundarias, la etimología, las acepciones asociadas a los lemas simples y un apartado para las formas complejas (sin diferenciar entre combinaciones estables, expresiones o locuciones). Para cada acepción, se trate de lemas simples o formas complejas, aparecen situadas las marcas gramaticales (categoría o subcategoría gramatical), las marcas de uso (vigencia, zona geográfica, lenguaje de especialidad y de nivel de uso de la lengua: coloquial, jergal, etc.), así como un espacio para los ejemplos.

El docente repartirá una fotocopia con una adaptación de la entrada de la palabra *saco* (Anexo III.A) a los alumnos, en la que aparecen representados muchos de los elementos de la microestructura mencionados. Al mismo tiempo, se proyectará la imagen descrita en el párrafo anterior. Recorriendo en orden el esquema, se pedirá a los alumnos que, analizando la entrada fotocopiada, traten de señalar qué información se proporciona en cada espacio; información que irá ampliando, precisando y explicando el docente y reflejándolo en el esquema *ad hoc*.

Una vez explicada la microestructura, se propondrá a los estudiantes que, de manera individual, identifiquen los componentes de la microestructura trabajados en una

versión reducida de la entrada de *amarrar* del *DLE* en línea, de la que también se proporcionará una fotocopia (Anexo III.A). Esta fotocopia se recogerá al finalizar la clase para que el docente pueda revisarla.

#### 4.6.3. Tercera sesión: El *CORPES*: palabras interconectadas

Comenzaremos la sesión haciendo un repaso colaborativo sobre las nociones imprescindibles manejadas los días previos acerca del *DLE*: cómo era su macroestructura (o lecionario) y su microestructura (qué información proporcionan las entradas, en qué orden, etc.). Además, se devolverá a los alumnos su tarea de identificación de los componentes de la entrada corregida para que la conserven y la puedan utilizar en actividades posteriores.

A continuación, se abordará la presentación de los nuevos contenidos: qué es un corpus, qué es el *CORPES* y cómo se utiliza. Con el fin de facilitar la comprensión del recurso y su pertinencia, se explicarán sus rasgos por contraste con el *DLE*. Partiendo de la microestructura del diccionario ya trabajada, se irá exponiendo, a grandes rasgos, qué información proporciona y cuál no, cuál amplía o mejora a partir de ejemplos de búsqueda. Un corpus, como el *CORPES*, no aporta etimología ni definiciones –puesto que, a diferencia de lo que sucede en la entrada léxica, no hay una esmerada labor filológica detrás–. Sin embargo, este recurso sí muestra las acepciones de uso vigente (como el caso de *enervar*); permite buscar las variantes ortográficas; informa de su categoría gramatical (no siempre de forma adecuada, dado que la anotación se realiza de manera automática); proporciona una información muy refinada sobre las marcas de uso a través de la opción “Estadística”; facilita ejemplos en los que observamos las palabras interactuando o conviviendo con otras, etc. A su vez, se relacionará con las particularidades de la macroestructura del *DLE* ya comentadas –la ausencia de algunos neologismos, extranjerismos, o palabras marcadas–. Para ello se mostrará cómo, con frecuencia, el *CORPES* sí da testimonio de aquellas siempre que sean de uso por los hablantes, a partir de ejemplos como *un like*, *machirulo* o *mamera* (‘que causa pereza’ en México).

En la siguiente parte de la sesión, se procederá a explicar y practicar con los alumnos la aplicación de consulta, limitándonos a utilizar las casillas vinculadas a las

búsquedas básicas. Recorriendo la ventana principal de izquierda a derecha, se irán explicando las utilidades de cada casilla: lema, forma, clase de palabra, grafía original, subcorpus, orden y ejemplos. Asimismo, se hará un somero repaso por las ventanas de “Estadística” y “Coapariciones”. Cada vez que se exponga la utilidad de una casilla, el docente invitará a algún alumno a practicar con una búsqueda similar. Al finalizar la clase, se pedirá a los alumnos que revisen en sus casas antes de la siguiente sesión una imagen interactiva creada con *Genially* en la que sintetizan las utilidades de las casillas explicadas que se subirá a la plataforma virtual del centro (<https://bit.ly/3BPfpex>).

#### 4.6.4. Cuarta sesión: Consultas básicas en el CORPES

Al entrar a la sala de informática, con los ordenadores de los alumnos todavía apagados, se proyectará la imagen interactiva creada con *Genially* donde se sintetizan las utilidades de la aplicación de consulta para las búsquedas básicas. El docente repasará brevemente, preguntando a los alumnos, la función de cada casilla y lo comprobará situando el ratón encima de cada icono. De seguido, les repartirá una hoja con ciertas preguntas que han de responder. Estas preguntas se ajustan a dos niveles: preguntas de comprobación y preguntas de ampliación (Anexo III.B).

Para las primeras, los alumnos trabajarán de manera individual. El docente leerá cada pregunta en voz alta (por ejemplo, ¿cuántos casos hay del infinitivo *vivir*?) y los alumnos levantarán la mano cuando tengan una respuesta y la hayan registrado en su hoja de ejercicios. Una vez todos hayan levantado la mano, se dirá la respuesta correcta y aquellos con otros resultados deberán volver a intentarlo, con una cierta ayuda del docente o sus compañeros, hasta que salga el número correcto. De esta manera comprobaremos que todos los alumnos entienden el funcionamiento de la aplicación y solventaremos los problemas de incomprensión.

Una vez resueltas estas preguntas de comprobación, se pedirá a los alumnos que formen grupos de tres integrantes con quien tengan a su lado para trabajar sobre las preguntas de ampliación, aunque todos los alumnos habrán de escribir, individualmente, las respuestas en la hoja de ejercicios entregada por el profesor. En estas preguntas, que también son breves, se incluye una reflexión sobre los resultados que guarda relación con las clases teóricas previas acerca del recurso y su contraposición al diccionario

académico. Seguiremos un método parecido al anterior: dejaremos un tiempo para cada pregunta y se irá pidiendo, de manera alternada, a alguno de los grupos que comparta su respuesta o que añadan o discutan la de sus compañeros si tenían otra propuesta. Todas las respuestas serán corregidas o ampliadas/explicadas por el profesor. Al finalizar la sesión, el docente recogerá las hojas de preguntas de los alumnos.

#### 4.6.5. *Quinta y sexta sesión: En la piel de un lexicógrafo*

El objetivo de estas dos sesiones es que los alumnos, agrupados en equipos heterogéneos de cuatro personas y actuando como un verdadero equipo de lexicógrafos, compongan una entrada de diccionario de una determinada palabra asignada, siguiendo el modelo de microestructura del *DLE* a partir de las búsquedas en el *CORPES*, aunque limitándonos a las acepciones de las formas simples de la palabra.

A cada grupo de alumnos se le asignará una palabra del *Observatorio de palabras* de la RAE. El *Observatorio* es un portal lingüístico *online* de la institución en el que se ofrece información sobre palabras o acepciones que no aparecen en el diccionario, pero de las que hay varios testimonios (neologismos recientes, regionalismos, extranjerismos, etc.) (<https://bit.ly/3vOtqGu>). Las palabras que se asignarán a los alumnos serán las siguientes: *ciberataque*, *funar*, *deseleccionar*, *spoiler*, *porfa*, *googlear*, *cubre bocas* y *barista*. Después, se explicará a los alumnos que habrán de ponerse en la piel de los lexicógrafos de la RAE y proponer una entrada de diccionario para alguna de estas palabras a la institución en formato digital atendiendo solo a las formas simples del lema.

Para ello, se les proporcionará una *Guía para la elaboración de entradas* (Anexo III.C), que se leerá en conjunto en clase con el fin de aclarar las dudas que puedan surgir antes de que empiecen a elaborar la entrada de manera independiente. El tiempo restante de la clase y la siguiente sesión al completo irán destinadas a que los alumnos aborden esta elaboración, completándolo con trabajo autónomo desde casa. El docente irá acompañando a los distintos equipos de lexicógrafos guiando el proceso de elaboración de la entrada y resolviendo las posibles dudas que se susciten.

#### 4.6.6. Séptima sesión: Reflexión sobre la propuesta

Cada grupo compartirá con el resto de los equipos su entrada de diccionario, defendiendo su propuesta como si fueran lexicógrafos: explicando el porqué de las informaciones incluidas y la pertinencia de su inclusión al diccionario. Una vez leídas todas las entradas, se pretende llevar al aula una discusión sobre la importancia de conocer las palabras que, aunque no estén recogidas, son habituales en distintos intercambios comunicativos (algún extranjerismo, algún regionalismo, alguna palabra adscrita a un grupo social concreto); los recursos de que se dispone para acceder a ellas (corpus, diccionarios, la red, etc.) y cómo el manejo eficaz de estos amplía nuestro conocimiento sobre estas palabras, considerando que conocer una palabra excede saber qué significa. Finalmente, se les pedirá que cumplimenten un cuestionario *online* en sus casas mediante el que valorarán algunos aspectos de la secuencia y su experiencia y valoración del uso del *CORPES* (<https://bit.ly/3wsYonx>).

#### 4.7. Evaluación

En este trabajo partimos de la idea de que evaluar no consiste únicamente en calificar un producto final, sino que una correcta evaluación ha de “ser un álbum de recortes en el que se van añadiendo instantáneas de todo un proceso de aprendizaje” (Wiggins y McTighe, 2015 citado en Moral, 2019, p. 117). Con tal fin, combinaremos una evaluación formativa y una evaluación sumativa.

La evaluación formativa, integrada en las rutinas de aula y desarrollada en el tiempo, nos permitirá adoptar medidas de carácter inmediato, en caso de ser necesario, para mejorar o regular el proceso de enseñanza-aprendizaje (Prado, 2016, pp. 99-100) y valorar el progreso de los alumnos, proporcionándoles el *feedback* oportuno para aproximarlos a los objetivos de aprendizaje planteados (Moral, 2019, p. 110). Para abordar esta forma de evaluación emplearemos técnicas informales como la observación (para la sesión de reflexión) y las preguntas de revisión al inicio de cada sesión acerca de las nociones teóricas (descripción del *DLE*, las casillas para las consultas básicas) y también técnicas formales como la revisión del trabajo individual escrito de los estudiantes (hoja de preguntas del aula de informática y actividad sobre la microestructura de *amarrar*).

La evaluación sumativa que emplearemos con el fin de verificar el logro alcanzado tras el proceso de aprendizaje, dada la naturaleza práctica de la secuencia didáctica, será una evaluación de desempeño, la cual “requiere que el alumno elabore un producto para demostrar su conocimiento o habilidad” (Moral, 2019, p. 119). En concreto, nos serviremos de una rúbrica (Anexo IV) para evaluar la entrada de diccionario elaborada por los grupos de alumnos; actividad que integra la parte más representativa de los objetivos-contenidos de la secuencia.

A lo anterior ha de añadirse que los estudiantes participarán en la evaluación del proceso de enseñanza-aprendizaje. Los alumnos habrán de responder a un cuestionario *online* con el que valorarán distintos aspectos de la secuencia (materiales, explicaciones, actividades), su experiencia de aprendizaje (dificultad, logro de objetivos) y el interés de los contenidos trabajados.

#### 4.8. Atención a la diversidad

Las medidas de atención a la diversidad han de estar ajustadas a cada contexto particular de aplicación de la propuesta con el fin de adecuarse correctamente a las distintas necesidades, capacidades e intereses del alumnado. En la propuesta didáctica se han planteado dos medidas de atención a la diversidad de carácter general. Una de ellas es la de comprobar que todos los estudiantes tienen acceso a los materiales digitales fuera del aula, sirviéndonos de los programas de préstamo de material informático de los centros, dando a conocer bibliotecas comunitarias, etc. La segunda es la de diseñar actividades de aula mediante el modelo de trabajo en grupo, el cual requiere “considerar distintos puntos de vista y colaborar con compañeros que presentan distintas capacidades” (Moral, 2019, p. 162) o habilidades para alcanzar unas metas comunes. Así, estableciendo grupos heterogéneos en cuanto al nivel de desempeño en la materia, talentos o capacidades, los alumnos podrán ayudarse entre sí para la realización de las tareas.

Junto a estas medidas, antes de la aplicación de la propuesta, se habrá de analizar el perfil concreto de los alumnos con los que se aplicará, considerando si entre estos hay alumnos con necesidades específicas de apoyo educativo para, en caso de ser necesario, realizar las adaptaciones oportunas, que permitan la realización de las tareas propuestas.

En el contexto concreto en el que se aplicó esta propuesta, uno de los alumnos estaba diagnosticado como TDAH, por lo que se insistió en verificar su comprensión de las exposiciones mediante preguntas y observación en el aula, se proporcionaron explicaciones individuales durante los momentos de trabajo autónomo en el aula y se hizo hincapié en la revisión de sus tareas.

## **5. Aplicación parcial de la propuesta didáctica**

La secuencia didáctica desarrollada en los apartados anteriores solo pudo aplicarse parcialmente durante el periodo de prácticas de la titulación. Concretamente, se llevó al aula una versión reducida de la sesión 3 (incorporando parte de las sesiones 1 y 2) y la sesión 4 al completo. La secuencia didáctica no pudo desarrollarse en su totalidad, ya que los contenidos abordados en esta propuesta no estaban previstos en la programación didáctica del departamento de Lengua del centro para ese periodo del curso. Sin embargo, los resultados de la aplicación parcial de la propuesta pueden servirnos para valorar su viabilidad y pertinencia.

El centro público de ESO y Bachillerato donde se realizaron las prácticas y que constituye el contexto de aplicación de la secuencia contaba con dos aulas de informática de las que se pudo hacer uso, además de un sistema de préstamo de material informático para los alumnos. El grupo de 4º de la ESO compuesto por 31 alumnos con el que se aplicó la propuesta mostraba, por lo general, un buen desempeño en la materia. Cabe resaltar que uno de los alumnos de este grupo, como se ha señalado, estaba diagnosticado como TDAH y requería de cierta atención individualizada.

Dado que uno de los objetivos principales en la aplicación de la propuesta era valorar la viabilidad de incorporar el uso del *CORPES* en este nivel educativo, se optó por dedicar las dos sesiones de las que se disponía al conocimiento y manejo de esta herramienta. En la primera sesión, a partir del modelo de enseñanza directa descrito, se expuso una definición y somera descripción del *CORPES* en comparación con el *DLE* como forma de facilitar su comprensión y resaltar su utilidad; además de hacer un breve repaso por la utilidad de las casillas vinculadas a las consultas básicas (sesión 3 y parte de las sesiones 1 y 2). Para esto último, además, se alojó en la plataforma virtual del

centro la imagen interactiva-manual de uso para que los estudiantes pudieran repasar las utilidades de manera autónoma antes de la sesión de práctica.

Finalmente, en la sesión en el aula de ordenadores, se resolvieron y pusieron en común todas las preguntas de comprobación y ampliación planteadas (mostramos un ejemplo de realización en el Anexo V.A) y se realizó, además, un cuestionario *online* de evaluación de la secuencia y el recurso. Ha de hacerse notar, no obstante, que esta última sesión se vio alterada por la celebración de unas jornadas culturales en el centro como cierre del periodo lectivo previo a las vacaciones de Semana Santa y solo asistieron 17 alumnos a clase. Esta circunstancia, sin embargo, permitió ofrecer una atención más individualizada a todos los alumnos y dialogar sobre los ejercicios realizados. La valoración de estas dos sesiones se integrará, a continuación, con las conclusiones.

## **6. Conclusiones**

A lo largo de este trabajo, se ha tratado de exponer la pertinencia de incorporar los corpus textuales –y, en concreto, el *CORPES*– como recurso didáctico para el desarrollo de los contenidos y destrezas propios de la materia de Lengua Castellana y Literatura en el segundo ciclo de Secundaria. Aunque se trata de un recurso de indudable interés para trabajar aspectos de los distintos bloques del currículo –cuestión en la que podría profundizarse en futuras investigaciones–, aquí nos hemos centrado en su utilidad como herramienta de consulta lingüística, especialmente, en el ámbito del léxico. En este ámbito se ha expuesto cómo, integrados con los diccionarios (puesto que estos constituyen una herramienta ineludible para el aprendizaje del léxico), los corpus permiten ampliar el conocimiento de las palabras y, por tanto, reforzar la competencia léxica del alumnado. Este refuerzo se debe, en esencia, a que los corpus muestran las palabras en un contexto lingüístico y discursivo y en un entorno comunicativo real, lo cual proporciona pistas acerca de cómo se combinan las palabras y cuáles son sus restricciones de uso.

La aplicación de estos recursos en la enseñanza de lenguas ya está, en cierta medida, extendida en el ámbito de la enseñanza del inglés como primera o segunda lengua, cuenta con varias propuestas en el ámbito del español como segunda lengua y con algunas otras, menos frecuentes, en el ámbito del español como lengua materna. De todas

las propuestas de uso didáctico de esta herramienta, aquí hemos optado por sugerir una secuencia didáctica orientada a un grupo de 4º de la ESO y dirigida al conocimiento y manejo del *CORPES* integrado con el *DLE*, con el fin de facilitar su aprendizaje y resaltar sus beneficios y limitaciones. En cuanto a las aproximaciones metodológicas, hemos optado por introducir de manera pautada el uso de esta herramienta para que, en un futuro, los estudiantes puedan abordar investigaciones lingüísticas con autonomía. Así, nuestra propuesta comenzaba con una introducción y contraste teórico de ambos recursos, continuaba con una sesión de práctica guiada destinada a que los alumnos desarrollasen la suficiente destreza en el uso de estos, llamando la atención sobre sus diferencias, y finalizaba con la integración de ambos recursos con el objetivo de crear una entrada de diccionario de ciertas palabras ausentes en el lecionario del *DLE* recurriendo al trabajo colaborativo.

Si bien esta propuesta no se pudo aplicar íntegramente en el centro educativo donde se realizaron las prácticas, sí se pudo poner en práctica la parte dirigida específicamente al conocimiento y uso del *CORPES*. Con estas dos sesiones se pretendía probar la viabilidad de la incorporación de este recurso en secundaria y, en efecto, los resultados de esta aplicación nos han permitido extraer varias conclusiones. La primera sesión teórica estuvo marcada por la limitación del tiempo del que se disponía. Esto supuso la dificultad de condensar en una sesión los contenidos mínimos necesarios para comprender la utilidad y pertinencia del *CORPES*, así como su contraposición al *DLE* (exponiendo, para ello, algunas nociones sobre este repertorio). Pese a tratar de exponer los contenidos esenciales, dado que los estudiantes desconocían en gran medida la microestructura del *DLE*, daba la impresión de que la sesión les resultaba una carga excesiva. Esto nos permite extraer como conclusión que, aparte del “manejo” del diccionario como herramienta de consulta –como prioriza el currículo–, habría de reforzarse el “conocimiento” de este para que los alumnos sepan aprovecharlo en profundidad. Asimismo, esto nos permitió resaltar la pertinencia de dedicar dos sesiones al conocimiento gradual de esta herramienta apostando por un modelo de aprendizaje dialógico y participativo.

La realización de la sesión práctica en el aula de ordenadores nos permitió constatar que, a pesar de lo apuntado, los alumnos eran capaces de realizar consultas básicas en el *CORPES* y contrastar la información obtenida mediante estas con la

proporcionada por el *DLE*. Todos los alumnos que asistieron pudieron resolver finalmente de manera correcta las preguntas planteadas en la hoja de ejercicios, aunque tres de ellos mostraron una mayor dificultad en el manejo de la herramienta (uno de ellos, el alumno con diagnóstico TDHA), por lo que se les proporcionó una ayuda individualizada. Asimismo, pudimos extraer varias conclusiones a partir del cuestionario *online* (<https://bit.ly/3sJun1X>) que se les proporcionó durante esta sesión para valorar algunos aspectos de la secuencia y el recurso<sup>12</sup>. La mayor parte de los alumnos valoraba positivamente las actividades realizadas en el aula y declaraba bastante o total conformidad con la consideración del *CORPES* como un recurso útil para el aprendizaje de la lengua (15 alumnos). Sin embargo, solo 12 alumnos decían estar bastante o totalmente convencidos de volver a utilizarlo para resolver dudas con respecto a la lengua. Por último, la mayoría de los alumnos no consideraba que el uso del *CORPES* le hubiera resultado difícil, aunque, de acuerdo con lo observado, tres de ellos señalaron bastante o cierta dificultad.

En definitiva, las consideraciones expuestas hasta el momento nos permiten concluir que la incorporación de los corpus textuales en el segundo ciclo de Secundaria al conjunto de herramientas de consulta de lingüística que los estudiantes manejan para la consulta del léxico –como los diccionarios o glosarios– es, en primer lugar, viable y, en segundo lugar, deseable para potenciar el desarrollo de la competencia léxica de los alumnos. En efecto, el reflejo de las palabras interconectadas en atmósferas discursivas y contextuales reales, tal y como habitan en los corpus textuales, junto a su rigurosa descripción en los diccionarios, fruto de un esmerado estudio lingüístico o lexicográfico, facilitará que nuestro alumnado alcance un conocimiento integral del léxico del español.

---

<sup>12</sup> En el Anexo V.B presentamos más detalladamente las respuestas del cuestionario, realizado por todos los alumnos que asistieron a clase ese día.

## 7. Bibliografía

- Ahumada, I. (1989). *Aspectos de lexicografía teórica: aplicaciones al diccionario de la Real Academia Española*. Universidad de Granada.
- Alonso-Ramos, M. (2017). Diccionarios combinatorios. *Estudios de Lingüística del Español (ELiEs)*, 38, 173-201. <https://bit.ly/39nEF19>
- Ávila, M.<sup>a</sup> C. (2010). Estadística y lingüística de corpus: implicaciones pedagógicas en la enseñanza y el aprendizaje del léxico. *Cauce: Revista Internacional de Filología, Comunicación y sus Didácticas*, 32-33, 163-175. <https://bit.ly/3wqKEtO>
- Battaner, P. (2019). El diccionario como punto de encuentro en la descripción de la lengua española. *Boletín de Información Lingüística de la Real Academia Española*, 13, 3-20. <https://bit.ly/3wmv4z2>
- Bordón, T. (2019). Gramática pedagógica. En J. Muñoz-Basols, E. Gironzetti y M. Lacorte (Eds.) *The Routledge Handbook of Spanish Language Teaching. Metodologías, contextos y recursos para la enseñanza de español L2* (pp. 213-228). Routledge Handbooks.
- Cortázar, J. (2013). *Rayuela (50<sup>a</sup> ed. conmemorativa)* (Trabajo original publicado en 1963). Alfaguara.
- Campos, M. y Pérez, J. A. (2003). El diccionario y otros productos lexicográficos. En A. M. Medina Guerra (Coord.) *Lexicografía española* (pp. 53-78). Ariel.
- Castillo, M.<sup>a</sup> A. (2013). La macroestructura del diccionario. En A. M. Medina Guerra (Coord.) *Lexicografía española* (pp. 79-102). Ariel.
- Decreto 48/2015, de 14 de mayo, del Consejo de Gobierno, por el que se establece para la Comunidad de Madrid el currículo de la Educación Secundaria Obligatoria. Boletín Oficial de la Comunidad de Madrid, núm. 118, de 20 de mayo de 2015, pp. 10 a 309. <https://bit.ly/3wtLkhT>
- Garriga, C. (2003). La microestructura del diccionario: las informaciones lexicográficas. En A. M. Medina Guerra (Coord.) *Lexicografía española* (pp. 103-126). Ariel.
- Haensch, G. y Omeñaca, C. (2004). *Los diccionarios del español en el siglo XXI (2.<sup>a</sup> ed. corregida y aumentada)*. Ediciones Universidad de Salamanca.
- Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. Boletín Oficial del Estado, núm. 340, de 30 de diciembre de 2020, pp. 1 a 86. <https://bit.ly/3FShwQf>

- Martín, R. A. (2013). Práctica de los diccionarios y corpus electrónicos en la enseñanza de la lengua española. *Didáctica, Innovación y Multimedia*, 26, 1-11. <https://bit.ly/32APdGI>
- Martínez, I. (2018). Lingüística de corpus y gramática inductiva: una aproximación. *Tonos Digital, Revista Electrónica de Estudios Filológicos*, 35, 1-20. <https://bit.ly/3wqwYzT>
- Moral, C. (2019). *Competencias para el diseño y desarrollo de experiencias de aprendizaje en la formación del profesorado*. Síntesis.
- Ortiz, D. (2015). El constructivismo como teoría y método de enseñanza. *Sophia: Colección de Filosofía de la Educación*, 19, 93-110. <https://bit.ly/3lfufmM>
- Parodi, G. (2010). *Lingüística de corpus: de la teoría a la empiria*. Iberoamericana Editorial Vervuert.
- Parodi, G. y Burdiles, G. (2019). Corpus y bases de datos. En J. Muñoz-Basols, E. Gironzetti y M. Lacorte (Eds.) *The Routledge Handbook of Spanish Language Teaching. Metodologías, contextos y recursos para la enseñanza de español L2* (pp. 596-612). Routledge Handbooks.
- Prado, J. (2016). *Didáctica de la lengua y la literatura para educar en el siglo XXI. (Tercera edición)*. La Muralla.
- Real Academia Española (s.f.-a). *Corpus del español del siglo XXI. Manual de consulta en línea (v 0.94 beta)* [Archivo PDF]. <https://bit.ly/3qpiQ9F>
- Real Academia Española (s.f.-b). *Diccionarios. Diccionario de la lengua española*. <https://bit.ly/3Pnf4pm>
- Real Academia Española (s.f.-c). *La institución*. <https://bit.ly/3PrpX9Q>
- Real Academia Española (2013). *Corpus del español del siglo XXI (CORPES). Descripción del sistema de codificación. Libros y prensa (texto revisado de 2006 a 2020)* [Archivo PDF]. <https://bit.ly/3J0MVCO>
- Real Academia Española (2014a). *Diccionario de la lengua española (23ª edición)*
- Real Academia Española (2014b). *Dossier de prensa. Diccionario de la lengua española. Vigésimotercera edición*. [Archivo PDF]. <https://bit.ly/3a0w5W3>
- Real Academia Española (2018). *Corpus del español del siglo XXI (CORPES). Descripción del sistema de codificación Textos orales* [Archivo PDF]. <https://bit.ly/3oScorp>
- Real Academia Española (2021a). *Acerca de la anotación de CORPES* [Archivo PDF]. <https://bit.ly/3OXymno>

- Real Academia Española (2021b). Diccionario de la lengua española, 23.<sup>a</sup> ed., [versión 23.5 en línea]. <https://dle.rae.es/>
- Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato. Boletín Oficial del Estado, núm. 3, de 3 de enero de 2015, pp. 169 a 546. <https://bit.ly/3PktmXL>
- Rojo, G. (2016). Corpus textuales del español. En J. Gutiérrez Rexach (Coord.) *Enciclopedia de Lingüística Hispánica*, Vol. 2, 285-296.
- Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Routledge.
- Ruhstaller, S. (2003). Las obras lexicográficas de la Academia. En A. M. Medina Guerra (Coord.) *Lexicografía española* (pp. 235-262). Ariel.
- Seco, M. (13 de marzo de 2014). *En el país de las palabras* [Sesión de conferencia]. Celebración del Día de la Fundación pro-RAE. <https://bit.ly/3a47zDE>
- Seco, M., Andrés, O. y Ramos, G. (1999). *Diccionario del español actual*. Aguilar
- Tolchinsky, L. (2014). El uso de corpus lingüísticos como herramienta pedagógica. *Textos de didáctica de la lengua y la literatura*, 65, 9-17. <https://bit.ly/3mtL7a7>
- Torruella, J. (2017). *Lingüística de corpus: génesis y bases metodológicas de los corpus (históricos) para la investigación en lingüística*. Peter Lang Edition.

## 8. Anexos

### 8.1. Anexo I. Material de apoyo a la explicación de los corpus

Atributos:

[autor\_título\_secundario], que recoge el autor del texto anidado. En el caso de la prensa, es el lugar previsto para el autor del artículo, si se conoce; en caso contrario, el atributo queda vacío.

Sintaxis:

```
<título_secundario autor_título_secundario="Benitez, Wilfrido">"Media docena de delincuentes asalta una tienda"</título_secundario> [Prensa]
<título_secundario autor_título_secundario="Muñiz, Carlos ... [et al.]">El proyecto geopirateria: el caso del Ecuador</título_secundario> [Libro]
```

Imagen 1. Codificación del título y autor de un documento en CORPES. Tomado de RAE (2013, p.14)

The screenshot shows the top part of the CORPES website. At the top left is the logo of the Real Academia Española. The main title is "Corpus del Español del Siglo XXI (CORPES)". To the right, it says "Versión beta (0.94)". Below the title is a navigation menu with items: "Concordancias", "Coapariciones", "Configuración", "Ayuda", "Estadística", "Modo de cita", "Sugerencias", and "Preguntas frecuentes". Below the menu are search filters: "Lema" (set to "anexo"), "Forma", "Clase de palabra", "Grafía original" (checkbox), "+ Subcorpus", and "+ Proximidad". At the bottom of the search area are buttons for "+", "-", "Concordancia", "Estadística", and "Nueva consulta".

4.525 casos en 2.012 documentos.

Imagen 2. Pantalla de la zona superior de la ventana inicial de consulta. Tomado de CORPES, v. beta 0.94

4.525 casos en 2.012 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por:	Año ascendente	sin crit
<input type="checkbox"/> 1 2001 Col.	La presente guía se complementa con un cuadernillo <b>anexo</b> de ejercicios técnicos, estudios aplicados; una sección con pieza			
<input type="checkbox"/> 2 2001 Arg.	En el mapa <b>anexo</b> (página 22) se muestran las áreas dominadas y zonas de influen			
<input type="checkbox"/> 3 2001 Arg.	Ver en el <b>Anexo</b> algunos ejemplos de comidas adecuadas.			
<input type="checkbox"/> 4 2001 Chile	Noviembre del presente mes bajo el título de "Desarrollo Nacional y Estrategia Naval" ( <b>Anexo</b> ).			
<input type="checkbox"/> 5 2001 Chile	llegaron a las puertas de la iglesia, a cuyo costado se hallaba una construcción <b>anexa</b> que servía de casa parroquial. Al instante apareció Betsabé quier			
<input type="checkbox"/> 6 2001 Chile	-Será suficiente y, más aún si las construcciones <b>anexas</b> se encuentran también en buenas condiciones -confirmó Durand			
<input type="checkbox"/> 7 2001 Chile	Pauti, con Blanche y Durand. Encontramos la vieja factoría y dos construcciones <b>anexas</b> . Todas se encuentran en buen estado, y estamos dispuestos a tr			
<input type="checkbox"/> 8 2001 Chile	comunidad. Los árboles de papaya y de otros frutales, que rodeaban las construcciones <b>anexas</b> a la antigua factoría, llegaron incluso a hacer pensar a algunos q			
<input type="checkbox"/> 9 2001 Chile	la colina desde donde pudieron ver la antigua factoría y las dos construcciones <b>anexas</b> : estaban de vuelta en casa. Tan pronto sus cabezas se asomaro			
<input type="checkbox"/> 10 2001 Chile	familiares colgados en el vestíbulo; la primera posada, sus remodelaciones, los edificios <b>anexos</b> y altillos, tatarabuelos, bisabuelos, abuelos, primos...			
<input type="checkbox"/> 11 2001 Méx.	iniciativa; debía dejarla hacer. Conocía que Luciana y su familia vivían en una casa <b>anexa</b> y comunicada al hotel, así que con seguridad la vería sin tener q			
<input type="checkbox"/> 12 2001 Guat.	En la intimidad, Johnson era remoto y prudente para evitar todos los roces <b>anexos</b> a la tediosa vida doméstica. A veces hablaba solo y como esper			
<input type="checkbox"/> 13 2001 Nic.	<b>Anexos</b>			
<input type="checkbox"/> 14 2001 Nic.	uso de la Alcaldía que actualmente ocupa, iniciándose la construcción de un nuevo <b>anexo</b> en 1996 finalizándose a principios de 1998, pero que aun no pre			
<input type="checkbox"/> 15 2001 Nic.	tras él, recorrieron varias calles, y en una calle que tenía contacto con calles <b>anexas</b> , se veían como caminitos, el extraño entró en uno de esos camir			

Imagen 3. Pantalla de la zona intermedia de la ventana inicial de consulta. Tomado de CORPES, v. beta 0.94

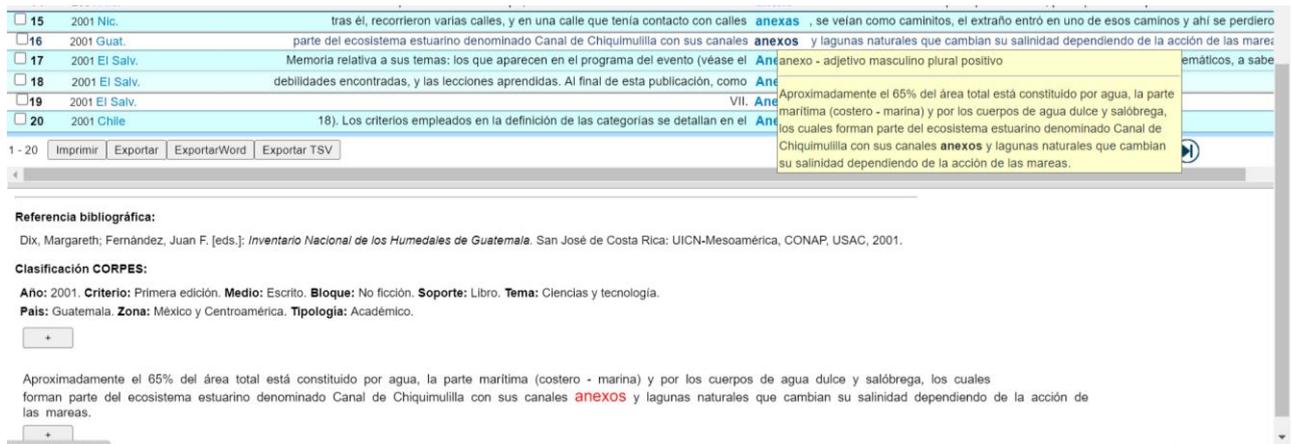


Imagen 4. Pantalla de la zona inferior de la ventana inicial de consulta. Tomado de CORPES, v. beta 0.94



Imagen 5. Lista de opciones de la casilla "Clase de palabra". Tomado de CORPES, v. beta 0.94.



Imagen 6. Búsqueda de conceptos lingüísticos abstractos a partir de la casilla "Clase de palabra". Tomado de CORPES, v. beta 0.94


**Corpus del Español del Siglo XXI**  
**(CORPES)** Versión beta (0.94)

[Concordancias](#) | [Coapariciones](#) | [Configuración](#) | [Ayuda](#) | [Estadística](#) | [Modo de cita](#) | [Sugerencias](#) | [Preguntas frecuentes](#)

Lema:  Forma:  período:  Clase de palabra:   Grafía original

27.850 casos en 12.601 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente
<input type="checkbox"/> 1 2001 Chile	han dormido tirados en la playa, han comido e incluso se han reído, y comienza un <b>período</b> gélido, un período aparentemente normal pero dom	
<input type="checkbox"/> 2 2001 Chile	en la playa, han comido e incluso se han reído, y comienza un período gélido, un <b>período</b> aparentemente normal pero dominado por unos día	
<input type="checkbox"/> 3 2001 Chile	que hice en mi primer club europeo: salí de putas y así fui capeando la lesión, el <b>período</b> de recuperación, la soledad. ¿Que si me acostumbi	
<input type="checkbox"/> 4 2001 Chile	formando ahí su extraño hogar. El es un vicioso jugador de Pool que pasa por un mal <b>período</b> y ella es una Prostituta que debe pagar con su cuer	
<input type="checkbox"/> 5 2001 Chile	Otro hito que marcó la práctica muralista de este <b>período</b> fue un mural pintado en 1971 por Roberto Matta, er	

Imagen 7. Búsquedas selectivas aplicando la opción "Grafía original". Tomado de CORPES, v. beta 0.94

[Concordancias](#) | [Coapariciones](#) | [Configuración](#) | [Ayuda](#) | [Estadística](#) | [Modo de cita](#) | [Sugerencias](#) | [Preguntas frecuentes](#)

Lema:  Forma:  Clase de palabra:   Grafía original

Proximidad

Lema: <input type="text" value=""/> Forma: <input type="text" value="se"/>	Clase de palabra: <input type="text" value=""/>	<input checked="" type="radio"/> Distancia <input type="radio"/> Intervalo	1	<input type="radio"/> Izquierda <input checked="" type="radio"/> Derecha <input type="radio"/> Izquierda o derecha
Lema: <input type="text" value=""/> Forma: <input type="text" value="un"/>	Clase de palabra: <input type="text" value=""/>	<input checked="" type="radio"/> Distancia <input type="radio"/> Intervalo	2	<input type="radio"/> Izquierda <input checked="" type="radio"/> Derecha <input type="radio"/> Izquierda o derecha
Lema: <input type="text" value=""/> Forma: <input type="text" value=""/>	Clase de palabra: <input type="text" value="sustantivo"/>	<input checked="" type="radio"/> Distancia <input type="radio"/> Intervalo	3	<input type="radio"/> Izquierda <input checked="" type="radio"/> Derecha <input type="radio"/> Izquierda o derecha

Dirección con respecto al lema o la f:

Número:  Género:  Tipo:

25 casos en 25 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente	sin criterio
<input type="checkbox"/> 1 2001 Esp.	tonta, ¿verdad? —continuó Fermin—. A usted lo que le gustaría es estar por ahí, <b>marcando</b> se un chotis. Porque tiene usted planta de bailarina, se lo debe de decir todo e		
<input type="checkbox"/> 2 2001 Esp.	guitarra, pues a tocar mucho. Si quieren ser famosos, ésa es otra carrera. Hay que <b>marcar</b> se un camino y hacer lo que te gusta", explica Gerardo Núñez.		
<input type="checkbox"/> 3 2001 Esp.	todavía Carlos Larrañaga, sin embargo, su ex mujer, María Luisa Merlo, no dudó en <b>marcar</b> se un baile en La Venta del Toro al son de las voces de la cantaora La Gallina y		
<input type="checkbox"/> 4 2003 R.Dom.	partir, estando asomada a la ventana que miraba al río, Toño acunaba un solo sueño: <b>marcar</b> se un destino distinguido como lo había hecho su abuelo. Cosechar fama y hor		
<input type="checkbox"/> 5 2003 Esp.	puede aferrarse todavía a ese guión que ha impuesto al tiempo. Gracias a él puede <b>marcar</b> se un farol de indiferencia si alguien entra en el bar, puede reír una ocurrencia		

Imagen 8. Búsquedas de estructuras sintácticas a partir de la opción "Proximidad". Tomado de CORPES, v. beta 0.94

[Concordancias](#) | [Coapariciones](#) | [Configuración](#) | [Ayuda](#) | [Estadística](#) | [Modo de cita](#) | [Sugerencias](#) | [Preguntas frecuentes](#)

Lema:  Forma:  Clase de palabra:   Grafía original

Subcorpus

Título: <input type="text" value=""/>	Autor: <input type="text" value=""/>	Fecha de clasificación: 2010 - 2022
Origen: <input type="text" value="Todos"/> América <input type="text" value="España"/>	Zonas lingüísticas: <input type="text" value="Chilena México y Centroamérica Río de la Plata"/>	Países: <input type="text" value="Todos"/> Argentina <input type="text" value="Paraguay"/>
Medio: <input type="text" value="Todos"/> Escrito <input type="text" value="Oral"/>	Bloque: <input type="text" value="Todos"/> Ficción <input type="text" value="No ficción"/>	Soporte: <input type="text" value="Libro"/> Miscelánea <input type="text" value="Prensa"/>
Tema: <input type="text" value="Todos"/> Actualidad, ocio y vida cotidiana <input type="text" value="Artes, cultura y espectáculos"/>		Tipología: <input type="text" value="Todos"/> Académico <input type="text" value="Carta al director"/>

Marcas:  Cursiva  Negrita  Subrayado  Versalitas

10 casos en 10 documentos.

REF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente	sin criterio
<input type="checkbox"/> 1 2011 Arg.	Lejos de <b>amilanar</b> se, Plaza dejó una mejor impresión, porque terminó jugando en terreno ajeno		
<input type="checkbox"/> 2 2011 Par.	- Anike jajaivy agóstope, ¡cháke la muerte repentina!... (No <b>amilanar</b> se a agosto, ¡cuidado de la muerte repentina!)		
<input type="checkbox"/> 3 2011 Ur.	Lejos de <b>amilanar</b> se el delincuente se bajó de la moto y tiró su arma a una canaleta. Metros adel		
<input type="checkbox"/> 4 2012 Arg.	de sus piernas. Sin embargo, antes de ser un excepcional campeón de tenis, no se <b>amilanó</b> para integrarse con sus amigos y jugar al fútbol con apenas seis años: "Me m		
<input type="checkbox"/> 5 2012 Arg.	ella. Sólo los fabricantes conocían los secretos de tan complejo aparato. Lejos de <b>amilanar</b> se por aquellas restricciones, los programadores locales tomaron cartas en el i		
<input type="checkbox"/> 6 2013 Ur.	El otro gladiador de los cuartos de final será Ferrer, quien no se <b>amilanó</b> cuando perdió el primer set ayer y ganó los tres siguientes. El primero de form		

Imagen 9. Búsquedas específicas a partir de la opción "Subcorpus". Tomado de CORPES, v. beta 0.94

REAL ACADEMIA ESPAÑOLA  
Corpus del Español del Siglo XXI (CORPES) Versión beta (0.94) Cerrar sesión

Concordancias | Coapariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerecias | Preguntas frecuentes

Lema:  Forma:  voy Clase de palabra:   Grafía original + Subcorpus + Proximidad

Concordancia Estadística Nueva consulta

68.941 casos en 16.952 documentos.

REF. (Clasificación, país) CONCORDANCIA Ordenar por: Año ascendente | sin criterio

<input type="checkbox"/> 1	2001 P.Rico.	Si, voy a seguir tocando en el futuro ().
<input type="checkbox"/> 2	2001 P.Rico.	Mañana mismo ir - verbo primera persona singular presente indicativo
<input type="checkbox"/> 3	2001 Arg.	53. Prefiero más que llegar pensar que ya Si, voy a seguir tocando en el futuro ().
<input type="checkbox"/> 4	2001 Arg.	23. En el tema ?Ahí

Imagen 10. Información morfosintáctica y lematización de una palabra. Tomado de CORPES, v. beta 0.94

REAL ACADEMIA ESPAÑOLA  
Corpus del Español del Siglo XXI (CORPES) Versión beta (0.94) Cerrar sesión

Concordancias | Coapariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerecias | Preguntas frecuentes

Lema:  Forma:  voy Clase de palabra:   Grafía original + Subcorpus + Proximidad

Concordancia Estadística Nueva consulta

en 16.952 documentos.

Clasificación, país) CONCORDANCIA Ordenar por: Año ascendente | sin criterio

P.Rico.	Si, voy a seguir tocando en el futuro ().	Año ascendente
P.Rico.	Mañana mismo voy a Valencia, a la gala de	Año descendente
Arg.	53. Prefiero más que llegar pensar que ya voy llegando. Es una de las	Pivote lema
Arg.	23. En el tema ?Ahí voy ? de Fito menciona a ?An	Pivote forma
Arg.	No voy al cine.	Pivote clase de palabra
Arg.	Siempre me traman ríos años, entre ríos en ríos en aunque haya en el medio, voy a tratar de mantener los	Autor
		Título

Imagen 11. Criterios de ordenación. Tomado de CORPES, v. beta 0.94

REAL ACADEMIA ESPAÑOLA  
Corpus del Español del Siglo XXI (CORPES) Versión beta (0.94) Cerrar sesión

Concordancias | Coapariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerecias | Preguntas frecuentes

Lema:  Forma:  amilanar Clase de palabra:   Grafía original + Subcorpus + Proximidad

Concordancia Estadística Nueva consulta

Frecuencia absoluta: 144 Documentos: 132 Frecuencia normalizada: 0,41 casos por millón

Distribución Zona

Zona	Freq	Fnorm.
España	58	0,47
Caribe continental	26	0,63
Andina	12	0,44
México y Centroamérica	12	0,18
Río de la Plata	12	0,26
Antillas	10	0,44
Chilena	6	0,29
Estados Unidos	4	0,98
Guinea Ecuatorial	4	4,42

1 - 9 de 9 página: 1

Distribución Zona

Imagen 12. Ventana de "Estadística". Tomado de CORPES, v. beta 0.94

REAL ACADEMIA ESPAÑOLA  
Corpus del Español del Siglo XXI (CORPES) Versión beta (0.94) Cerrar sesión

Concordancias | Coapariciones | Configuración | Ayuda | Estadística | Modo de cita | Sugerecias | Preguntas frecuentes

Lema:  Clase de palabra:  Tema: (Todos) Actualidad, ocio y vida cotidiana; Artes, cultura y espectáculos

Origen: (Todos) América España

Coapariciones Nueva consulta

	Clase	Freq	MI	LL SIMPLE	T-SCORE
naipe	sustantivo	14	11,4	84,11	3,74
hipótesis	sustantivo	144	11,29	859,51	12,00
posibilidad	sustantivo	347	9,64	1.739,86	18,62
opción	sustantivo	132	9,35	632,78	11,48
alternativa	sustantivo	52	8,83	232,30	7,21
cifra	sustantivo	52	8,14	210,60	7,21
nombre	sustantivo	156	8,04	626,05	12,48
carta	sustantivo	63	8,02	250,66	7,93
repartir	verbo	11	7,22	38,46	3,31
descartar	verbo	11	7,11	37,74	3,31
candidato	sustantivo	26	7,0	87,38	5,09
fórmula	sustantivo	12	6,96	40,07	3,46
fecha	sustantivo	29	6,83	94,65	5,38

Imagen 13. Búsqueda de las coapariciones del verbo "barajar". Tomado de CORPES, v. beta 0.94

## 8.2. Anexo II. Material de apoyo a la explicación del diccionario

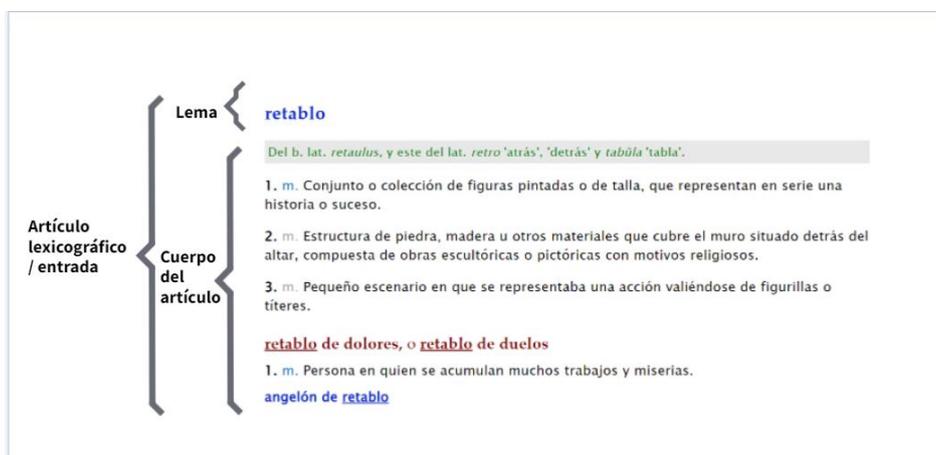


Imagen 1. Esquema de los elementos de la macroestructura del DLE a partir de la entrada de "retablo" en RAE (2021b). Elaboración propia.

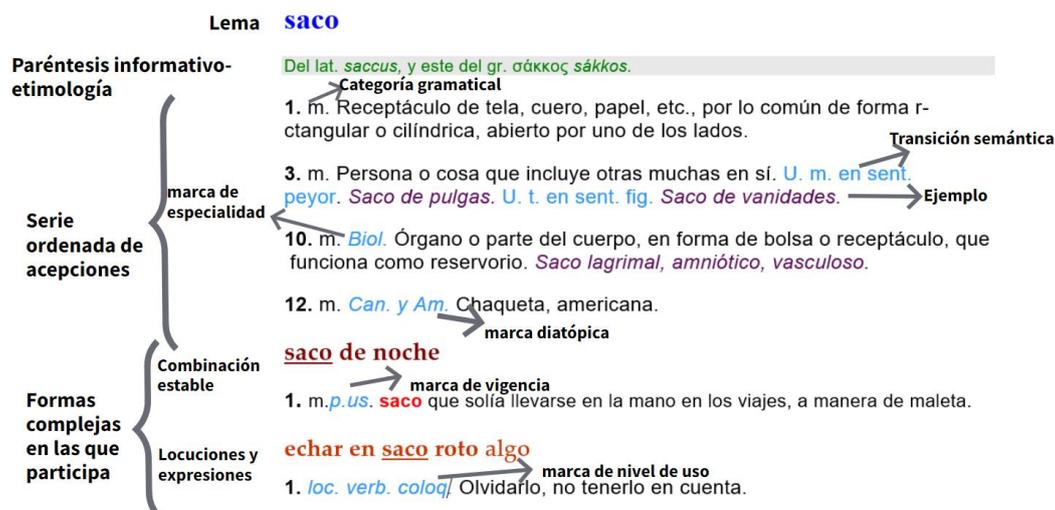


Imagen 2. Esquema de la microestructura del DLE en una reducción de la entrada de 'saco' en RAE (2021b). Elaboración propia.

### 8.3. Anexo III. Materiales de la propuesta didáctica

#### A) Versión simplificada de las entradas de *saco* y de *amarrar* (sesión 2)

##### **saco**

Del lat. *saccus*, y este del gr. *σάκκος sákkos*.

1. m. Receptáculo de tela, cuero, papel, etc., por lo común de forma rectangular o cilíndrica, abierto por uno de los lados.
3. m. Persona o cosa que incluye otras muchas en sí. **U. m. en sent. peyor.** *Saco de pulgas.* **U. t. en sent. fig.** *Saco de vanidades.*
10. m. **Biol.** Órgano o parte del cuerpo, en forma de bolsa o receptáculo, que funciona como reservorio. *Saco lagrimal, amniótico, vasculoso.*
12. m. **Can. y Am.** Chaqueta, americana.

##### **saco de noche**

1. m. **p.us.** **saco** que solía llevarse en la mano en los viajes, a manera de maleta.

##### **echar en saco roto algo**

1. **loc. verb. coloq.** Olvidarlo, no tenerlo en cuenta.

*Imagen 1. Versión simplificada de la entrada de “saco” en RAE (2021b).*

##### **amarrar**

Del fr. *amarre*, y este del neerl. *amarren* 'atar'.

1. tr. Atar y asegurar algo por medio de cuerdas, maromas, cadenas...
6. tr. **jerg. estud.** **empollar** (ll estudiar mucho).
8. tr. **Bol., Ec., El Salv., Hond., Méx., Nic. y Pan.** Vendar o ceñir.
9. **prnl. coloq.** Casarse o convivir maritalmente.
10. **prnl. Col., Méx. y R. Dom.** **embriagarse.**

##### **amarrársela**

1. **loc. verb. Col. y C. Rica.** **embriagarse.**

*Imagen 2. Versión simplificada de la entrada de “amarrar” en RAE (2021b)*

B) Consultas básicas en el *CORPES* (sesión 3)

PREGUNTAS DE COMPROBACIÓN

1. ¿Cuántos casos hay en el *CORPES* del paradigma verbal de *barajar*? ¿Cuántas veces aparece el infinitivo *barajar*?
2. ¿Cuántos casos hay del sustantivo *vino*? ¿Y de la forma *vino* del paradigma verbal de *venir*?
3. ¿Cuántos casos hay de la palabra *Estado* ('institución') desde 2012 hasta 2021? (Usar la opción "Subcorpus" para la fecha)
4. ¿Cuántos casos hay de la **forma** *saco* (sustantivo) en México?
5. ¿Cuándo aparece por primera vez integrada en español la palabra *creepy* en el sentido de 'malrollero' o 'que causa extrañeza, miedo, etc.?'

PREGUNTAS DE AMPLIACIÓN

6. ¿Cuántos casos hay de la forma verbal, no recomendada por la norma, *estuvistes*? ¿En qué tipo de textos es más frecuente? ¿Por qué creéis que es así?
7. Buscad todos los casos de la palabra *frentazo* en noticias de prensa, usando la opción de Subcorpus; concretamente, la opción de soporte y tipología. ¿Qué significado tiene esa palabra en ese contexto? ¿Está recogido en el *DLE*?
8. El diccionario de la RAE recoge las variantes *período* y *periodo*. Usando la opción "Grafía original" y "Estadística", ¿podríais decir cuál es la más usada por los hablantes en general? Desde 2016 hasta ahora, ¿cuál es la forma preferida o más usada por los hablantes?
9. En la canción "Sana Sana", Nathy Peluso dice: "tráeme la navaja filosa". Este adjetivo queda recogido en la RAE como "afilado, que tiene filo", entre otras acepciones. Revisando las estadísticas de esta palabra, ¿creéis que podríais añadirle a la definición alguna marca de la zona de uso de esta palabra?
10. ¿Cuáles son las tres palabras más relacionadas con la palabra "lluvia"? (Usad la opción "Coapariciones", con la ordenación por defecto) ¿Alguna de estas combinaciones aparece en las entradas complejas del *DLE*? ¿Cuál de estas asociaciones os parece lingüísticamente relevante?

### C) GUÍA PARA LA ELABORACIÓN DE ENTRADAS

1. La entrada de diccionario que elaboréis debe ajustarse al formato tipográfico del *DLE* en línea. Recordad que practicamos sobre ello con una imagen interactiva en el aula y con una fotocopia de la entrada de *amarrar* y de *saco*.
2. Proponed una definición de la palabra (o varias, si identificáis más de un valor) a partir de los 15 ejemplos más recientes. Vamos a procurar no utilizar la palabra definida en la definición o definiciones que proporcionemos. Podéis definirla con vuestras palabras o serviros de una búsqueda en *Google* o la web *Fundéu* (<https://bit.ly/3MQh08j>). Todas las palabras propuestas ya son de uso por determinados grupos sociales y, aunque el *DLE* no lo haya hecho, otras personas/diccionarios ya se han encargado de definir las.
3. Los ejemplos, como vimos, no siempre aparecen en el *DLE*, pero son muy importantes para entender el uso de una palabra y también para ver con qué elementos se suele combinar. Proponed un ejemplo en el que se incluya la palabra más relacionada estadísticamente y “lingüísticamente” con la palabra asignada.
4. La etimología, como vimos, puede referirse a su origen grecolatino, o a su proceso de formación: período (Del lat. ...) o contraatacar (*contra* + *atacar*). De nuevo, podéis basaros en búsquedas en *Google*, especialmente en el Observatorio de Palabras de la RAE (<https://bit.ly/3vOtqGu>) y el *Diccionario Etimológico Castellano en línea* (<http://etimologias.dechile.net>) para proponer la de vuestra palabra.
5. Con respecto a las marcas de uso, habéis de destacar su “vigencia de uso” (si es muy poco frecuente), su distribución geográfica (si hay un país/zona en el que se use especialmente), si es de uso especial en algún lenguaje técnico (medicina, tecnología...), y el nivel de lengua al que se asocia (jerga, coloquial).
6. Con respecto a las marcas de categoría gramatical, debéis señalar si es un sustantivo (masculino o femenino), si es un verbo (transitivo, intransitivo o pronominal), adjetivo, interjección/ expresión, etc. Para decidir esta cuestión,

debéis analizar cómo actúa en los ejemplos del corpus, cómo se relaciona con otras palabras: si suele tener complemento directo, si se acompaña de adjetivos/sustantivos masculinos/femeninos, etc.

#### 8.4. Anexo IV. Rúbrica de evaluación de la entrada de diccionario

	Sobresaliente	Notable	Aprobado	Insuficiente
Formato de la entrada	El formato de la entrada se ajusta de manera completa al del <i>DLE</i> en línea. Aparecen el lema y todos los elementos de la microestructura trabajados y requeridos en la disposición adecuada.	El formato de la entrada se ajusta de manera adecuada al del <i>DLE</i> en línea. Aparecen el lema y todos los elementos de la microestructura trabajados y requeridos, aunque la disposición no es muy adecuada.	El formato de la entrada se ajusta en cierta medida al del <i>DLE</i> en línea. Aparecen el lema y la mayor parte de los elementos de la microestructura trabajados y requeridos (+75%) pero su disposición es mejorable.	El formato de la entrada no se ajusta de manera adecuada al del <i>DLE</i> en línea. Aparecen menos del 75% de los elementos de la microestructura trabajados y requeridos su disposición es mejorable.
Etimología	La etimología propuesta es muy adecuada en contenido y forma. Su contenido se ajusta al propuesto en las fuentes de consulta recomendadas y su presentación reproduce al completo la del <i>DLE</i> (cursiva, abreviaturas)	La etimología propuesta es bastante adecuada en contenido y forma. Su contenido se ajusta al propuesto en las fuentes de consulta recomendadas y su presentación reproduce en cierta medida la del <i>DLE</i> (cursiva, abreviaturas)	La etimología propuesta es adecuada en contenido o forma. Su contenido se ajusta al propuesto en las fuentes de consulta recomendadas o su presentación reproduce en cierta medida la del <i>DLE</i> (cursiva, abreviaturas)	La etimología propuesta no es adecuada en contenido y forma. Su contenido no se ajusta al propuesto en las fuentes de consulta recomendadas y su presentación no reproduce la del <i>DLE</i> (cursiva, abreviaturas)
Categoría y subcategoría gramatical	La categoría y subcategoría gramatical propuestas para la acepción (o acepciones) son correctas y se expresan a través de las abreviaturas que emplea el <i>DLE</i>	La categoría gramatical propuesta para la acepción (o acepciones) es correcta, aunque hay algunos errores en la subcategoría o en el uso de las abreviaturas.	La categoría gramatical propuesta para la acepción (o acepciones) es correcta, aunque hay algunos errores en la subcategoría y en el uso de las abreviaturas.	La categoría y subcategoría gramatical propuesta para la acepción (o acepciones) es incorrecta y no se expresa a través de las abreviaturas que emplea el <i>DLE</i>
Marcas de uso	Incluye todas las marcas de uso requeridas para la palabra y las expresa mediante las abreviaturas del <i>DLE</i>	Incluye al menos el 75% de las marcas de uso requeridas para la palabra y la gran mayoría las expresa mediante las abreviaturas del <i>DLE</i>	Incluye entre el 50% y 75% de las marcas de uso requeridas para la palabra y algunas las expresa mediante las abreviaturas del <i>DLE</i>	Incluye menos del 50% de las marcas de uso requeridas para la palabra y no las expresa mediante las abreviaturas del <i>DLE</i>

Definición	Recoge todas las acepciones claramente diferenciables de la palabra según los quince primeros ejemplos y, además, están formuladas con corrección.	Recoge todas las acepciones claramente diferenciables de la palabra según los quince primeros ejemplos, pero la expresión no es muy cuidada.	Recoge al menos la mitad de las acepciones claramente diferenciables de la palabra según los quince primeros ejemplos y la expresión es bastante mejorable	Recoge menos de la mitad de las acepciones claramente diferenciables de la palabra según los quince primeros ejemplos y la expresión es inadecuada
Ejemplo	Proporciona un ejemplo del <i>CORPES</i> muy ilustrativo para la acepción (o total de las acepciones) incluyendo la palabra más relacionada estadística y lingüísticamente con el lema.	Proporciona un ejemplo del <i>CORPES</i> para la acepción (o total de las acepciones) incluyendo la palabra más relacionada estadística y lingüísticamente con el lema, pero no es muy ilustrativo	Proporciona un ejemplo del <i>CORPES</i> para la acepción (o total de las acepciones) incluyendo la palabra más relacionada estadísticamente con el lema, pero no es muy ilustrativo	Proporciona un ejemplo del <i>CORPES</i> para la acepción (o total de las acepciones), pero no es muy ilustrativo ni incluye la palabra más relacionada estadística o lingüísticamente con el lema.

## 8.5. Anexo V. Resultados de la aplicación parcial de la propuesta

### A) Ejemplo de realización de la hoja de preguntas de las consultas en CORPES

#### PREGUNTAS DE COMPROBACIÓN

1. ¿Cuántos casos hay en el CORPES del paradigma verbal de *barajar*? ¿Cuántas veces aparece el infinitivo *barajar*?
  - Hay 1994 paradigmas verbales del verbo *barajar*
  - Hay 256 veces en las que aparece el infinitivo *barajar*
2. ¿Cuántos casos hay del sustantivo *vino*? ¿Y de la forma *vino* del paradigma verbal de *venir*?
  - Hay 22184 en plural.
  - Hay 28846 casos del sustantivo *vino*
  - Hay 17230 casos del paradigma verbal de *venir*
3. ¿Cuántos casos hay de la palabra *Estado* ('institución') desde 2012 hasta 2021 (usar la opción Subcorpus para la fecha)?
  - Hay 37.096 casos de la palabra *Estado* (institución) en esas fechas
4. ¿Cuántos casos hay de la forma *saco* (sustantivo) en México?
  - Hay 797 de la forma *saco* (sustantivo) en México
5. ¿Cuándo aparece por primera vez integrada en español la palabra *creepy* en el sentido de 'malrollero' o 'que causa extrañeza, miedo, etc.'?
  - En el 2020.

## PREGUNTAS DE AMPLIACIÓN

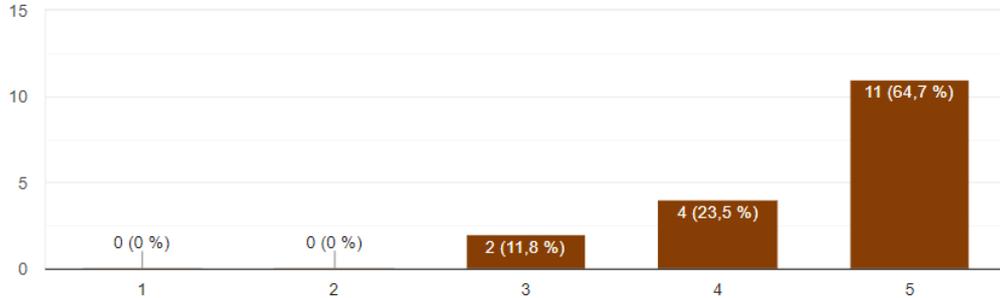
6. ¿Cuántos casos hay de la forma verbal, no recomendada por la norma, *estuvistes*?  
¿En qué tipo de textos es más frecuente? ¿Por qué creéis que es así?
- Hay 4 casos
  - En la forma oral.
  - Porque es incorrecta esa forma, y es poco recomendable
7. Buscad todos los casos de la palabra *frentazo* en noticias de prensa, usando la opción de Subcorpus, concretamente, la opción de soporte y tipología ¿Qué significado tiene esa palabra en ese contexto? ¿Está recogido en el *DLE*?
- Hay 8 casos
  - Tiene el significado de chutar un balón y lo marcar un gol (meter un gol pero con la frente).
  - No
8. El diccionario de la RAE recoge las variantes *período* y *periodo*. Usando la opción de "Grafía original" y "Estadística", ¿podrías decir cuál es la más usada por los hablantes en general? Desde 2016 hasta ahora, ¿cuál es la forma preferida o más usada por los hablantes?
- Es más usada la palabra *período* (27.850) y sin tilde (23.566).
  - Desde el 2016 hasta ahora, hay más palabras sin tilde de *período* (4.569) que en la palabra *periodo* <sup>con</sup> tilde (3.358).
9. En la canción "Sana Sana", Nathy Peluso dice: "tráeme la navaja filosa". Este adjetivo queda recogido en la RAE como "afilado, que tiene filo", entre otras acepciones. Revisando las estadísticas de esta palabra, ¿creéis que podrías añadirle a la definición alguna marca de la zona de uso de esta palabra?
- Se usa más en México y Centroamérica.

## B) Valoración de los estudiantes

Considero que el CORPES es una herramienta de consulta útil para mi aprendizaje de la lengua



17 respuestas

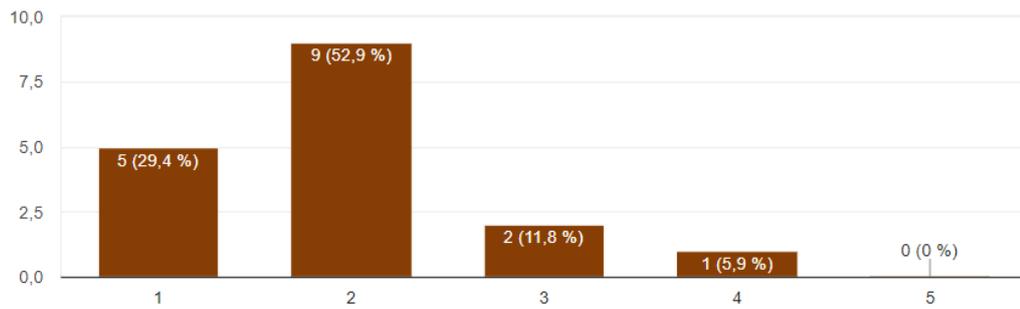


Escala 1= En total desacuerdo, 5=Totalmente de acuerdo

El uso del CORPES me ha resultado difícil



17 respuestas

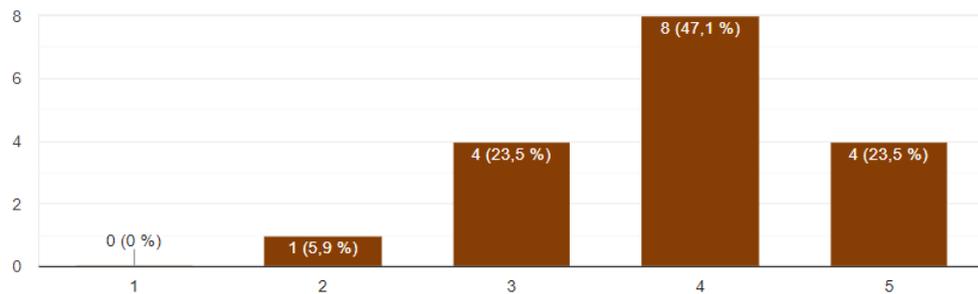


Escala 1= En total desacuerdo, 5=Totalmente de acuerdo

Volveré a utilizar el CORPES para resolver dudas acerca del uso del lenguaje



17 respuestas

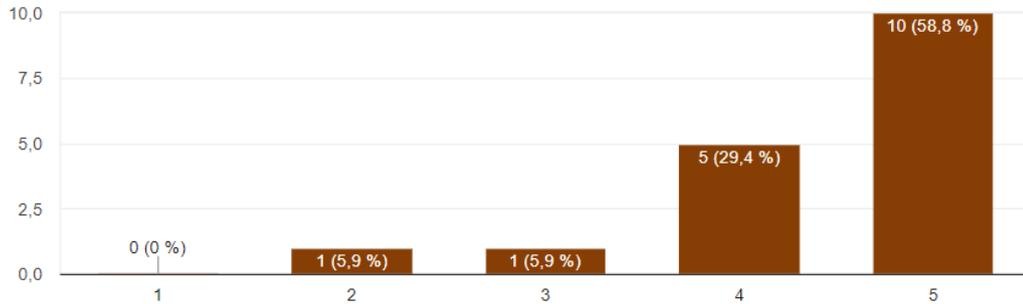


Escala 1= En total desacuerdo, 5=Totalmente de acuerdo

Las actividades de comprobación del aula de informática me han parecido adecuadas (dificultad, pertinencia) para para practicar las búsquedas simples en el CORPES



17 respuestas

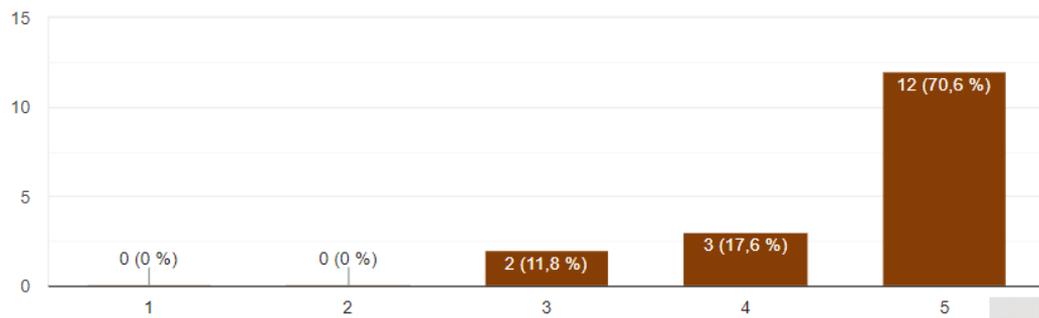


Escala 1= En total desacuerdo, 5=Totalmente de acuerdo

Las actividades de ampliación del aula de informática me han parecido adecuadas (dificultad, pertinencia) para reflexionar y constatar el contraste entre la información proporcionada por el DLE y el CORPES.



17 respuestas



Escala 1= En total desacuerdo, 5=Totalmente de acuerdo