
Morfología y Léxico en Ele: Análisis de Datos y Propuesta Didáctica Basada en Humanidades Digitales

Morfologia e Léxico em Ele: Análise de Dados e Proposta Didática Baseada em Humanidades Digitais

Morphology and Lexis in Sfl: Data Analysis and a Didactic Proposal Based on Digital Humanities

José Ramón Cruz Escudero¹

Resumen

En la enseñanza de español como lengua extranjera (ELE), el componente léxico se aborda habitualmente desde una perspectiva sincrónica, enfoque que suele postergar el análisis de la génesis y la productividad morfológica. Este trabajo presenta una propuesta didáctica basada en el uso de herramientas de las Humanidades Digitales y de corpus lingüísticos para el tratamiento de la competencia léxica en niveles intermedios y avanzados a través del estudio de los sufijos. La investigación se centra específicamente en los morfemas *-ción*, *-dad*, *-oso/-osa* y *-ear*, seleccionados por su rentabilidad y dinamismo en el español contemporáneo. Desde el punto de vista metodológico, se plantea un modelo de aprendizaje basado en datos que integra repositorios de referencia como el CORPES XXI y el CODEA, junto con el Corpus de Aprendices de Español (CAES). Asimismo, se incorpora el uso de herramientas de análisis y visualización textual, como AntConc y Voyant Tools, con el fin de explorar la frecuencia, distribución y productividad de estos derivados. La propuesta ofrece un marco de actuación didáctica fundamentado que aspira a favorecer la autonomía del estudiante, la identificación de relaciones morfológicas y semánticas y la reflexión crítica sobre la formación de palabras en el aula de ELE.

Palabras clave: ELE; morfología léxica; Humanidades Digitales; corpus lingüísticos.

Resumo

No ensino de espanhol como língua estrangeira (ELE), o componente lexical é habitualmente abordado a partir de uma perspectiva sincrónica, abordagem que tende a postergar a análise da génesis e da produtividade morfológica. Este trabalho apresenta uma proposta didática baseada no uso de ferramentas das Humanidades Digitais e de *corpora* lingüísticos para o tratamento da competência lexical nos níveis intermédios e avançados, por meio do estudo dos sufixos. A investigação centra-se especificamente nos morfemas *-ción*, *-dad*, *-oso/-osa* e *-ear*, seleccionados pela sua rentabilidade e dinamismo no espanhol contemporâneo. Do ponto de vista metodológico, propõe-se um modelo de aprendizagem baseada em dados que integra repositórios de referência como o CORPES XXI e o CODEA, juntamente com o Corpus de Aprendiz de Espanhol (CAES). Paralelamente, incorpora-se o uso de ferramentas de análise e visualização textual, como AntConc e Voyant Tools, com o objetivo de explorar a frequência, a distribuição e a produtividade desses

¹ Universidad Rey Juan Carlos, Espanha, joseamon.cruz@urjc.es, <https://orcid.org/0000-0002-5658-0441>

derivados. A proposta oferece um enquadramento didático fundamentado, que procura favorecer a autonomia do estudante, a identificação de relações morfológicas e semânticas e a reflexão crítica sobre os processos de formação de palavras no contexto do ensino de ELE.

Palavras-chave: ELE; morfologia léxica; Humanidades Digitais; *corpora* lingüísticos.

Abstract

In the teaching of Spanish as a foreign language (SFL), the lexical component is typically approached from a synchronic perspective, an approach that often postpones the analysis of morphological genesis and productivity. This article presents a didactic proposal based on the use of Digital Humanities tools and linguistic corpora for addressing lexical competence at intermediate and advanced levels through the study of suffixes. The study focuses specifically on the morphemes *-ción*, *-dad*, *-oso/-osa*, and *-ear*, selected for their productivity and relevance in contemporary Spanish. From a methodological perspective, the paper proposes a data-driven learning model that integrates reference corpora such as CORPES XXI and CODEA, together with the Corpus of Spanish Learners (CAES). In addition, text analysis and visualization tools such as AntConc and Voyant Tools are incorporated to explore the frequency, distribution, and productivity of these derived forms. The proposal provides a theoretically grounded didactic framework that seeks to encourage learner autonomy, the identification of morphological and semantic relationships, and critical reflection on word-formation processes in the SFL classroom.

Keywords: SFL; lexical morphology; Digital Humanities; linguistic corpora.

1. Introducción

En el marco actual de la enseñanza del español como lengua extranjera (ELE, en adelante), el aprendizaje del léxico ha pivotado hacia un enfoque comunicativo y sincrónico. Sin embargo, esta tendencia suele omitir la riqueza que subyace en el campo de la formación de palabras y su prolífica evolución. El presente estudio parte del supuesto de que, en los niveles intermedios y superiores de competencia (B2–C2), definidos por el Marco Común Europeo de Referencia para las Lenguas (Consejo de Europa, 2002), el alumnado cuenta con una madurez cognitiva idónea para abordar la lengua desde una perspectiva analítica. Por consiguiente, el foco de este trabajo se desplaza hacia la conciencia morfológica, entendida como la capacidad de identificar y manipular la estructura interna de las palabras, a la vez que potencia la autonomía léxica de los aprendientes de ELE.

A menudo, cuando trabajamos esta metodología en clase, la aplicación incorrecta de las reglas de formación de palabras (RFP, en adelante) da lugar a formaciones ilegítimas que, si bien demuestran una competencia morfológica por parte de los aprendientes, revelan, a su vez, una falta de consolidación en el uso de los sufijos. Un ejemplo claro se observa en la nominalización, donde el estudiante puede generar términos como *decidición* (en lugar de *decisión*) por una sobregeneralización del sufijo *-ción*, o *pobredad* empleada en lugar de la forma irregular *pobreza*, al aplicar la regla general de adjetivo + *-dad* (*pobre* + *-dad*). Del mismo modo, en el ámbito verbal y adjetival, encontramos creaciones como el verbo

cantarear (en vez de *canturrear*) bajo la influencia del productivo sufijo *-ear*, o el adjetivo *atunoso*, donde se aplica *-oso* a una base nominal de forma inusual.

Bajo esta hipótesis, la presente investigación tiene como propósito fundamental analizar cómo la integración de las Humanidades Digitales (HD, en adelante), y en particular, el uso de corpus lingüísticos transforma el aprendizaje de la morfología derivativa en el campo de la enseñanza de ELE. En este punto, se propone un cambio de paradigma: migrar de un modelo de clasificación rutinario y tradicional hacia un enfoque basado en el descubrimiento activo y el aprendizaje empírico por parte del discente. Al enfrentarse a repositorios en formato electrónico, como el Corpus del Español del Siglo XXI (CORPES XXI, en adelante), el Corpus de Documentos Españoles Anteriores a 1800 (CODEA, en adelante) o el Corpus de Aprendientes de Español como Lengua Segunda y Extranjera (CAES, en adelante), el estudiante no solo accede a una base de datos real, sino que puede rastrear la productividad de los sufijos mencionados anteriormente y comprender su comportamiento diacrónico y funcional. Para materializar este recorrido hacia un aprendizaje por descubrimiento, resulta imperativo el despliegue de una infraestructura metodológica que trascienda el libro de texto. En efecto, en muchos manuales de ELE el tratamiento del léxico derivado suele limitarse a actividades cerradas, tales como la identificación de sufijos en listas predeterminadas, la asociación de palabras con definiciones o el relleno de huecos, sin recurrir a datos reales de uso ni a la exploración autónoma de corpus electrónicos.

En este sentido, es necesario la implementación de secuencias didácticas que apoyen el uso estratégico de grandes repositorios de referencia y herramientas de procesamiento del lenguaje natural. De este modo, el manejo de los corpus lingüísticos digitales — mencionados anteriormente — deja de ser una labor técnica de investigación para convertirse en un entorno de consulta dinámico que permita al estudiante observar, de manera empírica, el comportamiento real y la productividad de los sufijos en el español actual. En definitiva, se propone una metodología donde la tecnología no es un fin en sí mismo, sino el vehículo para que los discentes de niveles intermedios y avanzados desentrañen la edificación del español y perfeccionen su precisión léxica.

A modo de reflexión final para esta apertura, conviene recordar que el estudio del lenguaje ha estado históricamente condicionado por el acceso limitado a los datos empíricos. La aparición de los corpus digitales ha supuesto un cambio sustancial en este panorama, al permitir el análisis sistemático de grandes volúmenes de producción lingüística. En este marco, la morfología deja de concebirse únicamente como un conjunto de principios descriptivos y se configura como una herramienta analítica fundamental para identificar regularidades formales, detectar patrones de uso y observar, a partir de datos reales, los procesos de cambio léxico en curso.

2. Bases Teóricas

2.1. Morfología Derivativa y Herramientas de Corpus

El español, por su rica estructura morfológica, ofrece ejemplos como *cabecear* (de *cabeza* + sufijo *-ear*), que ilustran cómo los sufijos aportan significado y permiten ampliar el léxico. Esta perspectiva fomenta una comprensión del carácter dinámico y evolutivo de la lengua. En este sentido, la formación de nuevas unidades léxicas por derivación se sustenta en la adquisición de las RFP, y no en una simple analogía. En esta línea, Baralo (2007) afirma que: “con el conocimiento léxico el aprendiente almacena la palabra en el lexicón con tal cantidad de etiquetas categoriales y conceptuales, que puede actualizarlas con la función sintáctica y semántica que le corresponda dentro del enunciado.” (p. 396).

En la enseñanza de ELE, la formación de palabras se presenta con frecuencia sin una reflexión didáctica sistemática sobre su finalidad en el aula. Es habitual encontrar listados de prefijos y sufijos organizados de manera meramente descriptiva o como recopilaciones de excepciones, enfoques que han privilegiado tradicionalmente el carácter excesivamente sistemático o asistemático de los procedimientos de formación de palabras en español, sin atender a su potencial didáctico. La revisión bibliográfica realizada, junto con la experiencia docente, confirma que la morfología léxica no suele ocupar un lugar central en la enseñanza del léxico en ELE, hecho ya señalado por diversos autores (Baralo, 2007; Liceras, 2009; Serrano-Dolader, 2006, 2018, entre otros).

Esta ausencia se explica, en parte, por la falta de familiaridad del profesorado con los procesos de formación de palabras y su aplicación didáctica, lo que dificulta una explotación sistemática de este recurso en el aula. No obstante, distintos estudios subrayan que el trabajo con la morfología léxica y el desarrollo de la conciencia morfológica pueden resultar especialmente productivos en ELE. En palabras de Serrano-Dolader (2018), la reflexión sobre la estructura interna de las palabras complejas favorece la inferencia léxica, la consolidación del vocabulario y el uso adecuado de las categorías gramaticales en contextos sintácticos diversos. Estos planteamientos respaldan la necesidad de integrar la morfología derivativa como un eje relevante en la didáctica del léxico.

Partiendo de estos planteamientos teóricos sobre la conciencia morfológica y su valor en la didáctica del léxico, el estudio se apoya en un enfoque metodológico basado en el aprendizaje a partir de datos (ABD, en adelante), mediante el empleo de corpus lingüísticos y herramientas digitales orientadas a fomentar la reflexión metalingüística en el aula de ELE. El estudio de la formación de palabras en el aula de ELE ha experimentado un cambio de paradigma gracias a la integración de las HD. Por consiguiente, las herramientas digitales idiosincrásicas de las HD permiten superar la enseñanza estática de las RFPs, al ofrecer acceso directo a la frecuencia de uso y a la vitalidad de la creación de nuevas palabras en contextos comunicativos reales. Asimismo, el procesamiento de datos masivos facilita que

el discente desarrolle una capacidad analítica inductiva, transforme la teoría léxica en una experiencia de aprendizaje tangible y basada en la evidencia.

Tradicionalmente, la enseñanza de la derivación se limitaba a listas de afijos estáticos y ejemplos paradigmáticos tales como:

- *-ción*: acción y efecto (ej. *clasificación, formación*).
- *-dad*: cualidad o estado (ej. *lealtad, brevedad*).
- *-ear*: acción repetida o proceso (ej. *golpear, pasear*).

En consonancia con lo expuesto, el enfoque tradicional tendía a ignorar tanto la productividad real de los procesos morfológicos como la generación de neologismos. En cambio, las herramientas que se emplean en este trabajo —como el CORPES XXI y el CODEA— permiten identificarlos de manera sistemática, tal como ocurre con la creación de verbos de uso técnico o informal (p. ej., *googlear, postear*). En este sentido, el análisis de la sufijación no debe entenderse como un ejercicio memorístico, sino como un proceso de "andamiaje cognitivo" que permite al alumnado desglosar la complejidad del léxico español².

Así pues, desde una perspectiva lingüística, los sufijos seleccionados para este análisis representan áreas de especial dificultad y, a su vez, de gran potencial creativo para el aprendiente:

- La nominalización (*-ción* y *-dad*): los sufijos *-ción* y *-dad* son fundamentales para la creación de sustantivos abstractos, esenciales en el discurso académico y formal. Mientras que *-ción* suele ligarse a bases verbales, *-dad* actúa sobre bases adjetivales, generando a menudo dudas en la selección de la base correcta (como vimos en la formación ilegítima *pobredad*).
- La adjetivación y verbalización (*-oso/-osa* y *-ear*): el sufijo *-oso/-osa* destaca por su alta productividad histórica y actual, mientras que *-ear* representa uno de los mecanismos más dinámicos de la lengua para la creación de nuevos verbos a partir de sustantivos, lo que lo convierte en un terreno fértil para las innovaciones (o errores) de los estudiantes.

Por otro lado, la bibliografía científica reciente subraya que el uso de bases de datos (como el CORPES XXI o el CAES) dota al estudiante de un papel investigador con el léxico. Esta metodología permite que los aprendientes observen la frecuencia de uso de determinadas formas en contextos reales. En el ámbito de la lingüística de corpus, el análisis sistemático

² Como se ha señalado en trabajos previos, la morfología presenta un notable potencial generativo en las lenguas, si bien su productividad varía considerablemente entre sistemas lingüísticos. Para una revisión aplicada al ámbito del ELE, especialmente en relación con la sufijación de unidades adjetivales, véase Cruz Escudero (2023).

de grandes volúmenes de datos auténticos —mediante concordancias, distribuciones de frecuencia y patrones de coocurrencia— facilita la comprensión de los mecanismos de formación del léxico actual, al evidenciar qué procesos morfológicos se encuentran realmente activos, en qué contextos aparecen y con qué grado de productividad.

En suma, la enseñanza de la morfología léxica en el ámbito de ELE y las tecnologías propias de las HD han adquirido una relevancia creciente en los últimos años, no solo por su papel estructural en la organización del léxico, sino también por su notable potencial para fomentar la reflexión metalingüística, la autonomía del aprendiente y la motivación intrínseca. En este contexto, la incorporación de recursos digitales y corpus históricos aporta rigor y aplicabilidad pedagógica, sitúa la morfología en un marco histórico y tecnológico, al tiempo que introduce una dimensión historiográfica que enriquece la comprensión del léxico como sistema dinámico.

2.2. Los Repositorios Digitales como Fuente de Evidencia Lingüística

En el campo de la enseñanza de ELE, la integración de grandes corpus textuales resulta esencial para documentar la naturaleza dinámica del cambio léxico actual. Más allá de su concepción tradicional como meras colecciones de textos, los corpus contemporáneos se definen por su naturaleza digital y estructurada. Los bancos de datos, organizados bajo criterios de registro, geografía y cronología, incorporan algoritmos de búsqueda que permiten procesar volúmenes masivos de información en un instante. La complejidad y la capacidad técnica propias de la lingüística de corpus permiten superar las limitaciones derivadas de la frecuencia natural. De este modo, fenómenos aparentemente periféricos adquieren relevancia estadística y se cuenta con un respaldo empírico sólido para el análisis de la variación (Alvar Ezquerro et al., 1994).

En la didáctica de lenguas, la utilidad de dichos repositorios es bidimensional: no solo aseguran que el *input* pedagógico se sustente en muestras de habla real, sino que permiten diseccionar la interlengua del estudiante mediante herramientas específicas como el CAES. Así, frente a un corpus restringido, que apenas permite observar la sufijación creativa, un repositorio de gran escala como el CORPES XXI ofrece al docente la posibilidad de seleccionar ejemplos representativos que ilustren la vitalidad y la oscilación formal de morfemas como *-oso/-osa*, *-ear*, *-ción* o *-dad*³. De este modo, el alumnado de ELE no aborda el análisis en los corpus lingüísticos en sentido estricto, sino que se aproxima, de

³ Cf. el caso del verbo *agendar*, cuya nominalización ilustra la competencia entre los sufijos *-ción* y *-miento*. Mientras que en muestras limitadas el derivado puede resultar inexistente, el CORPES XXI permite documentar la consolidación de *agendamiento* —especialmente en las variedades americanas y en los registros jurídico y médico— frente al carácter residual de *agendación*. Esta capacidad de mapear la oscilación formal y el desplazamiento geolingüístico de un neologismo solo es factible mediante el análisis de la «larga cola» (*long tail*) de frecuencias que ofrecen los macrocorpus.

forma guiada, a patrones de uso reales que favorecen la comprensión de la productividad morfológica.

Nuestra propuesta metodológica integra una perspectiva sincrónica y otra diacrónica mediante la combinación del CORPES XXI, que refleja la diversidad y polifonía del español actual, y el CODEA, orientado al estudio histórico de la lengua. Si bien el CODEA constituye una herramienta especializada destinada principalmente a la investigación en historia del español, su incorporación en el ámbito de ELE no responde a un uso directo por parte del estudiante, sino a una función mediada por el docente. En este sentido, el corpus permite seleccionar y adaptar ejemplos históricos que ilustran la evolución de determinados formantes derivativos, con el fin de contextualizar procesos morfológicos todavía productivos en la lengua contemporánea. De este modo, el alumnado de ELE accede a una comprensión más profunda de la morfología actual, al conocer el origen y la trayectoria de ciertas formas sin necesidad de realizar un análisis paleográfico o diacrónico especializado (Moreno Fernández, 2009).

Este análisis relaciona la productividad histórica con el uso actual y contribuye al desarrollo de la competencia metalingüística del estudiante en el ámbito de la afijación. No obstante, ningún corpus resulta autosuficiente, ya que cada tipología presenta limitaciones: los macrocorpus priorizan el volumen de datos, a menudo en detrimento del control filológico y la precisión cronológica, mientras que los corpus de documentación controlada ofrecen mayor rigor, aunque con menor representatividad. Este equilibrio condiciona los ejemplos lingüísticos a los que se expone el aprendiente de ELE y afecta a su percepción de la productividad, frecuencia y vigencia léxicas, lo que hace necesaria una mediación docente en la selección de los corpus.

Conviene subrayar que los corpus de gran tamaño proporcionan información valiosa sobre palabras de baja frecuencia, pues en ellos se registran mayores ocurrencias de esta categoría. Ello resulta particularmente útil para estudios morfológicos. Así lo demuestra el subcorpus *Historical/Genres* del Corpus del Español:

los 20 millones de palabras del Corpus Historical/Genres, existen sólo alrededor de 45 palabras terminadas en *-azo* (sufijo que alude a un movimiento o golpe con un objeto) que ocurren 20 veces o más; y las más bajas frecuencias son *arañazo*, *terrazo*, *espaldarazo*, *golazo*, *pinchazo*, *pelotazo*, *escopetazo*, *campanillazo*, *gustazo*, *trancazo*. Tenemos una información mucho más rica con los dos mil millones de palabras del Corpus Web/Dialects donde hay más de 460 palabras que ocurren al menos 20 veces, incluyendo las de baja frecuencia: *morongazo*, *postazo*, *teclazo*, *pollazo*, *autogolazo*, *libretazo*, *arcabuzazo*, *platanazo*, *trolazo*, *inventazo*, *peloterazo*, *guadañazo*, *potazo*, *fregadazo*, *cuponazo*. Obviamente con más de 20 veces en muchos tipos (formas distintas) se puede investigar el fenómeno de la creatividad léxica de mucho mejor manera que con un corpus pequeño. (Davies & Parodi, 2022, p. 18)

Finalmente, el recurso al CAES se revela como un componente estratégico para el estudio de la interlengua. Como sostiene Granger (2012), este tipo de repositorios facilitan la identificación de tendencias de error y reproducciones estructurales; y a la vez, permiten el

diseño de intervenciones pedagógicas alineadas con la progresión real del alumnado en cada fase de su aprendizaje.

2.3. La Integración de los Corpus en los Marcos de Referencia (Mcer y PciC)

La utilidad de los corpus lingüísticos en la didáctica de ELE se sustenta en su adecuación a las directrices establecidas por el *Marco común europeo de referencia para las lenguas* (MCER; Consejo de Europa, 2002) y a los inventarios del *Plan curricular del Instituto Cervantes* (PCIC; Instituto Cervantes, 2006). Ambos documentos enfatizan la necesidad de desarrollar una competencia léxica basada en el uso real de la lengua, la autonomía del aprendiente y la capacidad de establecer regularidades formales, objetivos que encuentran un respaldo metodológico sólido en el empleo didáctico de corpus lingüísticos.

En el marco establecido por el Marco común europeo de referencia para las lenguas (MCER; Consejo de Europa, 2002), que concibe el aprendizaje de lenguas como un proceso orientado a la acción y basado en el uso, el empleo de repositorios digitales contribuye a garantizar que el *input* pedagógico responda al principio de autenticidad. En efecto, estos recursos proporcionan muestras de lengua procedentes de contextos reales de comunicación, que complementan los ejemplos habitualmente seleccionados y adaptados con fines didácticos en los materiales impresos. Esta metodología favorece el desarrollo de la competencia comunicativa del alumnado, ya que permite observar la combinatoria léxica y las colocaciones en contextos de uso efectivo, aspectos fundamentales para alcanzar la fluidez y la adecuación exigidas en los niveles superiores de competencia.

Por su parte, el PCIC proporciona el marco de sistematización necesario para la aplicación de estos recursos. Mientras que este documento desglosa los objetivos y contenidos por niveles de competencia, los macrocorpus como el CORPES XXI permiten validar la productividad real de los inventarios de gramática y léxico desde una perspectiva pluricéntrica. De este modo, el estudio de la morfología derivativa deja de ser una lista estática de RFPs para convertirse en un análisis dinámico de la frecuencia y vitalidad de los afijos en las distintas variedades del español. Asimismo, la incorporación de corpus de aprendices, como el CAES, resulta esencial para contrastar la norma descrita en el PCIC con la interlengua real del estudiante, lo que facilita una transposición didáctica más precisa y orientada a las necesidades empíricas del aula de ELE.

3. Metodología: Instrumentos de Visualización y Análisis Digital

En el ecosistema de las HD, la visualización de datos ha trascendido la mera función estética para consolidarse como un vehículo de transferencia de conocimiento. Su potencial reside en la capacidad de convertir grandes volúmenes de información en estructuras gráficas que exponen nexos y regularidades que de otro modo permanecerían invisibles. Desde esta vertiente, el análisis visual constituye un giro metodológico sustancial: el hecho lingüístico deja de ser una unidad aislada para integrarse en un sistema de datos cuya representación espacial revela tendencias evolutivas y semánticas. Por tanto, toda representación gráfica implica un proceso de selección e interpretación de los datos lingüísticos, en el que se priorizan determinadas variables con el fin de apoyar una hipótesis de análisis. En la esfera de la enseñanza de ELE, este tipo de recursos facilita que el discente visualice procesos como la conversión léxica o la productividad morfológica a partir de datos reales, lo que contribuye a transformar el estudio de la morfología en una actividad guiada, dinámica y orientada al desarrollo de la autonomía en el aprendizaje del vocabulario.

3.1. Antconc

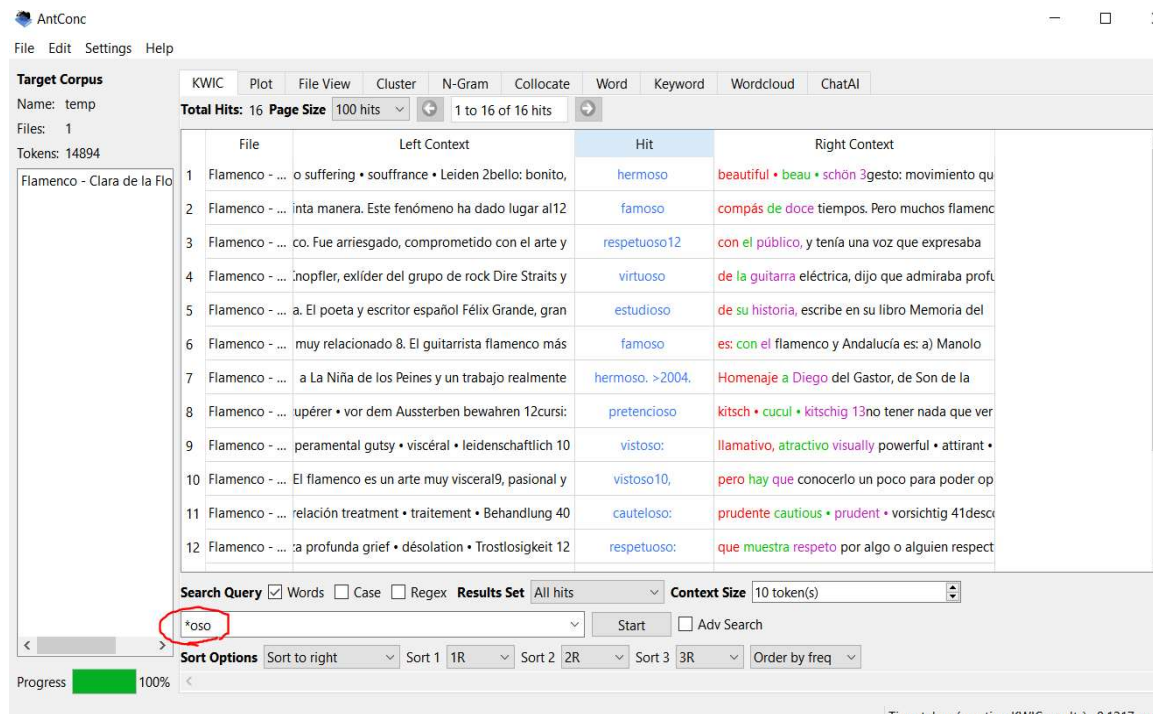
El *software* AntConc (Anthony, 2019), concebido por Laurence Anthony, se erige como una potente herramienta gratuita para el procesamiento de textos, que permite la extracción de concordancias, el cálculo de frecuencias y la detección de n-gramas⁴. Su operatividad es esencial para examinar la vitalidad de los morfemas tanto en estadios históricos como en el habla actual. Dentro del aula de ELE, este mecanismo favorece la monitorización de sufijos altamente productivos y el análisis de su comportamiento en contextos reales de comunicación. A modo de ejemplo, el docente puede rastrear la presencia de terminaciones como *-ción*, *-dad*, *-oso/-osa* o *-ear* en corpus específicos y observar cómo su frecuencia fluctúa según la tipología textual o el registro empleado. Dichos hallazgos sirven de base para articular tareas de descubrimiento guiado donde el alumnado deduce RFPs y reflexiona sobre el entramado de la lengua.

Como se observa en la **Figura 1**, el análisis de frecuencias y concordancias del sufijo *-oso/-osa* mediante el *software* AntConc permite identificar tanto su productividad como los contextos léxicos más frecuentes de aparición.

⁴ Entendidos como secuencias contiguas de *n* unidades léxicas que aparecen de manera recurrente en un corpus.

Figura 1

Análisis de frecuencias y concordancias del sufijo -oso/-osa mediante el software AntConc 4.3.1.



Fuente: elaboración propia a partir del análisis del texto fuente en AntConc.

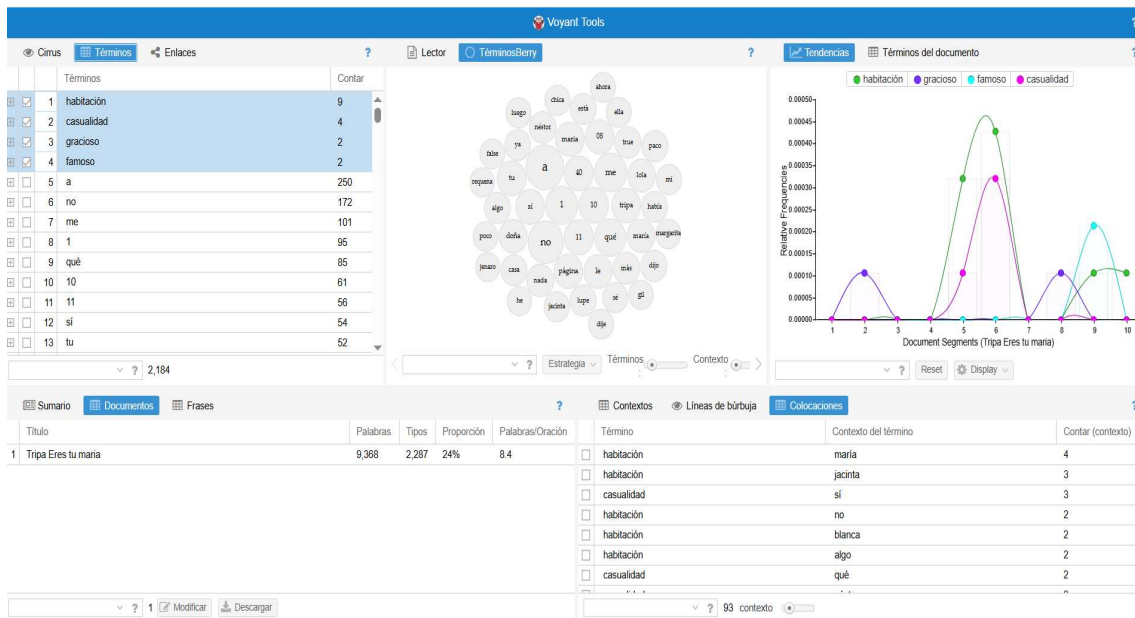
3.2. Voyant Tools

Voyant Tools representa una solución de código abierto orientada a la denominada lectura distante, entendida como un enfoque analítico que permite examinar grandes volúmenes de texto mediante métodos cuantitativos y visuales —como el análisis de frecuencias, tendencias léxicas o coocurrencias— sin centrarse en la lectura detallada de cada documento individual. Esta perspectiva facilita el análisis interactivo de textos y resulta especialmente útil para identificar patrones globales de uso léxico en corpus extensos. Su interfaz facilita el manejo de grandes masas documentales sin necesidad de recurrir a la programación compleja, además, ofrece herramientas como Cirrus (nubes de palabras de frecuencia relativa), gráficos de distribución temporal y esquemas de redes para las colocaciones. Estas funcionalidades son críticas para el estudio de la variación diatópica y diacrónica, hecho que permite comparar de forma intuitiva la evolución de los procesos derivados del léxico.

Como se aprecia en la **Figura 2**, el análisis de distribución y tendencias léxicas obtenido mediante Voyant Tools permite observar la frecuencia relativa de determinados ítems y su evolución a lo largo del texto, lo que resulta útil para el diseño de actividades de lectura guiada en ELE.

Figura 2

Análisis de distribución y tendencias mediante Voyant Tools aplicado a la lectura graduada: ¿Eres tú, María?



Fuente: elaboración propia

La integración de herramientas como AntConc y Voyant Tools en la didáctica de ELE favorece una metodología activa en la que el alumnado participa de forma directa en la exploración y análisis de los procesos morfológicos, apoyándose en datos reales de uso lingüístico. Al explorar la frecuencia y el uso de sufijos específicos de manera empírica, el alumnado no solo mejora su precisión léxica, sino que comprende la lengua como un organismo vivo y mutable. En última instancia, la digitalización de estos procesos analíticos dota al estudiante de herramientas estratégicas que refuerzan tanto su competencia lingüística como su madurez crítica frente al texto.

4. Análisis y Trayectoria de los Morfemas Derivativos Seleccionados

El siguiente apartado detalla los sufijos seleccionados para su aplicación en la propuesta didáctica objeto de este estudio; dicha selección responde a su elevada rentabilidad y a su valor didáctico en el aula de ELE⁵. Nuestra elección responde a un criterio de simplificación metodológica que prioriza el estudio de la forma prototípica de cada sufijo. Así, en el caso de *-ción*, *-dad*, *-oso/-osa* y *-ear*, se analizan exclusivamente sus realizaciones más frecuentes, prescindiendo de variantes alomórficas como *-sión* y *-xión*, *-tad*, *-ado/-ada*

⁵ Para un análisis más exhaustivo sobre la rentabilidad y el valor didáctico de estos sufijos en el aula de ELE, véase Cruz Escudero (2023).

condicionadas por alternancias fonológicas, o los ajustes gráficos asociados a *-ear* (p. ej., *postear-posteado*; *bloquear-bloqueado*). Esta delimitación permite una mayor nitidez en el análisis de la trayectoria semántica, el comportamiento pragmático y la recurrencia diacrónica de las voces derivadas, sin introducir complejidades formales ajenas a los objetivos didácticos del aula de ELE.

No obstante, es preciso reconocer que los materiales integrados en los corpus lingüísticos no constituyen una muestra absoluta de la lengua, sino que representan aquellos testimonios que han sobrevivido a las contingencias históricas y editoriales. En el caso del español, existen hiatos temporales —como el siglo XIV o el paso entre el XVII y el XVIII— con una representatividad menor, a lo que se suma el sesgo propio de la escritura, que suele filtrar los rasgos de la oralidad y los usos de grupos sociales tradicionalmente marginados hasta el siglo XIX. Pese a estas limitaciones, los corpus diacrónicos se erigen como el instrumento más eficaz para reconstruir estadios de lengua pretéritos. La frecuencia de uso permite validar la vigencia de una forma dentro del sistema lingüístico, siempre que el análisis se sustente en una muestra representativa que equilibre diversos registros y evite las distorsiones inherentes de estilos individuales.

Desde esta perspectiva, la propuesta desarrolla una explotación didáctica y sistemática de los sufijos seleccionados a partir de corpus con alcances temporales diferenciados: el CORPES XXI, centrado en textos contemporáneos, y el CODEA, que reúne documentación histórica entre los siglos XIII y XIX. Los datos extraídos se analizan mediante un enfoque cualitativo basado en la ejemplificación textual y en visualizaciones gráficas. La consulta del CODEA no tiene un objetivo formativo diacrónico para el aprendiente de ELE, sino una función contextualizadora mediada por el docente, orientada a comprender la productividad actual de los sufijos. Además, el acceso abierto y la trazabilidad de las fuentes garantizan la replicabilidad y el rigor del análisis.

4.1. El Sufijo *-ción*⁶

Derivado del latín *-tionem*, el sufijo *-ción* se consolida como uno de los mecanismos más potentes de la lengua española para la creación de sustantivos de acción y efecto a partir de bases verbales. Su rastreo en la documentación histórica —especialmente en tratados jurídicos y eclesiásticos de la época preclásica— evidencia una presencia predominante en discursos de alta formalidad. Términos como *abjuración* o *contravención* son ejemplos de una especialización léxica temprana con acepciones muy acotadas. No obstante, con el advenimiento de la modernidad y el desarrollo de las estructuras sociales a partir del siglo XIX, este sufijo cultiva el lenguaje político y administrativo, lo que consolida unidades como *organización* o *revolución*, que trascienden el significado técnico para adquirir densas connotaciones ideológicas. En el ámbito didáctico, esta trayectoria permite que el alumnado

⁶ Real Academia Española y Asociación de Academias de la Lengua Española. (2014). En *Diccionario de la lengua española* (23.^a ed.). Recuperado el 22 de diciembre de 2025, de <https://dle.rae.es/-ción>

no solo se ejercite en la transposición de categorías (verbo a sustantivo), sino que también se aproxime al contexto sociocultural y a la carga pragmática que estas voces han acumulado a lo largo de los siglos.

Como se observa en la Figura 3, la consulta del sufijo *-ción* en el CORPES XXI permite identificar su alta frecuencia de aparición en textos contemporáneos y su asociación con sustantivos abstractos derivados de bases verbales, lo que confirma su elevada productividad en el español actual. Los datos reflejan que *-ción* se mantiene como un mecanismo derivativo plenamente activo en distintos géneros discursivos, especialmente en registros expositivos y argumentativos, lo que justifica su selección como objeto de análisis en una propuesta didáctica orientada a niveles intermedios y avanzados de ELE.

Figura 3
Muestras de la consulta del sufijo *-ción* en el CORPES XXI.

Inventarios		
Descargar en TSV		
Número de resultados: 14 944 / Documentos: 388 357 / Frecuencia absoluta: 7 451 510 / Frecuencia normalizada: 17 057,21 por millón		
Formas	F.Abs.	F.Norm.
información	142 108	325,29
situación	127 632	292,16
población	99 915	228,71
atención	95 855	219,42
investigación	93 114	213,14
relación	93 008	212,9
producción	89 612	205,13
condiciones	88 884	192,01
educación	72 255	165,39
participación	61 014	139,66
acción	60 196	137,79
comunicación	58 556	134,04
acciones	56 374	129,04
relaciones	55 992	128,17
formación	55 362	126,72
construcción	54 655	125,11
organización	53 612	122,72
instituciones	51 108	116,99
función	50 751	116,17
posición	48 470	110,95
creación	46 752	107,01
dirección	45 327	103,75
elecciones	42 658	97,64
generación	37 331	85,45
protección	37 159	85,06

Fuente: elaboración propia a partir de datos del corpus CORPES XXI. Recuperado de <https://www.rae.es/corpes/>

4.2. El Sufijo -Dad⁷

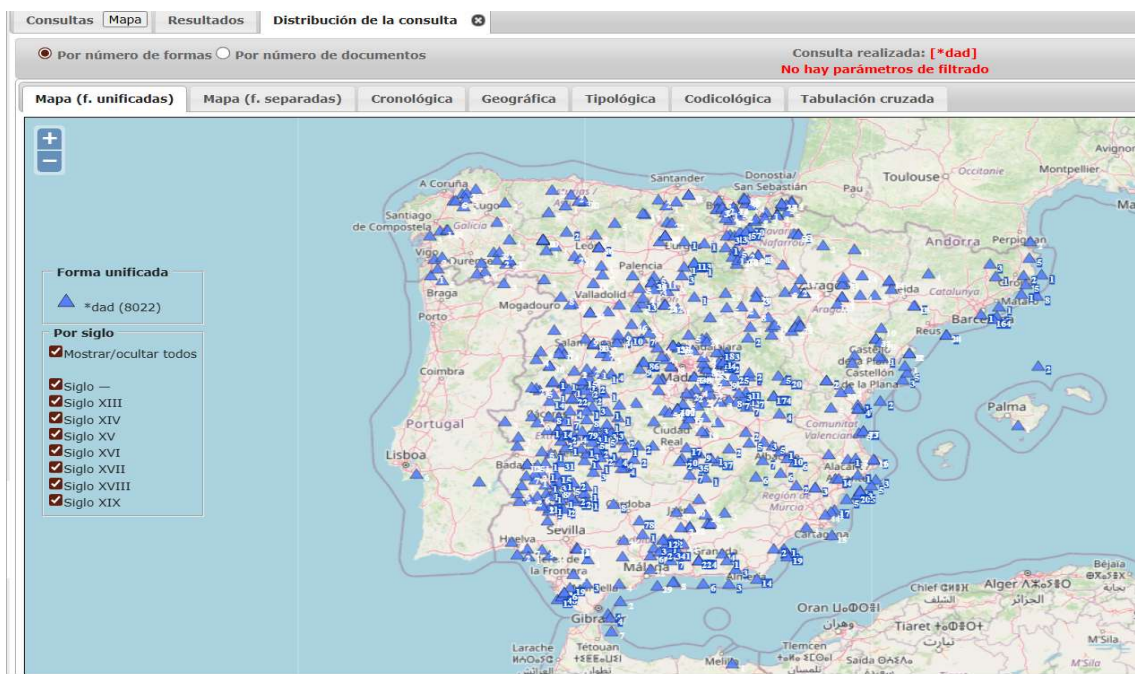
El morfema *-dad* (del latín *-tas*, *-tatis*) constituye el mecanismo paradigmático para la creación de sustantivos de cualidad a partir de bases adjetivales. Su productividad y estabilidad semántica lo convierten en un recurso especialmente representativo dentro del sistema morfológico del español. En el ámbito del currículo de ELE, el estudio de este sufijo resulta pertinente tanto desde una perspectiva descriptiva como didáctica, ya que permite abordar procesos de nominalización, profundizar en la relación entre forma y significado y favorecer el desarrollo de la competencia léxica y metalingüística del alumnado. Históricamente, el análisis en el CODEA posibilita observar su fijación en términos vinculados a la moralidad y la organización civil desde la etapa medieval, con voces como *lealtad* o *humanidad*.

No obstante, la complejidad de este sufijo para el aprendiente radica en su variación alomórfica (*-idad*, *-edad*, *-tad*), lo que genera dudas sobre la selección de la base, como se evidencia en el uso de formas no normativas detectadas en el CAES. La exploración del CORPES XXI revela que, en el español contemporáneo, este sufijo posee una vitalidad extraordinaria en el discurso académico y científico, facilitando la cohesión textual mediante la nominalización de adjetivos técnicos. Desde una vertiente pedagógica, trabajar con *-dad* a través de herramientas de visualización ayuda a que el alumno identifique las restricciones de combinación y comprenda que la elección de este sufijo frente a otros competidores (como *-eza* o *-ura*) responde a menudo a una cuestión de registro y tradición discursiva.

Como puede observarse en la Figura 4, las consultas realizadas en el CODEA permiten documentar la presencia y distribución de derivados en *-dad* en textos históricos, lo que facilita el análisis de su evolución formal y semántica en distintas etapas del español.

⁷ Real Academia Española y Asociación de Academias de la Lengua Española. (2014). En Diccionario de la lengua española (23.^a ed.). Recuperado el 24 de diciembre de 2025, de <https://dle.rae.es/-dad>

Figura 4
Muestras de la consulta del sufijo -dad en el CODEA.



Fuente: elaboración propia a partir de datos del corpus CODEA (2022). Recuperado de <https://doi.org/10.37536/CODEA.2015>. Imagen bajo licencia Creative Commons CC BY 4.0.

4.3. El Sufijo -Oso/-Osa⁸

El sufijo *-oso/-osa* (del latín *-osus*) destaca como uno de los morfemas adjetivadores más dinámicos y longevos del sistema lingüístico español. Su función primordial es la creación de adjetivos a partir de sustantivos que denoten abundancia, presencia o cualidad. En el nivel avanzado de ELE, el interés de este sufijo trasciende la mera acumulación léxica; se convierte en una herramienta para que el alumnado aprecie la gradación semántica y la carga evaluativa que el hablante imprime a su discurso. A través del corpus CODEA, es posible observar que dicha unidad ha mantenido una estabilidad notable desde el castellano medieval, con voces que hoy consideramos fundamentales como *hermoso* o *virtuoso*. No obstante, la consulta al CORPES XXI revela su faceta más innovadora y su plasticidad para adaptarse a nuevos dominios, desde el lenguaje coloquial hasta el científico.

Uno de los aspectos didácticos relevantes de esta propuesta es el análisis de la productividad morfológica de los sufijos derivativos. El uso del CAES en el aula permite al alumnado observar que el sistema léxico ofrece un margen para la creación contextual, más

⁸ Real Academia Española y Asociación de Academias de la Lengua Española. (2014). En *Diccionario de la lengua española* (23.ª ed.). Recuperado el 26 de diciembre de 2025, de <https://dle.rae.es/-oso>

allá de las unidades recogidas en los diccionarios. La aplicación del sufijo *-oso/-osa* a bases no convencionales, como *atunoso*, ilustra la generación de efectos descriptivos o expresivos mediante mecanismos formales reconocibles. De este modo, su estudio trasciende la norma y favorece la exploración de la creatividad lingüística en contextos de uso auténtico⁹.

Como se muestra en la Figura 5, los aprendientes de nivel C1 utilizan el CAES para contrastar hipótesis sobre su productividad y valor semántico.

Figura 5

La captura muestra la consulta del sufijo -oso tal como empleado por aprendientes de nivel C1.

Nº	Código	Fragmento de texto	Palabra resaltada
1	C1/Árabe	ha gustado en la peli porque el chico habla de cosas de las que los adultos pasan completamente como el acoso escolar , las penas de amor ... Y es+que es+que de verdad ¿ quién es este padre que se empeña	acoso
2	C1/Francés	anos sesenta , setenta tenía mucho mucho carisma en la película representaba el hombre cinico aventurero ambicioso	ambicioso
3	C1/Polaco	Llega hasta un punto en el que el cumbre parece estar ambicioso pero la manera de resolver lo no me gusta nada .	ambicioso
4	C1/Inglés	La original se trata de un ejecutivo que contrata a un ambicioso joven para trabajar en su empresa de inversiones .	ambicioso
5	C1/Portugués	se pasa en un barrio de Salvador en Bahía que nombra la película y cuenta la estoria de un triángulo amoroso entre tres personajes , una prostituta y dos amigos , populares de Salvador que viven de hacer	amoroso
6	C1/Portugués	En este sitio empieza un triángulo amoroso (casi cuadrado) muy sencillo , que envuelve Melanie , Wonderer , Jared y Ian .	amoroso
7	C1/Portugués	Es un buen programa para los niños y los adolescentes que les gustan muchísimo a_causa_de el triángulo amoroso entre una humana , un vampiro y un hombre+lobo hombre+lobo .	amoroso
8	C1/Portugués	empezar una amistad con beneficios , sexo sin compromiso , los dos no hacen lazos de relacionamiento amoroso .	amoroso
9	C1/Polaco	verdadero placer y diversión para todos los aficionados porque muestra el divertido , aunque seguramente ardoso , proceso de componer esa obra .	ardoso
10	C1/Portugués	A el saltar en una piscina natural local se chocó contra el suelo firme y arenoso , y se dañó en cuello , más específicamente tuvo vértebras cervicales rotas que le comprimiran la	arenoso
11	C1/Japonés	Lo que me encanta esta película es la música " Cu-cu-lu-cu _la_paloma " cantado por Caetano_Veroso , que me animó aprender la guitarra para cantar la algún día .	Caetano_Veroso
12	C1/Inglés	sus jugadoras pero tiene muchos problemas personales y le gusta tomar demasiado , algo que le causa ser caprichoso , severo o sensible , depende de el día .	caprichoso
13	C1/Portugués	es una mujer muy insegura , mientras Paul parece ser una persona tímida y reservada , pero muy gentil y cariñoso .	cariñoso
14	C1/Árabe	... la película cuenta la estoria de un matrimonio que estaba muy felices y enamorados , el marido era muy celoso que en un momento de enfado la pegó y la hecho de casa y se han divorciado a el pasar	celoso
15	C1/Polaco	En una fiesta , un chico intenta seducir a Tessa y a besar la , pero Hardin se vuelve celoso y le golpea .	celoso
16	C1/Inglés	Puede que este resumen de el argumento parezca poco cómico y hasta oscuro , pero lo chistoso es la manera en la que suceden los eventos .	chistoso
17	C1/Ruso	Fue curioso ver en la misma película por _ejemplo Judie_Dench , Jonnie_Depp y Michelle_Pfaifer .	curioso
18	C1/Polaco	He decidido ver la por su título , que me parecía bastante curioso .	curioso
19	C1/Polaco	Además , es curioso llegar a saber que si no hubiera sido por la obstinación de Freddie , la canción podría no haber aki	curioso
20	C1/Polaco	indemnización tanto por haber me engañado proponiendo una oferta imposible de cumplir como por el trato escandaloso que recibí por parte de los que trabajan para su empresa .	escandaloso

Fuente: elaboración propia a partir de datos del corpus CAES (2022), Instituto Cervantes y Universidad de Santiago de Compostela. Recuperado de <https://galvan.usc.es/caes/>.

4.4. El Sufijo -Ear¹⁰

El sufijo *-ear* representa uno de los mecanismos de verbalización más prolíficos y flexibles del español contemporáneo. Su función primordial es la creación de verbos a partir de bases nominales (denominativos) o adjetivales, a menudo con un matiz de acción reiterada o

⁹ Para profundizar en este aspecto, pueden consultarse otros trabajos del autor (Cruz Escudero, 2022).

¹⁰ Real Academia Española y Asociación de Academias de la Lengua Española. (2014). En *Diccionario de la lengua española* (23.ª ed.). Recuperado el 27 de diciembre de 2025, de <https://dle.rae.es/-ear>

durativa. En el contexto de ELE, este sufijo es de vital importancia, ya que constituye la puerta de entrada a la neología y a la adaptación de préstamos lingüísticos, especialmente en el lenguaje técnico y de las redes sociales. A través del análisis diacrónico en el CODEA, se observa que *-ear* ha sido un recurso constante para la adaptación de vocabulario cotidiano desde los orígenes del idioma. Sin embargo, su verdadera explosión se documenta en el CORPES XXI, donde se evidencia su papel protagonista en la era digital. Verbos como *escanear*, *cliquear*, *tuitear* o *postear* demuestran que este morfema integra términos extranjeros bajo las RFPs del español.

Desde la perspectiva del ABD, el uso de Voyant Tools permite al alumnado observar la "nube de tendencias" de estos verbos, identificar cuáles están en auge y cuáles pertenecen a registros más informales o jergales. Asimismo, el estudio de errores en el CAES revela que los estudiantes a menudo recurren a *-ear* como una estrategia de compensación, hecho que los lleva a formas como el mencionado *cantarear* cuando desconocen el término preciso. Trabajar este sufijo en el aula no solo refuerza la gramática, sino que dota al estudiante de una competencia estratégica fundamental: la capacidad de innovar léxicamente de manera coherente con la estructura de la lengua meta.

Como puede observarse en la Figura 6, las consultas realizadas en el CORPES XXI permiten identificar muestras representativas de formas verbales derivadas con el sufijo *-ear*, lo que facilita el análisis de su frecuencia y de los contextos de uso en el español contemporáneo a partir de datos empíricos.

Figura 6
Muestras de la consulta del sufijo *-ear* en el CORPES XXI.

Concordancias						
<input type="button" value="Descargar en TSV"/> <input type="button" value="Descargar en documento"/> <input type="button" value="Descargar en texto"/> <input type="button" value="Imprimir"/>		Ordenar por: Año ascendente				
Sin criterio						
Documentos: 75 215 / Frecuencia absoluta: 179 701 / Frecuencia normalizada: 411,35 por millón						
<input type="checkbox"/>	Doc.	País	Año	Contexto izquierdo	Término	Contexto derecho
<input type="checkbox"/>	IE2001_0004	Esp.	2001	...a estar al amanecer en el puerto, y	pelear	se por trabajar en los muelles a cam...
<input type="checkbox"/>	IE2001_0005	Esp.	2001	...cratas que no se cortan un pelo en	pasear	se con las corbatas verdes y morada...
<input type="checkbox"/>	IE2001_0011	Esp.	2001	...n a estar doblemente motivados a	putear	a sus subordinados. ¿Dónde están l...
<input type="checkbox"/>	IE2001_0011	Esp.	2001	...mpo que tendrían mil maneras de	emplear	. ¿Tiene sentido? Ummmmm.
<input type="checkbox"/>	IE2001_0011	Esp.	2001	...y Easyjet- que han empezado ya a	putear	a sus trabajadores, haciendo que ca...
<input type="checkbox"/>	LA2001_0001	Bol.	2001	... desde la cintura a la derecha, por	banquetear	se con la sirvienta no bien terminó ...
<input type="checkbox"/>	LA2001_0001	Bol.	2001	...aco, de todo punto estúpida. Sabía	rodear	se de políticos y personajes de prest...
<input type="checkbox"/>	LA2001_0001	Bol.	2001	...egolatría y arrogancia les induce a	rodear	se de adulones viles en vez de aseso...
<input type="checkbox"/>	LA2001_0001	Bol.	2001	...e zapateaban donairosas y hacían	flamear	el pañuelo provocativamente cerca ...
<input type="checkbox"/>	LA2001_0001	Bol.	2001	...ean oficiales o diplomáticos, para	delinear	políticas nacionales o posiciones int...

Fuente: elaboración propia a partir de datos del corpus CORPES XXI. Recuperado de <https://www.rae.es/corpes/>

5. Propuesta de Innovación Docente: Una Aproximación Digital a la Morfología de Ele

La aplicación práctica de esta investigación se articula a través de una secuencia didáctica diseñada para los niveles comprendidos entre el B2 y el C2, definidos por el Marco Común Europeo de Referencia para las Lenguas (Consejo de Europa, 2002), cuyo eje vertebrador es la autonomía del estudiante en la exploración del léxico. A diferencia de enfoques didácticos basados en la presentación explícita de las RFPs, la memorización de listas de derivados o la realización de ejercicios cerrados de clasificación y cumplimiento de espacios en blanco, esta propuesta pedagógica aspira a que el alumnado asuma un rol activo como lingüista ingenuo, emplee el uso guiado de corpus lingüísticos para contrastar y validar sus propias hipótesis sobre la morfología derivativa.

5.1. Diseño De La Secuencia: Del Dato A La Competencia¹¹

La metodología se divide en tres fases estratégicas que integran las herramientas analizadas previamente:

1. **Fase de Detección (Uso del CAES):** el punto de partida es el análisis de errores reales. Se presenta al alumnado una selección de muestras del CAES que contienen formaciones no normativas con los sufijos objeto de estudio (p. ej., *mentabilidad*, *pegación* o *cantarear*). El objetivo es que los estudiantes identifiquen la anomalía y formulen hipótesis sobre por qué se ha producido el error (sobregeneralización de la regla o transferencia de la lengua materna). Esta fase está específicamente orientada a activar la conciencia morfológica del aprendiente, pues favorece la identificación de patrones formales, sobregeneralizaciones y regularidades derivativas a partir de datos reales de interlengua.
2. **Fase de Contraste (Uso de CORPES XXI y CODEA):** una vez detectado el problema, el alumnado recurre a los macrocorpus de la RAE (CORPES XXI) y al CODEA para verificar la vigencia, la frecuencia y los contextos de uso de los sufijos objeto de estudio. En esta etapa, se emplean búsquedas avanzadas para observar, por ejemplo, cómo el sufijo *-ción* presenta una elevada productividad en el ámbito administrativo o cómo *-ear* se ha consolidado como un mecanismo especialmente activo en la creación de verbos de carácter tecnológico (*tuitear*, *resetear*). Este mismo procedimiento se aplica de forma guiada al resto de sufijos seleccionados (*-dad*, *-oso/-osa*), permitiendo al estudiante contrastar patrones de uso y consolidar el conocimiento morfológico a partir de evidencia empírica.

¹¹ La propuesta didáctica ampliada, que incluye el tratamiento de todos los sufijos analizados, se presenta en el anexo de este trabajo.

3. **Fase de Visualización e Interpretación (*AntConc* y *Voyant Tools*):** en la fase final, se propone el uso de herramientas de visualización textual para mapear la productividad de los formantes seleccionados. Mediante *AntConc*, el alumnado genera líneas de concordancia con el fin de observar los contextos de uso de distintos sufijos y analizar, por ejemplo, en el caso de *-oso/-osa*, si los adjetivos resultantes presentan una carga semántica preferentemente positiva o negativa. Este mismo procedimiento se aplica de forma guiada a otros sufijos de la secuencia (*-ción*, *-dad*, *-ear*), se atiende a sus patrones combinatorios y valores semánticos. De manera complementaria, con *Voyant Tools* se elaboran nubes de palabras y gráficos de tendencias que permiten comparar visualmente la riqueza léxica de diferentes géneros textuales, lo que fomenta una reflexión metalingüística sobre la plasticidad y productividad del léxico.

5.2. Objetivos De Aprendizaje Y Beneficios Pedagógicos

La incorporación didáctica de herramientas digitales, aplicadas al análisis lingüístico en ELE, tiene como objetivo favorecer el desarrollo de una competencia léxico-constructiva y, por otro lado, aumentar la motivación intrínseca del estudiante, al vincular el aprendizaje del español con las competencias digitales propias del siglo XXI. Sin embargo, para que el enfoque sea riguroso, las herramientas empleadas deben seguir la cronología real de los textos, para evitar así, el sesgo de datar las obras por su composición en lugar de por su copia original. Solo así el aprendiente podrá situar correctamente los cambios lingüísticos y desarrollar la capacidad crítica para desentrañar cualquier palabra nueva.

6. Balance Y Horizontes: Hacia Una Pedagogía Digital De La Morfología

A tenor de la explicación teórica y la aplicación práctica desarrolladas en las secciones precedentes, es posible extraer un conjunto de directrices pedagógicas orientadas a la enseñanza de la morfología en ELE combinada con herramientas propias de las HD. Asimismo, se perfilan horizontes de indagación que inviten a profundizar en la simbiosis entre la lingüística aplicada y el ecosistema digital.

6.1. Recomendaciones Estratégicas Para El Aula

Para maximizar el impacto de la morfología léxica mediante la integración de recursos tecnológicos, se propone un modelo estructurado sobre tres pilares fundamentales que garantizan un aprendizaje profundo y significativo. En primer lugar, resulta imperativo establecer una priorización basada en la productividad y la progresión curricular de los sufijos propuestos. La selección de formantes debe responder a criterios de rentabilidad en el discurso contemporáneo —como ocurre con los sufijos *-ción* o *-ear*—, con ello se asegura una integración escalonada que acompañe al discente desde los estadios iniciales hasta alcanzar la sofisticación léxica característica de los niveles C1 y C2. Por tanto, esta

secuenciación permite que el estudiante asimile la red de formantes léxicos de manera armónica y coherente con el resto de las unidades de la lengua y con sus necesidades comunicativas.

En segundo lugar, la metodología debe fundamentarse en una praxis inductiva que desplace la instrucción directa tradicional hacia el ABD. Bajo este paradigma, los corpus lingüísticos trascienden su rol accesorio para erigirse en el núcleo rector de la actividad docente, para esto, plantean un escenario donde el estudiante asume el rol de investigador activo de la lengua. Dicho enfoque exige, no obstante, una actitud crítica frente a este tipo de herramientas, ya que la configuración de los repositorios no siempre permite verificar con exactitud cada ocurrencia, ni descartar errores derivados de las fases de procesamiento técnico. Desde tal rigor metodológico, el estudiante adquiere la competencia necesaria para contrastar sus hipótesis morfológicas frente a la evidencia empírica. Así, el aprendizaje léxico trasciende la memorización para convertirse en una indagación autónoma, dotada de rigor y sustento científico.

En última instancia, la metodología empleada requiere un análisis contextual que evite el estudio de la derivación como un fenómeno aislado. Se busca incentivar una reflexión profunda sobre la lengua, dada la naturaleza intrincada de los procesos morfológicos, que a menudo manifiestan irregularidades y comportamientos asistemáticos difíciles de categorizar. En este sentido, la capacidad de inferencia semántica puede abordarse mediante la exposición sistemática a la diversidad de registros documentados en corpus como el CORPES XXI o el CODEA, al tiempo que se incorpora al análisis una reflexión crítica sobre la procedencia y naturaleza de las fuentes. Esta aproximación contribuye a una interpretación más precisa de los datos lingüísticos y favorece el desarrollo de una actitud analítica ante la variación discursiva. Ello implica que el estudiante aprenda a discernir los riesgos de la datación textual, ya que, al clasificarse los testimonios por su fecha de producción y no por la de su copia, existe la posibilidad de atribuir a la lengua original fenómenos propios de copistas posteriores. De este modo, al interactuar con el entorno natural de la palabra y su realidad material, se consolida una conciencia morfológica, capaz de interpretar la vitalidad de un sufijo dentro de sus coordenadas históricas y filológicas exactas.

6.2. Horizontes Y Líneas De Indagación Futura

Con el propósito de consolidar la intersección entre las HD y la didáctica de lenguas, se identifican tres líneas de investigación prioritarias. En primer lugar, cobra relevancia el análisis del entramado cognitivo en el empleo de la morfología y la visualización de datos en el aula; para ello, se parte de la premisa de que los textos constituyen la fuente principal para el estudio de la evolución lingüística del léxico. Dado que la investigación histórica exige el dominio de herramientas metodológicas que permitan interpretar y comparar testimonios de diversos periodos y registros, resulta necesario indagar en qué medida la representación gráfica mediante recursos como Voyant Tools o AntConc optimiza estos

procesos. El objetivo es determinar si las visualizaciones digitales consolidan la memoria léxica con mayor eficacia que los métodos tradicionales. Al optimizar el procesamiento cognitivo, estas herramientas permiten al alumnado navegar con mayor soltura a través de la complejidad de los cambios históricos y estructurales de la lengua.

En segundo lugar, resulta necesario establecer protocolos claros para la integración del ABD, ante la ausencia de directrices en muchas guías docentes de ELE. En este marco, el uso de recursos como el CODEA, mediado por el docente, cumple una función contextualizadora al mostrar la trayectoria histórica de determinados formantes y explicar su productividad actual. El desafío metodológico consiste en diseñar estrategias que traduzcan la complejidad técnica de estos corpus en aplicaciones pedagógicas accesibles para el aula de ELE.

Finalmente, cualquier avance en esta disciplina requiere abordar la alfabetización digital crítica del profesorado¹². En definitiva, nuestra propuesta aspira a que el estudiante desarrolle una capacidad analítica para entender la forma de las palabras. Gracias a las herramientas digitales, el aprendiente de ELE puede desentrañar el uso real de los sufijos y organizar su vocabulario de manera empírica. Al integrar la visualización de datos y el análisis de corpus, la morfología se vuelve tangible y dota al hablante no nativo de una mayor independencia y rigor en su expresión con la lengua.

7. Conclusiones

Para concluir, esta propuesta plantea que el trabajo sistemático de la morfología derivativa, apoyado en herramientas idiosincrásicas de las HD, puede constituir un marco didáctico propicio para el desarrollo de la competencia léxica y de una mayor autonomía del aprendiente. La integración de repositorios digitales y la gestión estratégica de sufijos de alta rentabilidad —como *-ción*, *-dad*, *-oso/-osa* y el neológico *-ear*— se concibe como una vía potencial para favorecer la conciencia morfológica y la organización del léxico. La incorporación de una dimensión diacrónica, mediada por el docente, pretende además fomentar una actitud crítica y reflexiva hacia los procesos de formación léxica del español.

De forma simultánea, la explotación de corpus lingüísticos como CORPES XXI y CODEA, en combinación con herramientas de análisis y visualización como AntConc y Voyant Tools, se configura como un recurso didáctico relevante que favorece un ABD y promueve la reflexión metalingüística. Este enfoque facilita que los aprendientes analicen y comprendan los procesos de formación de palabras. Esta metodología trasciende la mera memorización

¹² Es fundamental explorar las barreras y actitudes de los docentes frente a la minería de textos, toda vez que la adopción de las HD en el aula de ELE depende de que el profesorado perciba estas herramientas no como una carga técnica adicional, sino como un aliado estratégico para la enseñanza del sistema léxico.

de las unidades léxicas para generar una conciencia capaz de descodificar discursos de alta complejidad en la lengua meta. Al combinar el conocimiento formal con escenarios comunicativos auténticos, la propuesta se articula conforme a los principios del aprendizaje significativo y contribuye a que los discentes de niveles avanzados desarrollen estrategias discursivas adecuadas para la producción de textos en contextos académicos y profesionales.

En definitiva, la convergencia entre morfología y HD en el currículo de ELE requiere que el docente desarrolle un criterio sólido para seleccionar y aplicar los recursos tecnológicos disponibles. La competencia para verificar la validez empírica de los datos es un requisito indispensable para explotar el potencial de los corpus; solo así se garantiza que la naturaleza cambiante del español contemporáneo se estudie con el rigor, la profundidad y la exactitud que el ámbito académico exige.

Referencias Bibliográficas

- Alvar Ezquerro, M., Azorín Fernández, D., & Sánchez Lancis, C. (1994). *Léxico y diccionarios*. Arco/Libros.
- Anthony, L. (2020). *AntConc (Version 4.3.1)* [Software de computación]. Waseda University. <https://www.laurenceanthony.net/software>
- Baralo, M. (2007). Adquisición de palabras: redes semánticas y léxicas. En *Actas del Foro de Español Internacional: Aprender y enseñar léxico* (pp. 384–399).
- Consejo de Europa. (2002). Marco común europeo de referencia para las lenguas: Aprendizaje, enseñanza, evaluación. Instituto Cervantes. https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/
- Cruz Escudero, J.R (2022). Sobre la alta productividad y empleabilidad del sufijo -oso/-osa en ELE. *MarcoELE: Revista de Didáctica del Español Lengua Extranjera*, 35
- Cruz Escudero, J.R. (2023). *Las unidades adjetivales por derivación en la enseñanza del español como LE: tratamiento, productividad y clasificación del error* [Tesis doctoral, Universidad de Jaén]. <https://hdl.handle.net/10953/230>
- Davies, M., y Parodi, C. (2022). *Corpus del Español: Historical/Genres y Web/Dialects*. Brigham Young University.
- Granger, S. (2012). Learner corpora. En C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1–8). John Wiley & Sons.
- GITHE. (2022). *CODEA+ (Corpus de documentos españoles anteriores a 1900)* [Base de datos]. <https://doi.org/10.37536/CODEA.2015>

- Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes: Niveles de referencia para el español* (3 vols.). Biblioteca Nueva.
https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/
- Instituto Cervantes & Universidade de Santiago de Compostela. (2022). Corpus de Aprendices de Español como Lengua Segunda y Extranjera (CAES) [Corpus].
<https://galvan.usc.es/caes/>.
- Liceras, J.M. (2009). La morfología léxica del español y el llamado problema lógico de la adquisición del lenguaje no nativo. En D. Serrano-Dolader, M.^a A. Martín Zorraquino y J.F. Val Álvaro (eds.), *Morfología y español como lengua extranjera (E/LE)* (pp. 21-66). Prensas de la universidad de Zaragoza.
- Moreno Fernández, F. (2009). *La lengua española en su geografía*. Madrid: Arco/Libros.
- Real Academia Española. (s. f.). *Banco de datos (CORPES XXI) [Corpus del Español del Siglo XXI]*. <https://www.rae.es/corpes/>
- Real Academia Española & Asociación de Academias de la Lengua Española. (2014). Entradas -ico, -ista, -oso, -udo. En *Diccionario de la lengua española*.
<https://dle.rae.es>
- Serrano-Dolader, D. (2006). La morfología léxica en el ámbito del español como lengua extranjera. En E. Felú Arquíola (ed.): *La morfología a debate*, (pp.145-169). Universidad de Jaén.
- Serrano-Dolader, D. (2018). *Formación de palabras y enseñanza del español LE/L2*. Routledge Advances.
- Sinclair, S., y Rockwell, G. (2016). *Voyant Tools* [Software]. <https://voyant-tools.org/>
- Zholobova, A. (2014). El CORDE como herramienta para el estudio diacrónico del español. *Lingüística Histórica*, 12(1), 33–52

Nota sobre el uso de la IA

Este trabajo ha contado con la asistencia de la IA (Gemini, Google) para las tareas de corrección de estilo en algunos fragmentos confusos, como asistente de traducción en los idiomas portugués e inglés y la verificación del formato de citación APA. El proceso de supervisión y la redacción original han sido realizados íntegramente por el autor

Recibido 21/01/2026

Aceite 07/05/2026

Publicado 19/05/2026

Este artículo está disponible bajo una licencia [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

ANEXO I

PROPUESTA DIDÁCTICA: DEL DATO A LA COMPETENCIA MORFOLÓGICA

1. Contextualización

- Nivel: B2–C2 (MCER)
- Duración: 2–3 sesiones (90 minutos cada una)
- Tipo de curso: ELE general / fines académicos
- Modalidad: presencial o virtual

2. Objetivos

➤ Objetivo general

Desarrollar la conciencia morfológica del alumnado mediante el análisis guiado de datos reales, con el fin de favorecer la comprensión y el uso de los procesos de formación de palabras en español.

➤ Objetivos específicos

- Identificar formaciones léxicas no normativas y explicar su origen.
- Reconocer patrones de uso de los sufijos *-ción*, *-dad*, *-oso/-osa* y *-ear*.
- Analizar la productividad y los valores semánticos de dichos sufijos.
- Utilizar corpus y herramientas digitales para la exploración lingüística.
- Aplicar los conocimientos adquiridos a la producción léxica contextualizada.

3. Contenidos

➤ Lingüísticos

- Morfología derivativa: sufijación
- Reglas de formación de palabras (RFP)
- Productividad y restricciones morfológicas

➤ Estratégicos

- Inferencia léxica
- Análisis de errores
- Observación de patrones lingüísticos en corpus

➤ Tecnológicos

- Consulta de corpus (CAES, CORPES XXI, CODEA)
- Uso de herramientas de análisis (AntConc, Voyant Tools)

4. Secuencia didáctica

➤ Fase 1. Detección (CAES)

Actividad 1: “Detectives del léxico”

- **Objetivo:** activar la conciencia morfológica mediante el análisis de errores reales.
- **Material:** fragmentos de textos del CAES con errores léxicos (ej.: *mentabilidad*, *pegación*, *cantarear*)

- **Organización:** trabajo en parejas
 - **Procedimiento:**
 - Los estudiantes leen los ejemplos proporcionados.
 - Identifican las formas no normativas.
 - Proponen posibles formas correctas.
 - Formulan hipótesis sobre el origen del error:
 - sobregeneralización,
 - analogía,
 - influencia de la L1.
 - **Producto esperado:**

Tabla con:

 - error detectado
 - forma correcta
 - explicación del proceso
- **Fase 2. Contraste (CORPES XXI + CODEA)**
Actividad 2: “¿Existe o no existe?”
- **Objetivo:** verificar patrones de uso y productividad mediante evidencia empírica.
 - **Material:** acceso guiado a CORPES XXI y CODEA
 - **Organización:** grupos
 - **Procedimiento:**
 - Los estudiantes buscan las formas analizadas en los corpus.
 - Identifican:
 - frecuencia de aparición,
 - contextos de uso,
 - variantes correctas.
 - Analizan ejemplos reales de:
 - *-ción* en textos administrativos,
 - *-ear* en contextos tecnológicos (*tuítear, resetear*).
 - **Producto esperado:**

Informe breve con:

 - capturas o ejemplos de corpus
 - conclusiones sobre uso y productividad
- **Fase 3. Visualización e interpretación (AntConc + Voyant)**
Actividad 3: “Radiografía de los sufijos”
- **Objetivo:** representar y analizar patrones léxicos de forma visual.
 - **Material:** corpus descargado / pequeños textos + software
 - **Organización:** individual o parejas
 - **Procedimiento:**
 - Con AntConc:
 - generación de concordancias,
 - análisis de contexto (*-oso/-osa, -ear...*).
 - Con Voyant Tools:
 - creación de nubes léxicas,
 - análisis de frecuencia.
 - Discusión guiada:

- ¿qué valores semánticos predominan?
 - ¿hay tendencias positivas o negativas?
 - ¿qué sufijos son más productivos?
 - **Producto esperado:** breve exposición oral o escrita con interpretación de resultados
- 5. Actividad final (transferencia)**
- Los estudiantes producen un breve texto (200 palabras) en el que:
 - utilicen al menos 5 derivados trabajados,
 - incluyan una forma innovadora (neologismo controlado),
 - justifiquen su formación.
 - Permite evaluar:
 - competencia léxica,
 - conciencia morfológica,
 - creatividad controlada.
- 6. Evaluación**
- **Instrumentos**
 - Rúbrica de análisis de errores
 - Informe de corpus
 - Producción final
 - **Criterios**
 - Identificación de patrones morfológicos
 - Uso adecuado de derivados
 - Interpretación de datos