



# LA FORMACIÓN EN LINGÜÍSTICA DE CORPUS DENTRO DEL MARCO DE LAS INDUSTRIAS DE LA LENGUA

Education in Corpus Linguistics within the framework of Language Industry

ERIKA VEGA MORENO<sup>1</sup>

<sup>1</sup> Universidad de Cádiz, España

---

## KEYWORDS

Linguistics  
Applied Linguistics  
Language Industry  
Corpus Linguistics  
Education  
Methodology  
Research

---

## ABSTRACT

*Education in Corpus Linguistics is part of the subject "Language Industry" from the degree in "Linguistics and Applied Languages" of the University of Cadiz. The main difficulty of this subject is that, in addition to being a subject itself, some of its contents are also studied in other specific courses of the same degree. That is why this study is of interest, because its significant results showed that the teaching staff's research experimentation and the implementation of acquired knowledge improved the students' final marks and also achieved greater student involvement. Furthermore, this study contributes to the enrichment of professors' research work.*

---

## PALABRAS CLAVE

Lingüística  
Lingüística aplicada  
Industrias de la lengua  
Lingüística de corpus  
Formación  
Metodología  
Investigación

---

## RESUMEN

*Dentro del grado en "Lingüística y lenguas aplicadas" de la Universidad de Cádiz, la formación en lingüística de corpus se enmarca en la materia "Industrias de la lengua". La complejidad de esta reside en que, a pesar de contar con una asignatura propia, también se aborda en otras específicas. Es ahí donde estriba el interés de este estudio, con el que obtuvimos interesantes resultados, ya que observamos que la experimentación investigadora y la puesta en práctica de conocimientos adquiridos optimizaron las notas finales y consiguieron una mayor implicación del alumno, además de enriquecer la labor investigadora del profesorado.*

---

Recibido: 07/ 07 / 2022

Aceptado: 20/ 09 / 2022

## 1. Introducción

La formación en lingüística de corpus es una de las líneas de actuación que abordamos en la asignatura de “Industrias de la Lengua”, materia optativa del tercer curso del grado en “Lingüística y lenguas aplicadas” de la Universidad de Cádiz.

Las *Industrias de la lengua* es campo de investigación cuyo objeto de estudio es el diseño, la producción y la comercialización de herramientas, productos y servicios relacionados con el tratamiento informatizado de las lenguas. Son muchos los campos y áreas de actuación de este sector, que nace principalmente de la relación entre la lingüística y la informática, entre los que podemos destacar principalmente: la lingüística computacional, la lingüística de corpus, la neología, la terminología y la lexicografía.

Hay que tener en cuenta que, dentro del citado grado, encontramos asignaturas específicas dedicadas al estudio de algunas de estas disciplinas aplicadas, concretamente:

- Dentro del módulo dirigido a las “Aplicaciones de las lenguas”, localizamos la materia “Lingüística computacional”, asignatura obligatoria en el segundo curso dentro del cuarto semestre;
- En la sección que se ocupa de las “Lenguas para fines específicos”, hallamos la de “Terminología y variación especialidad”, también obligatoria, pero en el tercer curso dentro del quinto semestre, y,
- En el grupo de asignaturas destinadas a la “Evaluación y tratamiento del lenguaje y las lenguas”, se oferta la de “Técnicas en la elaboración de diccionarios”, optativa en el cuarto curso dentro del séptimo semestre.

Sin embargo, observamos que, en el plan de estudio de la citada titulación, la “Neología” y la “Lingüística de corpus”, no cuentan con asignaturas propias, por lo que estas se convierten en los pilares fundamentales de esta materia optativa.

Así, en esta aportación, nos centraremos concretamente en el aprendizaje/enseñanza de la segunda, este es, el de la Lingüística de corpus, puesto que ya dedicamos diferentes investigaciones y publicaciones al análisis del aprendizaje/enseñanza teórico, práctico y metodológico de la disciplina neológica (cf. Vega Moreno 2020a y 2020b), completando de esta forma el estudio del proceso de aprendizaje/enseñanza en la asignatura “Industrias de la lengua” en dicha institución.

Esta contribución, al igual que las citadas con anterioridad (cf. Vega Moreno 2020a y 2020b), está vinculada al Proyecto de Innovación y Mejora Docente “Acciones para la potenciación del enfoque científico en las asignaturas del grado de Lingüística y lenguas aplicadas a través del Instituto de Lingüística Aplicada de la Universidad de Cádiz” (Ref. sol-202000161929-tra), dirigido por la Profa. Dra. Carmen Varo Varo y desarrollado por miembros del área de Lingüística General de la Universidad de Cádiz, e inserta en la línea de investigación en “Lingüística de corpus” del “Gabinete de Industrias de la Lengua” del “Instituto Universitario de Investigación en Lingüística Aplicada” de la Universidad de Cádiz.

## 2. Objetivos

En líneas generales, en esta asignatura, materia optativa del tercer curso del citado grado, impartida dentro del quinto semestre, dentro del módulo “Teoría lingüística y tecnologías del lenguaje”, se plantea, entre otras cuestiones, como objetivo general, elaborar un trabajo de investigación en lingüística de corpus, requisito indispensable para la superación de la asignatura, experimentando así, la experiencia científica en dicho campo de actuación.

Para la consecución de este fin, es necesario establecer otros objetivos específicos, entre los que podemos destacar los siguientes:

- a) definir la *lingüística de corpus*, analizando esta como metodología científica, viendo sus antecedentes, la primera y la segunda generación de lingüística de corpus, y su renacer actual;
- b) explicar el *corpus* y señalar las características que este presenta en la actualidad, revisando además sus posibles tipologías atendiendo a diferentes parámetros y analizando su posible desarrollo y elaboración, y
- c) revisar distintos corpus y bases de datos existentes en diferentes campos lingüísticos.

Para la puesta en práctica del procedimiento de aprendizaje/enseñanza planteada, en esta asignatura se establecieron cuatro fases en la instrucción del estudiante universitario:

1º) en la primera, se estudió el concepto *lingüística de corpus*, analizando esta como metodología de investigación, viendo sus antecedentes, la primera y la segunda generación de lingüística de corpus, y su renacer actual, cumpliendo así con el objetivo específico a;

2º) en la segunda, se analizó la noción de *corpus* y su evolución a lo largo de la historia, se examinó su concepción actual y sus posibles características, se revisó su tipología atendiendo para ello a distintos criterios y se explicó su posible desarrollo y elaboración, ejecutando de esta forma el objetivo específico b;

3º) en la tercera, se revisaron distintas bases de datos existentes para diferentes disciplinas lingüísticas, sus usos y aplicaciones potenciales en diversos campos de la lingüística aplicada, obedeciendo de este modo al último objetivo específico, concretamente, el c, y;

4º) en la cuarta y última, se solicitó al alumnado que eligiera una disciplina lingüística y que planteará una investigación en la que la lingüística de corpus formara parte de su metodología de trabajo, poniendo así en práctica los conocimientos adquiridos, respondiendo al objetivo general expuesto.

### 3. Metodología

La metodología planteada para llevar a cabo el procedimiento de aprendizaje/enseñanza trazado fue estructurado, como ya avanzamos anteriormente (cf. apdo. 2), en cuatro fases, las cuales serán descritas a continuación.

#### 3.1. Primera etapa

Dentro de la primera etapa, dedicada al estudio de la *lingüística de corpus*, se establecieron, a su vez, cuatro pasos en su desarrollo:

a) en primer lugar, analizar sus antecedentes desde su origen hasta el siglo XIX y desde finales del siglo XIX hasta mediados del siglo XX;

b) en segundo lugar, examinar su evolución en la primera generación de lingüística de corpus y comentar las críticas recibidas tanto las de carácter teórico como las de índole práctica;

c) en tercer lugar, revisar su desarrollo en la segunda generación de lingüística de corpus y refutar los comentarios teóricos y aplicados anteriormente recibidos, y

d) en cuarto y último lugar, comentar el renacer actual de la lingüística de corpus con la aparición del ordenador como parte de su metodología de trabajo.

De este modo, para abordar el primero, partimos de los precedentes planteados por Villayandre Llamazares (2008, pp. 330-331), diferenciado, por un lado, una primera etapa, que iba desde su origen hasta el siglo XIX y, por otro, una segunda, que engloba desde finales del siglo XIX hasta mediados del siglo XX, estableciendo así un primer concepto de corpus previo al ordenador.

Igualmente, en el tratamiento del segundo, siguiendo a Villayandre Llamazares (2008, pp. 331-333), revisamos la *primera generación de lingüística de corpus* basada en la observación de los datos como metodología empírica; plantemos el segundo concepto de corpus, que es mucho más estructuralista, apuntado por la cita autora, y señalamos las críticas realizadas por Chomsky de carácter teórico y las formuladas por Abercrombie de índole aplicada.

De igual manera, en la ejecución del tercero, analizamos, como comentaba Villayandre Llamazares (2008, pp. 333-336), la *segunda generación de lingüística de corpus* marcada por la presencia del ordenador; comentamos su tercer concepto de corpus, y contraargumentamos los comentarios recibidos por parte de Chomsky y Abercrombie a partir de la opinión expuesta por Leech (1992).

Para cerrar esta etapa, en la realización del cuarto, según señala Villayandre Llamazares (2008, pp. 336-340), explicamos el renacer actual de la lingüística de corpus con la aplicación de las nuevas posibilidades que ofrecía el ordenador y señalamos su cuarto concepto de corpus, entendido en el sentido más actual del término.

#### 3.2. Segunda etapa

Asimismo, en la segunda fase, destinada a la revisión de la noción de *corpus*, se plantearon, a su vez, tres pasos:

a) en primer lugar, revisar el desarrollo del concepto de *corpus* a lo largo de la historia;

b) en segundo lugar, analizar las características que debe presentar un corpus en la actualidad, y

c) en tercer lugar, plantear las diversas tipologías existentes para su clasificación.

En este sentido, para trabajar el primero, examinamos la transformación que había sufrido el concepto de *corpus* con el paso del tiempo, partiendo de sus antecedentes hasta llegar a su concepción actual, de manera que explicamos, como señalaba Villayandre Llamazares (2008, pp. 330, 331, 334 y 338), las diferentes concepciones del mismo:

corpus<sup>1</sup>

a) Ser un conjunto de textos escritos (datos).

b) Tener como finalidad el estudio y descripción de lenguas muertas (latín, griego...).

c) Ser necesario, desde un punto de vista metodológico, para llevar a cabo los estudios lingüísticos: esos datos constituían el único acercamiento posible, pues esas lenguas ya no contaban con hablantes vivos que utilizaran la lengua como vehículo de comunicación en una sociedad.

(Villayandre Llamazares, 2008, p. 330).

corpus<sup>2</sup>

a') Ser un conjunto de grabaciones y transcripciones en papel (datos), en la mayoría de los casos fichas almacenadas en cajas.

b') Tener por finalidad el estudio de lenguas vivas, pero no documentadas previamente por escrito (lenguas amerindias).

c') Ser necesario, ya que la recogida de datos orales era la única forma de acceder al conocimiento de esas lenguas, nunca antes estudiadas.

d') Centrarse en los aspectos fonéticos y (morfo)fonológicos, niveles en los que es posible realizar un inventario de todos los elementos implicados dada la naturaleza finita de sus paradigmas.

e') No atender a cuestiones de representatividad debido a que el análisis de los datos debía efectuarse de forma manual y visual era imposible manejar un número elevado de datos, de ahí una de las principales críticas que recibirá la metodología, por su parcialidad a la hora de describir la realidad.

(Villayandre Llamazares, 2008, p. 331).

### corpus<sup>3</sup>

a'') La presencia del ordenador: solo en los años sesenta los ordenadores alcanzaron una potencia de procesamiento y una capacidad de almacenamiento suficientes para poder albergar grandes cantidades de texto, aunque en un principio no todos los proyectos para recopilar corpus se concebían pensando en su informatización. No obstante, el vínculo entre los corpus y los ordenadores ya había sido establecido a finales de los cuarenta por R. Bussa.

b'') Carácter representativo de los datos: la mayoría de los proyectos de elaboración de corpus pretendía recoger textos escritos que dieran cuenta del estado de la lengua en ese momento. Durante la década de los cincuenta A. Juilland estableció los conceptos de marco de la muestra, representatividad y equilibrio, básicos en el concepto actual de corpus.

c'') Tendencia a desfavorecer los datos orales por las dificultades técnicas y de transcripción. Predominan los corpus de textos escritos, aunque con notables excepciones.

d'') Tamaño: un millón de palabras.

(Villayandre Llamazares, 2008, p. 334).

### corpus<sup>4</sup>

a''') Formato electrónico: conjunto de textos informatizados.

b''') Tamaño en aumento progresivo, hasta superar los cien millones de palabras, aunque existen corpus de tamaños inferiores.

c''') Carácter abierto: corpus no cerrados, sino en continua actualización (corpus monitor).

d''') Vertiente comercial: los corpus no se limitan a centros de investigación o instituciones vinculadas a la lengua, sino que muchos proyectos son desarrollados por consorcios comerciales, principalmente editoriales, o empresas de telecomunicaciones en el caso de los corpus orales.

e''') Se amplía el repertorio de lenguas que disponen de corpus, y también se elaboran corpus multilingües.

f''') Automatización de las diferentes tareas de procesamiento de los textos de un corpus gracias a los avances técnicos que permiten realizar, de forma automática o semiautomática, la asignación de categoría gramatical a cada una de las palabras del corpus, la desambiguación, la extracción de concordancias o ejemplos en contexto, el alineamiento de las grabaciones de audio con su correspondiente transcripción, etc.

g”) Estímulo para el desarrollo de nuevos modelos y campos de investigación en Lingüística, así como para la realización de estudios sobre los más variados aspectos lingüísticos, desde los gramaticales hasta los discursivos, pasando por los históricos, los psicolingüísticos o los culturales.

(Villayandre Llamazares, 2008, p. 338).

De igual modo, para atender el segundo, nos basamos, como afirma Villayandre Llamazares (2008, pp. 340-341), en las características que debe presentar un corpus en la actualidad, entre las que podemos destacar:

- El formato electrónico, así que este debe estar informatizado, permitiendo automatiza tareas como la búsqueda de información, recuperar datos, medir frecuencia y clasificar contenidos en función de los parámetros que el investigador considere;
- La autenticidad de los datos, de manera que los datos recogidos en la base de datos deben ser muestras reales y originales de uso de la lengua que es objeto de estudio;
- La elección de unos criterios de selección, que pueden ser lingüísticos y/o extralingüísticos, los cuales se aplicarán de manera similar en todo el proceso de recogida y regularán la elección del registro de las informaciones que se recojan en él;
- La representatividad, la cual garantizan que las evidencias recolectadas de dicha variedad de la lengua son lo suficientemente representativas de dicha utilización, y
- El tamaño finito de estos, que se suele fijar al principio de la investigación y cuantificar en millones de palabras.

Para finalizar esta fase y ejecutar el tercero, partimos, siguiendo a Villayandre Llamazares (2008, pp. 342-343), en la aplicación de, al menos, siete parámetros, distinguiendo:

- Según la modalidad de la lengua, *corpus escritos, orales y mixtos*;
- Según el número de lenguas, *corpus monolingües, bilingües y multilingües*;
- Según los límites y capacidad de actualización, *corpus cerrados y abiertos*,
- Según la especificidad delos textos por su contenido, *corpus generales y especializados*;
- Según el período cronológico que abarca, *corpus diacrónicos y sincrónicos*;
- Según el tamaño de los textos, *corpus de referencia y textuales, y*
- Según al tratamiento informático al que se someta, *corpus simple y codificados o anotados*.

### 3.3. Tercera etapa

Así, en la tercera, empleada en el examen de diferentes corpus existentes, se revisaron diversas bases de datos disponibles en la web para diferentes campos lingüísticos y se comentaron algunos de sus posibles usos en diversas actuaciones de la lingüística aplicada, de manera que,

- Dentro de la lexicografía, señalamos, entre otros:
  - El *CREA. Corpus de Referencia del Español Actual* (cf. <http://corpus.rae.es/creanet.html>),
  - El *CORDE. Corpus Diacrónico del Español* (cf. <http://corpus.rae.es/cordenet.html>),
  - El *CORPES XXI. Corpus del Español del Siglo XXI* (cf. <http://web.frl.es/CORPES/view/inicioExterno.view>), y
  - El *CDH. Corpus del Nuevo diccionario histórico del español* (cf. <http://web.frl.es/CNDHE/view/inicioExterno.view;jsessionid=BDCA17DB5A991FD09667AA9E2BCA177C>);
- dentro del campo de la neología, estudiamos, entre otros:
- El *BOBNEO. Banco de datos de neologismos del Observatorio de Neología* de la Universitat Pompeu Fabra (cf. <http://obneo.iula.upf.edu/bobneo/index.php>) y
  - el *Banco de neologismos* del Centro Virtual Cervantes (cf. [https://cvc.cervantes.es/lengua/banco\\_neologismos/busqueda.asp](https://cvc.cervantes.es/lengua/banco_neologismos/busqueda.asp));
  - dentro de la disciplina terminológica, consultamos, entre otros:
  - el *CercaTerm. Base de datos del Centro de Terminología (TermCat)* (cf. <https://www.termcat.cat/es/cercaterm>),
  - el *EuskalTerm. Banco terminológico público vasco* (cf. <https://www.euskadi.eus/euskalterm/>),
  - el *IATE. Interactive Terminology for Europe* (cf. <https://iate.europa.eu/home>),
  - el *TermOTAN. Base de datos terminológica oficial de la OTAN* (cf. <https://nso.nato.int/natoterm/content/nato/pages/home.html?lg=fr>) y
  - el *UBTerm. Base de datos terminológica de la Universidad de Barcelona* (cf. <http://www.ub.edu/ubterm/>);
  - dentro de la lingüística clínica, comentamos, entre otros:
  - el *Corpus PerLA. Percepción, Lenguaje y Afasia* (cf. <https://www.uv.es/perla/CorpusPerla.htm>), y
  - el *Corpus de trastornos del lenguaje* de la Universidad de Cádiz;

- Dentro de la materia sociolingüística, vimos, entre otros:
  - El *CSCM. Corpus sociolingüístico de la ciudad de México* (cf. [https://lef.colmex.mx/corpus\\_sociolinguistico.html](https://lef.colmex.mx/corpus_sociolinguistico.html)) y
  - El *COSER. Corpus Oral y Sonoro del Español Rural* (cf. <http://www.corpusrural.es/index.php>), y
- Dentro de la enseñanza de lenguas, encontramos, entre otros:
  - El *Indexador de Corpus de Aprendices de Español* (cf. [http://repositorios.fdi.ucm.es/corpus\\_aprendices\\_Espa%c3%b1ol/view/paginas/view\\_paginas.php?id=1](http://repositorios.fdi.ucm.es/corpus_aprendices_Espa%c3%b1ol/view/paginas/view_paginas.php?id=1)),
  - El *CAES. Corpus de aprendices de español* (cf. [https://www.cervantes.es/lengua\\_y\\_ensenanza/tecnologia\\_espanol/caes.htm](https://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm)) y
  - El *CORLSE. Corpus de la Lengua de Signos Española* (cf. <https://corpulse.es/>).

### 3.4. Cuarta etapa

Por último, en la cuarta fase, se solicitó al alumnado que eligiera una disciplina lingüística y que planteará una investigación en la que la lingüística de corpus formará parte de su metodología de trabajo, poniendo así en práctica los conocimientos adquiridos, a través de una tarea de evaluación.

Este trabajo científico debía estar estructurado en cuatro apartados, que son los que se detallan a continuación:

- Una introducción, que tenía que recoger al menos el tema de investigación, los objetivos generales y/o específicos, la estructura a seguir, describiendo brevemente sus contenidos (marco teórico, apartado metodológico y parte aplicada, todos estos con sus respectivos apartados y subapartados), y si el estudiantado lo consideraba necesario, podía además añadir una justificación y unas hipótesis;
- Un desarrollo, el cual estaba obligado a centrarse fundamentalmente en la parte metodológica, en la que se describiría las estrategias de trabajo empleadas y, sobre todo, el papel del corpus dentro del trabajo práctico, añadiendo, si fuera necesario, las fases del procedimiento propuesto y las herramientas necesarias;
- Unas conclusiones, que debían contener una valoración de lo más relevante del tema y todo lo que el alumno considerase más significativo de su contribución particular, de manera que los resultados recogidos en ellas se derivasen de la discusión realizada en los apartados anteriores, cumpliendo lo anunciado en la introducción, y
- Unas referencias, en la que se debían señalar los datos bibliográficos de los estudios citados a lo largo de esta investigación.

De esta forma, la sección más importante de esta tarea era el desarrollo, puesto que era donde se detallaba el procedimiento de trabajo, pudiendo seguir varias posibilidades de estudio:

- a) el empleo nuclear del corpus como eje central de la investigación, en el que se debía describir su elaboración, los criterios empleados en su confección, sus características, algunos ejemplos, etc.;
- b) el uso parcial de una base de datos externa al estudiante para el desarrollo de cualquier otra cuestión importante, en la cual debía especificarse qué corpus emplearía, su modo de uso, su papel en el estudio concreto, etc., y
- c) la utilización combinada de los mismos, siendo este modo el más altamente valorado, en el que se empleaba estos conjuntos de datos tanto de manera completa como forma superficial.

En resumen, el trabajo sugerido al alumno debía seguir dos pasos en su confección en el aula:

- a) en la primera, debía seleccionar un campo de actuación de la lingüística aplicada y pensar una posible investigación en la que la lingüística de corpus formará parte de su metodología de trabajo, bien de forma completa, bien de modo parcial, bien de manera combinada, siendo esta investigación inédita y original, y
- b) en la segunda, a partir de este trabajo científico, tenía que elaborar su propia propuesta metodológica, que podía estar o no apoyada en las vistas y trabajadas en clase, y aplicarla al campo de estudio seleccionado, presentando su trabajo y explicando el papel que tiene la lingüística de corpus dentro de sus técnicas de trabajo.

Para finalizar esta última fase, se expuso al alumnado dos ejemplos de casos prácticos, en el que se aunaron los dos campos de actuación fundamentales de esta asignatura, es decir, la neología y la lingüística de corpus, dando cohesión a todo lo explicado en ella.

Igualmente, en esta ejemplificación, se tomó como base el tema de investigación de la creación neológica en el lenguaje de la publicidad, en el que se plantearon:

- Como objetivo principal, analizar las tendencias de las unidades neológicas en dicho ámbito, y
- Como objetivos específicos, entre otros, delimitar el concepto de *neologismo*, estudiar el papel de la neología dentro de la publicidad, elaborar un corpus de neologismos, establecer criterios para detectar unidades neológicas, etc.

Así, en nuestra metodología, se combinaba el uso nuclear de la lingüística de corpus, a la vez que su utilización parcial, ya que planteamos:

- Por un lado, la creación de un corpus propio de neologismos extraídos de la publicidad para estudiar las Tendencias de este tipo de unidades dentro de este sector y,

- Por otro, el empleo de dos corpus para medir neologicidad de una voz y detectar candidatos a neologismos, para crear el corpus señalado en la tarea anterior, concretamente:
  - El *BOBNEO. Banco de datos del Observatorio de Neología* (cf. <http://obneo.iula.upf.edu/bobneo/index.php>) y
  - El *CREA. Corpus de referencia del español actual* (cf. <http://corpus.rae.es/creanet.html>).

De esta forma, para completar la explicación, mostramos las fases de trabajos seguidas en esta investigación, además de las herramientas necesarias para su ejecución, en la que el uso de los corpus tenía un papel primordial y combinado, que era la característica de la tarea más altamente calificable. De este modo, a partir de toda la información expuesta,

- Mientras que el alumno llevaba a cabo su propia investigación en lingüística de corpus, seleccionado un campo de actuación de la lingüística aplicada, el cual no podía repetirse, dándole libertad de elección;
- Comenzaba la labor enriquecedora tanto del profesor como del alumnado, ya que, a partir de las diferentes dudas y/o propuestas planteadas, se producía en clase un debate muy beneficioso sobre el análisis y la reflexión de las cuestiones presentadas.

#### 4. Resultados

Con la aplicación de la metodología anteriormente propuesta, obtuvimos interesantes resultados, que comparamos con los datos del curso anterior, teniendo en cuenta tres cuestiones:

- a) el índice de alumnos que ejercieron su derecho a la evaluación,
- b) a partir de la asistencia a la evaluación, el porcentaje de alumnado que superó la asignatura, y
- c) teniendo en cuenta el total de alumnos que tenían ese derecho a evaluación, los datos referentes a las calificaciones finales obtenidas en la materia.

De esta manera, si atendemos al índice de asistencia a la evaluación, observamos en dicho análisis los siguientes datos:

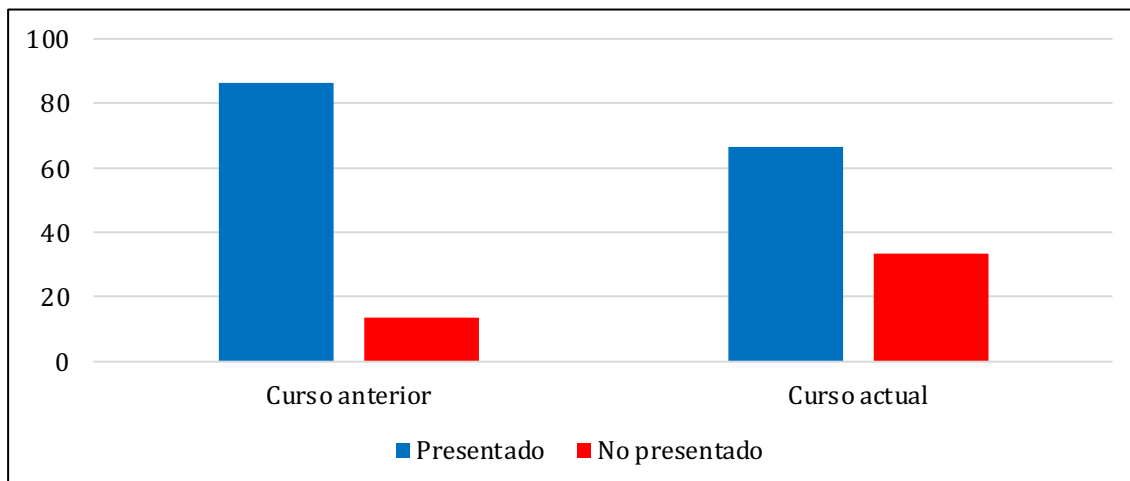
Tabla 1. Asistencia a la evaluación

	Curso anterior	Curso actual
<b>Presentado</b>	86,36%	66,67%
<b>No presentado</b>	13,74%	33,33%
<b>Total</b>	100,00%	100,00%

Fuente: Universidad de Cádiz, 2022.

Así, encontramos que hay una diferencia cuantitativa significativa entre ambos cursos, ya que, mientras que en el curso pasado ejercieron su derecho a evaluación un 86,36% del alumnado, frente al otro 13,64% que no lo hizo; en el curso actual ha habido un descenso considerable en el número de alumnos que ejecutaron esta opción, bajando hasta el 66,67%, frente al 33,33% que no lo ejerció, subiendo la no participación del estudiantado en las actividades de evaluación, mostrando tal vez que, con las metodologías plateadas en el curso actual, el alumno no se sentía tan capaz de realizar dichos ejercicios para la superación de la asignatura. Estos datos son recogidos en el siguiente gráfico:

Gráfico 1. Asistencia a la evaluación



Fuente: Universidad de Cádiz, 2022.

Asimismo, a partir del porcentaje de alumnos que ejecutaron su derecho a evaluación, obtuvimos en dicho estudio los siguientes resultados:

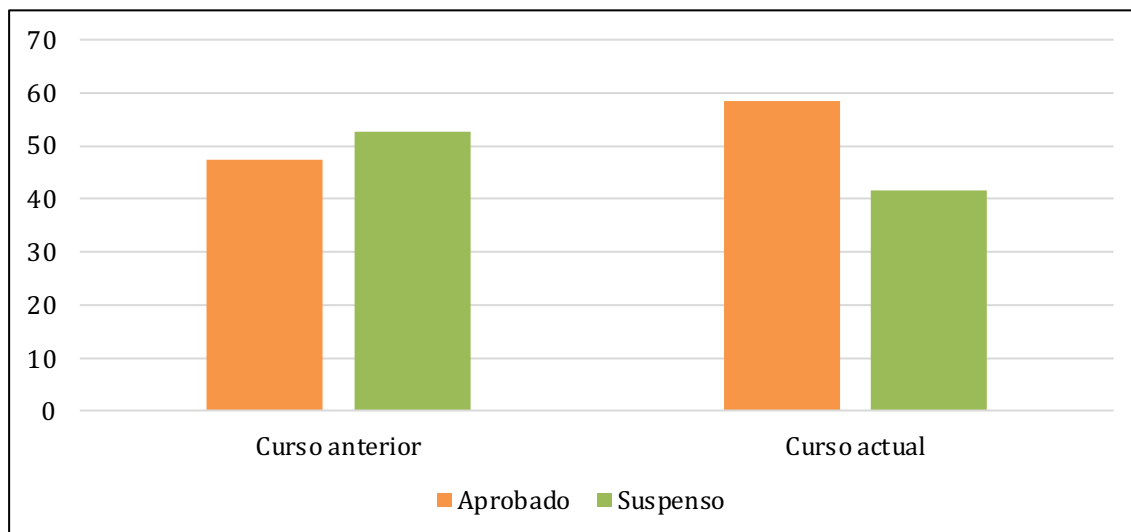
Tabla 2. Superación de la asignatura

	Curso anterior	Curso actual
<b>Aprobado</b>	47,37%	58,33%
<b>Suspenseo</b>	52,63%	41,67%
<b>Total</b>	100,00%	100,00%

Fuente: Universidad de Cádiz, 2022.

De este modo, vimos que, a diferencia del caso anterior, los resultados de superación de la misma fueron bastante opuestos. Mientras que en el curso pasado el 47,37% de los alumnos que asistieron a la evaluación obtuvieron un resultado satisfactorio, frente al 52,63% que no lo consiguió; en el curso actual hubo un ligero incremento, alcanzando un 58,33% de datos positivos, frente al 41,67% que no lo hizo, habiendo una reducción significativa de resultados negativos, demostrando que, aunque el alumnado no se sentía tan cualificado para superar el curso, si había adquirido mucho mejor las destrezas necesarias para realizar los ejercicios de evaluación con este planteamiento metodológico, frente al empleado en el curso anterior. Estos datos son recogidos en el siguiente gráfico:

Gráfico 2. Superación de la asignatura



Fuente: Universidad de Cádiz, 2022.

Finalmente, teniendo en cuenta el total de alumnos que tenían ese derecho a evaluación, que algunos aprovecharon y otros no, en dicha observación extrajimos los siguientes datos:

Tabla 3. Notas finales

	Curso anterior	Curso actual
<b>Matrícula de honor</b>	0,00%	0,00%
<b>Sobresaliente</b>	13,64%	16,67%
<b>Notable</b>	22,73%	16,67%
<b>Aprobado</b>	4,55%	5,56%
<b>Suspenseo</b>	45,45%	27,78%
<b>No presentado</b>	13,64%	33,33%
<b>Total</b>	100,00%	100,00%

Fuente: Universidad de Cádiz, 2022.

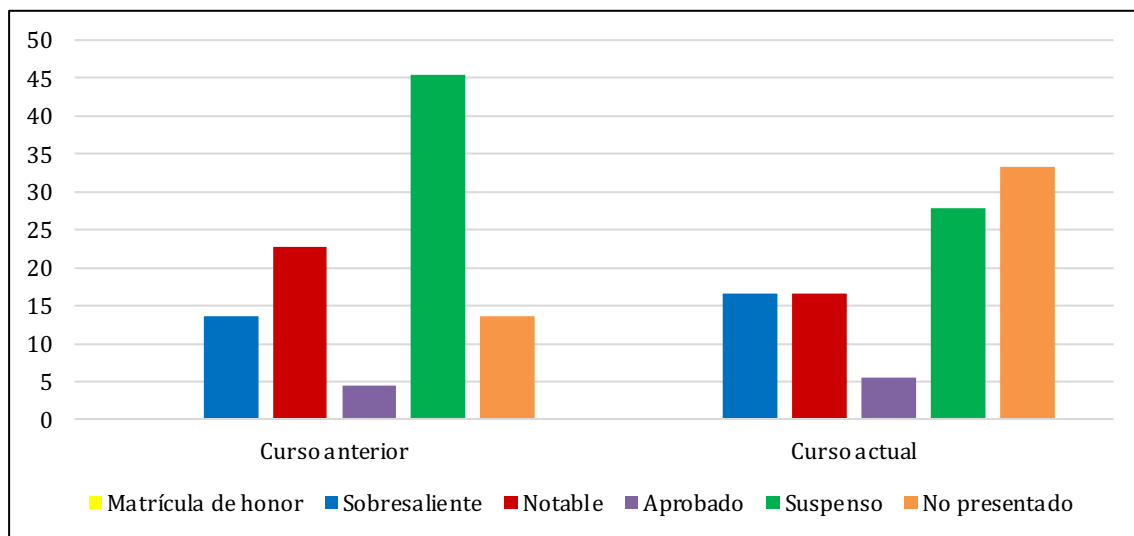


De esta manera, notamos que, mientras que entre el curso pasado y el actual no hubo una diferencia considerable en lo que respecta a la obtención de matrículas de honor; si se produjo un aumento leve destacable en el porcentaje de sobresalientes, encontrando un 13,64% en el curso anterior y un 16,67% en el actual.

No obstante, algo distinto ha ocurrido en lo que respecta a las calificaciones de notable y aprobado, mientras que encontramos una reducción en los que se refiere a la obtención de notables con un 22,73% en el curso pasado y un 16,67% en el actual; si se dio un ligero incremento en lo que se refiere al número de aprobados, con un 4,55% en el curso anterior y un 5,56% en el actual.

Del mismo modo, observamos en lo que hace referencia al resto de puntuaciones, estas son las de suspenso y no presentado, ya que, mientras que se duplica la cantidad de alumnos que no ejercieron su derecho a examinarse con un 13,64% de no presentados en el curso pasado y un 33,33% en el actual; si disminuyó considerablemente la cantidad de suspensos con la nueva metodología, pasando de un 45,45% de alumnos que les quedo pendiente el año pasado a 27,78% en el actual. Estos datos son recogidos en el siguiente gráfico:

Gráfico 3. Notas finales



Fuente: Universidad de Cádiz, 2022.

## 5. Conclusiones

Como hemos podido comprobar a lo largo de estas páginas, los resultados obtenidos en ambos cursos no difieren mucho unos de otros, ya que hay cierta compensación entre ambos. Mientras que con uno se incrementa el porcentaje de las calificaciones de sobresaliente en lo que respecta al otro, con este otro disminuye la cantidad de notables, frente a este uno en el que abundaban.

Asimismo, mientras que con ambas metodologías se mantiene el número de matrículas de honor y aumenta ligeramente el de aprobados, con el procedimiento actual se multiplican los no presentados, pero se reducen considerablemente la cantidad de suspensos.

Ahora bien, los datos que más nos motivan con la aplicación de la metodología de enseñanza/aprendizaje teórico, práctico y metodológico en la formación en lingüística de corpus se resumen básicamente en dos cuestiones. Por un lado, comprobamos que, mientras ha aumentado el número de alumnos que superan la asignatura y con ello la evaluación, ha disminuido considerablemente los que no, y, por otro, observamos que tanto la experimentación investigadora como la puesta en práctica de conocimientos, la aplicación de recursos informáticos y la ejemplificación de diferentes casos de dicha tarea de evaluación optimizaron los resultados del trabajo final y consiguieron una mayor implicación del alumno, además de enriquecer la labor investigadora del profesorado. Todo ello fomentó la participación e impulsó la motivación del alumnado, ayudando de esta forma al trabajo autónomo.

No obstante, a nuestro juicio, este tipo de formación que aquí esbozamos no solo presenta beneficios, sino también algunas limitaciones, tanto a nivel general como específico, de manera que, entre los inconvenientes que podemos denominar “generales”, hallamos los que detallamos a continuación.

- Considerando las diferentes salidas profesionales que presenta la comercialización de productos lingüísticos, resulta llamativo que esta materia solo se imparta en el grado en “Lingüística y lenguas aplicadas” de la Universidad de Cádiz y no en el resto de universidades españolas donde se oferta la formación lingüística, como son: la titulación de “Lingüística” en la Universidad Autónoma de Barcelona,

en la de “Lingüística y lenguas aplicadas” en la Universidad Complutense de Madrid y en la de “Lingüística aplicada y ciencias del lenguaje” en la Universidad Nebrija.

- Asimismo, es curioso que en la única institución española donde se da la posibilidad de formarse en esta línea sea solo una materia optativa en la instrucción lingüística, que ni siquiera se oferte en los dobles grados en combinación con esta, ni tampoco en los distintos títulos filológicos de la Universidad de Cádiz, aunque sea como optativa.
- Igualmente, teniendo en cuenta el papel tan destacado que tiene la lingüística de corpus dentro de la investigación asociada con el lenguaje y las lenguas, esta formación tan específica solo es ofrecida como asignatura propia en la titulación en “Lingüística aplicada y ciencias del lenguaje” de la Universidad Nebrija y como parte de otra, según ya comentamos en la introducción (cf. apdo. 1), en el de “Lingüística y lenguas aplicadas” de la Universidad de Cádiz, pero no en la de “Lingüística” en la Universidad Autónoma de Barcelona, ni en la de “Lingüística y lenguas aplicadas” de la Universidad Complutense de Madrid.

De igual modo, entre las desventajas que podemos calificar como “específicas”, observamos las siguientes cuestiones.

- Teniendo en cuenta la productividad profesional que ofrecen las Industrias de la lengua, pensamos que esta materia no debería ser repartida de una forma tan confusa y dispersa entre distintas asignaturas tanto generales como específicas de este título en diferentes cursos y semestres. Estamos de acuerdo con que debe haber una asignatura introductoria antes de la impartición de conjunto de disciplinas aplicadas que la componen que sirva como hilo conductor y, a partir de ahí, el resto de materias específicas que la componen, dándole cohesión a todo lo trabajado en el grado.
- En el caso de la formación en la lingüística de corpus en esta asignatura compartida que es el caso que estamos comentando, creemos que esta no debería ser ofrecida únicamente dentro de una asignatura tan abarcadora, sino que se debería darse a un nivel más especializado.
- Viendo que el aprendizaje en lingüística de corpus se produce en un marco repartido, observamos que encima se hace con otra disciplina desconocida por el público general, de la que apenas se tiene conocimiento fuera del ámbito lingüístico y que está en pleno auge, como es la neología, lo que limita la formación de ambas a una instrucción del estudiante universitario solo en cuestiones básicas.
- A todo esto, debemos sumarle que ambas líneas de formación son distribuidas en el mismo semestre, que es, como ya adelantamos (cf. apdo. 2), el quinto, dentro del tercer curso del citado grado, por lo que solo disponemos de muy pocas horas de clase para ambas, lo que no nos deja mucho margen para profundizar en ambos campos temáticos.

## Referencias

- Centro de Normalización Lingüística de la Lengua de Signos Española (en línea). *CORLSE. Corpus de la Lengua de Signos Española*. <https://corpuslse.es/>.
- Centro de terminología de la lengua catalana (en línea). *CercaTerm. Base de datos del Centro de Terminología (TermCat)*. <https://www.termcat.cat/es/cercaterm>.
- Centro Virtual Cervantes (en línea). *Banco de neologismos del Centro Virtual Cervantes*. [http://cvc.cervantes.es/lengua/banco\\_neologismos/](http://cvc.cervantes.es/lengua/banco_neologismos/).
- Díaz Sánchez, A. (en línea). *Indexador de Corpus de Aprendices de Español*. [http://repositorios.fdi.ucm.es/corpus\\_aprendices\\_espa%3%b1ol/view/paginas/view\\_paginas.php?id=1](http://repositorios.fdi.ucm.es/corpus_aprendices_espa%3%b1ol/view/paginas/view_paginas.php?id=1).
- Fernández-Ordóñez, I. (en línea). *COSER. Corpus Oral y Sonoro del Español Rural*. <http://www.corpusrural.es/index.php>.
- Gallardo Pauls, B. (2009). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 5. Trastorno por déficit de atención y/o hiperactividad. Materiales y análisis pragmático*. Universitat-AVaLCC.
- Gallardo Paúls, B. y Moreno, V. (2005). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 2. Afasia no fluente. Materiales y análisis pragmático*. Universitat/ AVaLCC.
- Gallardo Paúls, B. y Sanmartín, J. (2005). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 1. Afasia fluente. Materiales para su estudio*. Universitat-AVaLCC.
- Gallardo-Paúls, B.; Moreno-Campos, V. y Pablo Manuel, M. R. (2011). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 7. Lesionados de Hemisferio Derecho. Materiales y análisis pragmático*. Universitat-AVaLCC.
- Garayzábal, E. (2005). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 3. Síndrome de Williams. Materiales y análisis pragmático*. Universitat-AVaLCC.
- Gobierno Vasco (en línea). *EuskalTerm. Banco terminológico público vasco*. <https://www.euskadi.eus/euskalterm/>.
- Hernández, C.; Serra, E. y Veyrat, M. (2007). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 4. Afasia: casos mixtos. Materiales y análisis lingüístico*, Universitat-AVaLCC.
- Instituto Cervantes (en línea). *CAES. Corpus de aprendices de español*. [https://www.cervantes.es/lengua\\_y\\_ensenanza/tecnologia\\_espanol/caes.htm](https://www.cervantes.es/lengua_y_ensenanza/tecnologia_espanol/caes.htm).
- Instituto Universitario de Lingüística Aplicada de la Universidad Pompeu Fabra (en línea). *Banco de datos del Observatorio de Neología*. <http://obneo.iula.upf.edu/bobneo>.
- Leech, G. (1992). Corpora and theories of linguistic performance. En J. Svartvik (ed.), *Directions in Linguistics: Proceedings of Nobel Symposium 82 (Stockholm, 4-8 August 1991)* (pp. 105-122), Mouton de Gruyter.
- Martín Butragueño, P. y Lastra, Y. (coords.) (en línea). *CSCM. Corpus sociolingüístico de la ciudad de México*. [https://lef.colmex.mx/corpus\\_sociolingustico.html](https://lef.colmex.mx/corpus_sociolingustico.html).
- OTAN (en línea). *TermOTAN. Base de datos terminológica oficial de la OTAN*. <https://nso.nato.int/natoterm/content/nato/pages/home.html?lg=fr>.
- Paredes Duarte, M. J. (Coord.) (2018). *Corpus de transtornos del Ineguaje*. Editorial UCA.
- Parlamento Europeo (en línea). *IATE. Interactive Terminology for Europe*. <https://iate.europa.eu/home>.
- Real Academia Española (en línea). *CDH. Corpus del Nuevo diccionario histórico del español*. <http://web.frl.es/CNDHE/view/inicioExterno.view;jsessionid=BDCA17DB5A991FD09667AA9E2BCA177C>.
- Real Academia Española (en línea). *CORDE. Corpus Diacrónico del Español*. <http://corpus.rae.es/cordenet.html>.
- Real Academia Española (en línea). *CORPES XXI. Corpus del Español del Siglo XXI*. <http://web.frl.es/CORPES/view/inicioExterno.view>.
- Real Academia Española (en línea). *CREA. Corpus de referencia del español actual*. <http://corpus.rae.es/creanet.html>.
- Rodríguez Muñoz, F. J. (2010). *Corpus PerLA. Percepción, Lenguaje y Afasia. Vol. 6. Síndrome de Asperger*. Universitat-AVaLCC.
- Universidad de Barcelona (en línea). *UBTerm. Base de datos terminológica de la Universidad de Barcelona*. <http://www.ub.edu/ubterm/>.
- Universidad de Cádiz (2022). Actas de la asignatura de “Industrias de la lengua”. *Web de la Universidad de Cádiz*.
- Vega Moreno, É. (2020a). Aportaciones metodológicas para la formación en neología. En J. J. Gázquez Linares, M. Molero Jurado, Á. Martos Martínez, A. B. Barragán Martín, M. Simón Márquez, M. Sisto, R. M. del Pino Salvador y B. M. Tortosa Martínez (Comps.), *Innovación Docente e Investigación en Arte y Humanidades. Avanzando en el proceso de enseñanza-aprendizaje* (pp. 1055-1065). Editorial Dykinson.
- Vega Moreno, É. (2020b). Herramientas digitales para el aprendizaje/enseñanza de la neología aplicada. En M. Camino Carrasco, B. Fernández de Castro y L. Jiménez Ríos (Coords.), *III Congreso Motiv-ARTE: Retos de la teledocencia en Humanidades (Cádiz, del 2 al 4 de diciembre de 2020)*. Universidad de Cádiz.
- Villayandre Llamazares, M. (2008). Lingüística con corpus (I). *Estudios humanísticos. Filología*, 30, 329-349. <http://revpubli.unileon.es/ojs/index.php/EEHFFilologia/article/view/2847>.