Giulia Nalesso

# La competencia léxica de estudiantes de ELE

## Un estudio sobre la disponibilidad y la riqueza léxica

Incipit

PADOVA UP

PADOVA UNIVERSITY PRESS

# INCIPIT

*Tesi*

# INCIPIT

*è una collana di tesi di dottorato in*
*Scienze linguistiche, filologiche e letterarie*

---

La collana Incipit accoglie due serie distinte: le *Tesi*, selezionate fra quelle discusse all'interno del Dottorato in Scienze Linguistiche, Filologiche e Letterarie dell'Università di Padova e/o sotto la supervisione di docenti del Dipartimento di Studi Linguistici e Letterari dell'Università di Padova (DiSLL); i *Colloqui*, gli atti dei convegni organizzati annualmente da allievi e allieve del Dottorato.

Per entrambe le serie la scelta delle pubblicazioni avviene mediante *peer review* e *double blind*.

La collana è finanziata dalla Commissione Ricerca Scientifica del DiSLL, con un contributo del Dipartimento di Scienze Storiche, Geografiche e dell'Antichità (DiSSGeA).

Prima edizione 2022, Padova University Press
Titolo originale *La competencia léxica de estudiantes de ELE. Un estudio sobre la disponibilidad y la riqueza léxica*


Università degli Studi di Padova
via 8 Febbraio 2, Padova

www.padovauniversitypress.it
Redazione Padova University Press
Progetto grafico Padova University Press

This book has been peer reviewed

ISBN 978-88-6938-331-1

Giulia Nalesso

# La competencia léxica de estudiantes de ELE.

*Un estudio sobre la disponibilidad y la riqueza léxica*

PADOVA UP

**Índice**

**Siglas y acrónimos**

ASALE: Asociación de Academias de la Lengua Española
CAES: Corpus de aprendices de español
CI: Centro de interés
CREA: Corpus de referencia del español actual
DL: Disponibilidad léxica
DLE: Diccionario de la lengua española
ELE: Español lengua extranjera
IAT: Intervalo de aparición de palabras nocionales
IC: Índice de cohesión
ID: Índice de disponibilidad léxica
LM: Lengua materna
L1: Primera lengua
L2: Segunda lengua
L3: Tercera lengua
LE: Lengua extranjera
MCER: Marco común europeo de referencia para las lenguas
PCIC: Plan curricular del Instituto Cervantes
PPHDL: Proyecto Panhispánico de Disponibilidad Léxica
RAE: Real Academia Española
RL: Riqueza léxica
TTR: *Type/Token Ratio*

## Introducción

Este trabajo, que procede fundamentalmente de nuestra tesis doctoral (Nalesso 2019a), se enmarca en el ámbito de los estudios sobre el aprendizaje del léxico en español lengua extranjera y tiene como objetivos la medición y la evaluación de la competencia léxica de un grupo de aprendientes italófonos de nivel universitario mediante la aplicación de dos metodologías complementarias que, por primera vez, se emplean en un único estudio. La primera consiste en el análisis de la disponibilidad léxica, la cual se refiere al cómputo de palabras que un sujeto es capaz de actualizar ante un estímulo temático dado; la segunda se denomina análisis de la riqueza léxica y mide la variedad y la densidad del vocabulario utilizado en un texto, oral o escrito. La correlación entre estos dos métodos posibilita un conocimiento profundo de la competencia léxica, ya que la perspectiva de la disponibilidad se coloca en el plano paradigmático del léxico y analiza no solo de qué palabras dispone un aprendiente, sino cómo se organizan conceptualmente en su lexicón mental (léxico potencial). Por su parte, la riqueza se ubica en el plano sintagmático del léxico y observa cuántas palabras en total y distintas se emplean en un educto (léxico utilizado).

Conscientes de la importancia que ha cobrado el componente léxico en la didáctica de lenguas extranjeras, elegimos este tema debido al interés que despierta la competencia léxica, entendida como el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo (MCER 2002: 108). En este sentido, nuestra principal aportación es el hecho de abordar el asunto de una manera novedosa que conjuga los dos métodos encaminados a analizar el vocabulario que el aprendiz sabe y utiliza en una situación comunicativa. Reunimos bajo un mismo marco metodológico

procedimientos distintos para lograr una mejor comprensión de un fenómeno tan complejo como la competencia léxica, proporcionando un instrumento de análisis empírico innovador para la comunidad científica. Asimismo, realizamos la recogida de datos en dos momentos distintos para la misma muestra, lo que nos permitió analizar la evolución del vocabulario de los informantes tras la asistencia a un curso de español de un año. Este tipo de análisis, del que no tenemos constancia que haya sido realizado hasta ahora, pretende obtener las claves necesarias para observar el desarrollo temporal en la adquisición del léxico y, por tanto, la efectividad de la acción didáctica.

En definitiva, la originalidad de la investigación consiste en conciliar la disponibilidad y la riqueza léxica en un estudio de tipo transversal y longitudinal[1] que sea lo más aglutinador posible y permita avalar una serie de hipótesis: en el plano didáctico, creemos que la competencia léxica aumentará a medida que avanza el nivel de español, a saber, los resultados de los estudiantes más avanzados y de la encuesta suministrada a final del año académico presentarán índices mejores, tanto en la activación de léxico disponible como en la producción escrita. Existiría, por tanto, una relación asociativa entre el dominio de ELE y la cantidad de lexías aprendidas (activadas y utilizadas en las pruebas). Es más, considerado el nivel lingüístico adquirido, los indicadores de la riqueza expresarán una buena variación de vocabulario gracias al desarrollo de una adecuada competencia léxica que se repercute positivamente en la competencia comunicativa de los estudiantes. Asimismo, el análisis comparativo demostrará que el léxico disponible de nuestros informantes es mayor que el de otros aprendientes de distinta procedencia y que poseen mejores capacidades expresivas gracias a la afinidad tipológica entre italiano y español, aunque existirá un desfase entre las listas de los sujetos que estudiaron en contexto de inmersión (deberían poseer un conocimiento superior).

De entre los factores sociolingüísticos analizados, el sexo no surtirá ningún efecto en la producción de los informantes, ni cuantitativa ni cualitativamente. El nivel de ELE influirá positivamente en la competencia del grupo: es esperable que disponibilidad y riqueza léxica de los estudiantes de nivel más avanzado sean mayores con respecto a los que tienen

[1] No conocemos ningún estudio de carácter longitudinal sobre disponibilidad léxica en ELE que tenga en cuenta el desarrollo de la competencia de los mismos informantes además de nuestros análisis (Nalesso 2018a, 2019a). Laufer (1991) examinó la riqueza léxica de 47 aprendientes de inglés L2 en diferentes etapas de su carrera universitaria, averiguando que es posible detectar una evolución del vocabulario en intervalos de 14 y 28 semanas.

un menor dominio lingüístico. Según el conocimiento de otras lenguas extranjeras los participantes que conocen más idiomas conseguirán mejores resultados según la *Hipótesis de la Interdependencia Lingüística* de Cummins (1979), aunque se detectará la influencia de dichas lenguas, en particular del inglés al ser el primer idioma extranjero estudiado en este contexto académico.

La edición de los datos revelará los siguientes aspectos: frecuentes interferencias procedentes del italiano; errores de ortografía; palabras de la sintopía hispanoamericana y del sociolecto coloquial porque la programación curricular prevé la enseñanza de los rasgos peculiares de dichas variedades diatópicas y diafásicas.

En lo que atañe al análisis del léxico disponible, teniendo en cuenta las investigaciones ya publicadas, los centros de interés producirán un mayor número de palabras y de vocablos en función de su rentabilidad para el alumnado desde el momento en el que algunos de estos estímulos se revelan poco apropiados para cumplir con las necesidades de un aprendiente del siglo XXI. Los sustantivos serán la clase gramatical predominante y en menor medida aparecerán otras categorías morfosintácticas. Igualmente, aparecerá una escasa cantidad de unidades multipalabra.

En sintonía con estos planteamientos, el trabajo se estructura en cuatro capítulos repartidos en dos partes. La primera sirve de marco para la investigación, está dedicada a los cimientos y a los patrones teóricos y metodológicos que la rigen. El capítulo 1 expone el estado de la cuestión en lo que se refiere a los orígenes de la disponibilidad y la riqueza léxica en la segunda mitad del siglo XX en Francia y al desarrollo actual de los estudios en el mundo hispano y su aplicación en el ámbito de ELE. El capítulo 2 se centra en los aspectos metodológicos: describe la selección de la muestra y de los índices de cálculo; las técnicas de elicitación de los datos y la encuesta que diseñamos para la medición de la competencia léxica; los criterios de edición y el tratamiento –informático y estadístico– del material recogido.

La segunda parte constituye el bloque central del trabajo donde se presentan los análisis que nos permitieron alcanzar los objetivos y corroborar o refutar las hipótesis iniciales. El apartado analítico inicia con el capítulo 3 que desarrolla, en primera instancia, el estudio cuantitativo del léxico disponible repartido en los análisis transversal, longitudinal y comparativo. Como todos los trabajos de disponibilidad léxica, se empieza por los datos generales extraídos de la primera suministración de la prueba (número total de palabras y de vocablos, media de palabras por

sujeto y por centro de interés, índice de cohesión, densidad léxica) que, a continuación, se desglosan en función de las variables consideradas (sexo, nivel de ELE, conocimiento de otras LE). Los resultados por variable siguen el mismo protocolo y se complementan por el estudio de los datos estadísticos descriptivos de cada condicionante para obtener información sobre el comportamiento intragrupal e intergrupal de los encuestados. El análisis longitudinal, llevado a cabo gracias a la segunda aplicación de la encuesta, permitió percatar la evolución en el tiempo de la competencia léxica y deducir en qué medida los procesos didácticos inciden en la adquisición del vocabulario. Por último, el análisis comparativo coteja nuestros resultados con los de otras investigaciones realizadas con aprendientes de ELE de idéntica competencia, pero de distinta procedencia y lengua materna, con el objetivo de indagar puntos comunes o diferencias en su bagaje léxico. En segunda instancia, este capítulo aborda el estudio cualitativo de las listas de léxico disponible centrándose en el tipo de vocabulario activado a partir del análisis de los vocablos más disponibles para detectar las temáticas predominantes en cada centro de interés y las categorías gramaticales más difundidas en las respuestas. Para ello, cuantificamos también la cardinalidad del conjunto con el propósito de medir la compatibilidad existente entre las variantes de una misma variable. Por último, comparamos el material obtenido con el *Corpus de Referencia del Español Actual* y el *Corpus de aprendices de español* para comprobar si el léxico disponible de los participantes está incluido en estos bancos de datos elegidos como corpus de referencia.

El estudio de la riqueza prosigue el bloque analítico para completar la información adquirida mediante el análisis de textos escritos por los informantes. En el capítulo 4 examinamos, en primer lugar, los indicadores de la variedad y de la densidad del vocabulario (*Type/Token Ratio*, variación léxica, índice de hápax, densidad léxica, intervalo de aparición de palabras nocionales) siguiendo el mismo orden de los capítulos anteriores, empezando por el análisis transversal. Seguidamente, el análisis longitudinal –realizado por primera vez en este marco metodológico– permitió observar el desarrollo de la competencia en relación con el tiempo de aprendizaje lingüístico. Sigue el análisis comparativo con otras investigaciones centradas en la producción escrita de estudiantes de ELE, cuya finalidad es corroborar la hipótesis de que discentes no nativos de distinta procedencia que comparten el dominio lingüístico poseen iguales capacidades expresivas. El capítulo termina con la propuesta de un estudio cualitativo de la riqueza léxica: siendo conscientes de los límites que con-

lleva un trabajo preliminar, tenemos la pretensión de proponer una nueva línea de encuesta para futuras investigaciones. Planteamos un análisis tipológico de las unidades léxicas del conjunto textual que vaya más allá de las cifras, aun a sabiendas de que los índices cuantitativos aportan por sí mismos un tipo de información que no se limita a los datos numéricos, como es la «calidad de la escritura» (López Morales 2011). Para cumplir con este cometido, estudiamos la lista de los vocablos más frecuentes del corpus a fin de determinar morfológicamente las lexías más utilizadas, el área semántica a la que pertenecen y el grado de asociación con el tema propuesto para la redacción del texto.

Finalmente, las últimas páginas presentan las conclusiones de los resultados de la investigación.

Esperamos que el trabajo cumpla los objetivos fijados y, por tanto, pueda abrir nuevos caminos relacionados con la competencia léxica de los aprendientes italianos de español y, en general, de los aprendices de ELE cualquiera que sea su procedencia y que sirva también para la determinación de un nuevo protocolo experimental aplicable a diferentes entornos investigativos y educativos.

# Capítulo 1.
# Orígenes y desarrollo actual de la disponibilidad y de la riqueza léxica

En este primer capítulo reseñamos el origen y la difusión de los estudios de disponibilidad y riqueza léxica, trazamos la evolución de estas dos metodologías en una perspectiva historiográfica hasta llegar a su aplicación en el ámbito de las investigaciones en ELE propia del marco teórico en el que se sitúa el presente trabajo.

## 1.1. La disponibilidad léxica

La disponibilidad léxica cuenta con una amplia trayectoria de estudios desde su nacimiento en Francia a mediados de los años cincuenta del siglo pasado cuando un grupo de lingüistas realizó una revisión metodológica encaminada a perfeccionar la selección del vocabulario para la enseñanza del francés L2/LE que solucionase las limitaciones de un proceso basado exclusivamente en el criterio de frecuencia.

Hasta la fecha se utilizaban los diccionarios de frecuencia (Käding 1897, Thorndike 1921, Henmon 1924, Buchanan 1927, Bakonyi 1933, Van der Beke 1935, West 1935, Rodríguez Bou 1952, García Hoz 1953, Juilland y Chang-Rodríguez 1964) para la elaboración de materiales didácticos finalizados al aprendizaje del léxico de un idioma, tanto L1 como L2/LE. Consistían en listas de palabras estadísticamente frecuentes en un corpus textual representativo, las cuales permitían la recopilación de listados para al aprendizaje memorístico del vocabulario (en la época no se consideraba el conocimiento del léxico como factor de éxito en el proceso

de aprendizaje lingüístico por lo que no se elaboraban estrategias más eficaces). En poco tiempo se puso de relieve que el mero criterio de frecuencia no era proficuo para el aprendiente extranjero ya que los ítems propuestos aparecían descontextualizados y, a menudo, se trataba de palabras funcionales del discurso, de poco, o nulo, contenido semántico, lo que no permitía el desarrollo de un conocimiento "natural" de la lengua objeto, próximo al de un nativo (Gougenheim *et al.* 1964: 31-35).

Tanto la didáctica como la lingüística aplicada criticaron estos recursos porque consistían en el recuento del léxico de una comunidad solamente en función de la mayor o menor aparición de las palabras, método que, en realidad, no era eficaz porque se había detectado que lexías que presentaban índices de frecuencia bajos o muy bajos (que, por tanto, no se incluían en tales diccionarios) eran, por el contrario, usadas comúnmente por los nativos, por lo que basarse solo en este criterio no resultaba tan provechoso como para ofrecer al aprendiente el caudal y el tipo de léxico necesario para un adecuado aprendizaje. De ahí que se plantease una nueva metodología, la disponibilidad léxica, que muy pronto se aplicó a la enseñanza de otras lenguas extranjeras. Es en el ámbito hispánico donde tuvo su mayor desarrollo, donde se apreció la mayor difusión de los trabajos y donde se abrió un amplio abanico de posibilidades de aplicación en distintos campos.

### 1.1.1 El estudio pionero

El estudio pionero sobre disponibilidad léxica fue realizado en Francia por un grupo de lingüistas nombrado por el *Ministère de l'Éducation Nationale* francés que vio la luz en su forma definitiva en 1964 bajo el nombre de *Français Fondamental* y sentó las bases para esta nueva metodología. Aunque ya en un momento anterior había empezado a plantearse el concepto de *léxico disponible*, el cual abarca las palabras que se activan en una situación comunicativa concreta y se compone, en la mayoría de los casos, de unidades léxicas de contenido nocional (Michéa 1953).

La disponibilidad léxica se elaboró a partir de la distinción entre *palabras temáticas* (lexías concretas relacionadas con el tema de la comunicación, que pueden ser más o menos frecuentes, es decir palabras inestables) y *palabras atemáticas* (lexemas gramaticales que aparecen en cualquier texto independientemente del tema comunicativo, dotadas de un alto índice de frecuencia, es decir palabras estables):

> En una situación determinada, las primeras palabras que vienen a la mente son aquellas que están relacionadas con tal situación y que la

> caracterizan [...]. Una palabra disponible es una palabra que, sin ser particularmente frecuente, siempre está lista para ser utilizada, y acude inmediatamente a la mente cuando parece necesario (Michéa 1953: 342).

Sin embargo, las palabras temáticamente disponibles para cualquier hablante (por ejemplo, *tenedor*) pueden ser muy infrecuentes en un corpus de textos escritos, incluso de grandes proporciones, lo cual se imponía como una de las limitaciones del criterio de frecuencia en la selección del vocabulario destinada al aprendizaje. López Morales (1999: 11) afirma:

> Existe en el lexicón mental una serie de términos que no se actualizan a menos que sea necesario para comunicar una información muy específica. Se trata de un léxico 'disponible', cuyo estudio no puede emprenderse manejando frecuencias, porque este factor es pertinente solo en el caso de las actualizaciones léxicas efectivas, no de las potenciales.

Este proyecto nació a partir del presupuesto de que el vocabulario de un idioma cualquiera es inagotable para los hablantes nativos y aún más para los extranjeros. Por ello, se elaboró una lista de palabras que contenía los elementos esenciales para permitir un rápido acceso al léxico de la lengua objeto, el léxico fundamental del francés: se completaron los datos obtenidos de la frecuencia (*léxico frecuente*) contextualizando las palabras que aparecían en mencionada lista (*léxico disponible*). Era indispensable llevar a cabo una selección ulterior de los contenidos didácticos previamente propuestos porque se había puesto de relieve que algunas palabras, aun siendo perfectamente conocidas por los usuarios nativos de una lengua, solo se empleaban en determinadas situaciones comunicativas y, por esta razón, no constituían el léxico frecuente de una lengua, por lo que no podía ser la única vía para la adquisición del vocabulario. De esta forma, se seleccionaron las unidades léxicas más apropiadas para la enseñanza en los niveles iniciales de aprendizaje del francés.

Para la recogida de datos se organizaron pruebas escritas de tipo asociativo suministradas a informantes[2] que recopilaban listas de palabras relacionadas con ciertos estímulos temáticos, definidos centros de interés (Michéa 1950: 189). Según López Morales (1999: 32) se trata de «[...] las únicas [pruebas] que hacen posible en condiciones experimentales, que

[2] La primera prueba de disponibilidad léxica se suministró a un grupo de escolares franceses, elegidos porque parece que en aquella edad los niños ya tienen un vocabulario bien desarrollado, pero libre de todo tipo de especialización sectorial (Gougenheim *et al.* 1964). Hernández Muñoz (2015) corrobora esta idea al afirmar que la formación previa al acceso al mundo laboral es uno de los factores ligados a la experiencia que modifican el conocimiento léxico de los hablantes de una lengua.

se realicen en la actuación lingüística las unidades léxicas con poca estabilidad estadística». La edición de los datos constituyó la etapa sucesiva que se consiguió sumando las palabras más frecuentes a las primeras palabras disponibles procedentes de los análisis.

El *Français Fondamental* supuso un punto de inflexión en el marco de la léxico-estadística y un proceso de reajuste metodológico que culminó en la planificación de un nuevo sistema de recogida de datos para la elaboración del léxico disponible de una comunidad lingüística. El valor de la disponibilidad léxica y la elaboración de los vocabularios fundamentales fue ampliamente reconocido y pronto dio frutos en forma de investigaciones derivadas. Disponemos de un gran número de estas últimas, dedicadas al francés y a otros idiomas que se consideran pioneras desde el momento en el que el método empírico estaba todavía configurándose (Dimitrijevic 1969, Mackey 1971, Njock 1979, Pfeffer 1964).

A estas alturas cabe señalar que los centros de interés (CI) son los elementos nucleares de los trabajos sobre disponibilidad léxica ya que permiten recoger el léxico disponible del grupo encuestado. El establecimiento de estos estímulos temáticos es una de las novedades más destacables del *Français Fondamental* y más importantes de cada investigación realizada en este campo al configurarse con el punto inicial para el diseño de la prueba.

Con el objetivo de contextualizar el vocabulario y facilitar su aprendizaje, los investigadores se preguntaron cuáles fuesen las áreas semánticas más adecuadas para activar los lexemas que no aparecían en los diccionarios de frecuencia. Propusieron dieciséis temas que, en su opinión, reunían el vocabulario universal de una lengua (Gougenheim *et al.* 1964: 152-153):[3]

1. Las partes del cuerpo,
2. La ropa (hombre y mujer),
3. La casa (sin muebles),
4. Los muebles de la casa,
5. Los alimentos y las bebidas,
6. Los objetos colocados en la mesa para la comida,
7. La cocina, sus muebles y utensilios,
8. La escuela, sus muebles y material escolar,
9. La calefacción y la iluminación,
10. La ciudad,
11. El pueblo o el burgo,

[3] La traducción es nuestra.

12. Los medios de transporte,
13. Los trabajos del campo y del jardín,
14. Los animales,
15. Los juegos y las distracciones,
16. Las diferentes profesiones.

Estos centros se utilizan todavía en los experimentos de disponibilidad léxica, pese a que con frecuencia se haya cuestionado, en primera instancia, su supuesta universalidad temática (ya en 1999 López Morales se interrogaba sobre si fuera posible circunscribir todos los intereses humanos en esos dieciséis campos nocionales y cuántos de ellos debería tener en cuenta un estudio que pretende ser exhaustivo) y, en segunda instancia, su dependencia tanto de la naturaleza de cada proyecto y de las necesidades del investigador (Rubio Lastra 2018), como de los cambios socioculturales y los intereses de los informantes. Además, otra de las cuestiones denunciadas es la limitación de los datos recogidos a la clase de los sustantivos y a la unidad palabra, entendida como unidad monoverbal.

Efectivamente, en el ámbito de la investigación en ELE, algunos de los dieciséis estímulos originales fueron objeto de crítica por la falta de interés y de significatividad para los participantes (Samper Padilla, Bellón y Samper Hernández 2003) o, en nuestro ámbito de acción, por no abarcar los temas más comunes utilizados por un aprendiente no nativo, como las acciones de la vida cotidiana y las relaciones familiares (Izquierdo Gil 2005). Así, por ejemplo, González Fernández (2014: 52) asevera:

> [...] se pone de manifiesto como algunos centros de interés clásicos no son idóneos para los trabajos de disponibilidad léxica aplicada a la enseñanza del español como lengua extranjera. En las investigaciones de esta corriente se debería eliminar algunos centros de interés y modificar otros para poder obtener resultados más relevantes.

La consecuencia que se deriva es que los centros están sujetos a arbitrariedad, por lo que es posible reformularlos, reducirlos o sustituirlos conforme a los objetivos propuestos en cada trabajo (Juilland 1970, Mackey 1971, Benítez Pérez 1994, López Morales 1999, Bombarelli 2005, Moreno Fernández 2012, González Fernández 2014, Paredes García 2014, Tomé Cornejo 2015, Sánchez-Saus 2016, Hidalgo 2019). Asimismo, en la actualidad, como producen cantidades mayores o menores de vocablos en función de la rentabilidad que tienen para los informantes, algunos CI se revelan poco adecuados para cumplir con las necesidades y los intereses culturales y comunicativos del siglo XXI propios de la sociedad digitalizada y fuertemente intercultural en la que vivimos. Esto justificaría la

necesidad de tomar nuevas decisiones que solventen la inadecuación de ciertos centros, establecidos como *input* para la elicitación hace más de setenta años. Se trataría de eliminar los pocos productivos, ampliar o unir otros, incluir nuevos que impulsen la actualización de otras categorías morfosintácticas y de unidades complejas.

### 1.1.2 La disponibilidad léxica en el mundo hispánico

Como ya mencionado, la investigación sobre la lengua española es la que más ha contribuido al desarrollo de la disponibilidad léxica, sobre todo a partir de los años setenta del siglo pasado, cuando se propusieron numerosos trabajos en el ámbito del español lengua materna. López Morales, el promotor de la aplicación del método a los estudios hispánicos, fue quien primero analizó, siguiendo la propuesta del *Français Fondamental*, el léxico disponible de hispanohablantes. Planteó un proyecto de gran extensión dedicado a los habitantes de Centroamérica y del Caribe que culminó en la publicación de *Léxico disponible de Puerto Rico* (1999) con la finalidad de determinar el vocabulario disponible en toda la comunidad puertorriqueña. Con ello, estimuló el interés de otros investigadores que empezaron a administrar la prueba de disponibilidad léxica a muchas de las comunidades de habla española en todo el mundo hasta llegar a la creación del *Proyecto Panhispánico de Disponibilidad Léxica* (PPHDL) coordinado por el propio López Morales, cuyo objetivo último es la recopilación de un diccionario del léxico disponible referido a la totalidad de los hispanohablantes nativos y que permite cotejar las diferencias lingüísticas y culturales entre distintas áreas.

Para cumplir con este fin se requería la aplicación de un *modus operandi* compartido, sustentado en un mismo conjunto de normas: en 1999 se organizó una reunión en Bilbao para establecer una metodología común de recogida y tratamiento de los datos que garantizase la comparabilidad y la fiabilidad de cada estudio.[4]

Las primeras fórmulas empleadas no tenían en consideración el orden de aparición de las palabras en los listados aportados por los informantes, por lo que se desarrolló una fórmula que calculase el índice de disponibilidad léxica en función de un nuevo parámetro (López Chávez y Strassburger Frías 1987):[5]

[4] Cfr. Samper Padilla (1998), Samper Padilla, Bellón Fernández y Samper Hernández (2003), Samper Padilla y Samper Hernández (2006).

[5] Donde: D(Pj)= disponibilidad de la palabra j; n= máxima posición alcanzada en el CI en

$$D(Pj) = \sum_{i=1}^{n} e^{-23\left(\frac{i-1}{n-1}\right)} \times \frac{fji}{I_1}$$

Figura 1. Fórmula de la disponibilidad léxica (1987).

La mejora aportada consistía en cuantificar la frecuencia y la posición en la que se actualizaban las palabras para poder detectar aquellas que acudían primero a la mente ante un cierto estímulo (grado de espontaneidad). Todo esto llevó también al desarrollo de un *software* para el tratamiento informático de los datos, el *Dispolex* (Bartol Hernández *et al.*).

De ahí en adelante los experimentos aumentaron notablemente difundiéndose en todas las áreas hispanohablantes del mundo y contribuyendo al avance de las investigaciones (análisis cuantitativos, estadísticos, cualitativos) y de los datos recogidos. A este propósito, parece obligado citar al grupo chileno encabezado por Echeverría que elaboró en 1987 una nueva fórmula finalizada a la medición del *índice de cohesión* (IC) para profundizar la perspectiva cualitativa de los estudios, junto al programa *DispoGrafo* (Echeverría *et al.* 2008) que permite el análisis de las redes semánticas que organizan el lexicón mental, conjugando las labores de la disponibilidad léxica con la psicolingüística. La gran evolución de esta metodología permitió, asimismo, notables innovaciones teóricas y prácticas a partir de las múltiples aplicaciones en distintas disciplinas como sociolingüística, dialectología y etnolingüística, además de estudios de lingüística, psicolingüística y didáctica de lenguas (véase §1.1.4).

### 1.1.3 La disponibilidad léxica en ELE

En lo que se refiere a la didáctica de lenguas, desde los últimos años del siglo XX se propusieron trabajos centrados en el léxico disponible de aprendientes no nativos de español. Apreciamos una cantidad notable de proyectos que, inicialmente, adoptaron las directrices del PPHDL, pero que fueron ajustando algunas pautas a las exigencias específicas propias

la encuesta; i= número de posición de que se trata; j= índice de la palabra tratada; e= número natura *e* (2,718281828459045); fji= frecuencia absoluta de la palabra j en la posición i; I1= número de informantes que participan en la encuesta. Se consideran los siguientes elementos: la frecuencia absoluta de cada palabra en cada posición de la lista; la frecuencia absoluta de cada palabra resultante de la suma de las diferentes frecuencias alcanzadas en cada posición; el número de encuestados; el número de posiciones alcanzadas en los centros de interés analizados; la posición en que aparece cada palabra.

de la investigación y, en particular, a las de sus informantes que, por ser discentes extranjeros, necesitaban un tratamiento diferente, lo que ha llevado a una renovación metodológica.

El iniciador de esta nueva y muy fructífera corriente fue Carcedo, quien trabajó con estudiantes finlandeses. Planteó numerosos estudios en los que analizaba: el proceso de aprendizaje mediante el análisis del léxico disponible tras la asistencia a cursos de ELE (1998); el desarrollo de la competencia léxica en distintas etapas de la adquisición lingüística (1999a); la interlengua y los errores más comunes (1999b); el cotejo de sus resultados con los de hablantes nativos (1999c, 2000a) y, por último, la cuestión del componente cultural del vocabulario en la evolución del conocimiento del español (2000b). Su proyecto más influyente y extenso (2000c) trata una muestra de 350 aprendientes de diferentes niveles, cuya finalidad es ofrecer una imagen exhaustiva del léxico disponible del estudiantado finlandés. Para ello, utilizó en su prueba los dieciséis centros de interés del *Proyecto Panhispánico*, pero aportó algunos cambios en las variables extralingüísticas: además de los factores clásicos (sexo, edad, nivel sociocultural, zona geográfica, centro de estudios) introdujo algunos nuevos como la lengua materna, el nivel de estudios, el conocimiento de otras lenguas románicas y el libro de texto utilizado. Esto le permitió corroborar que el léxico disponible de los alumnos no era conforme a sus reales exigencias comunicativas en español y, por ende, puso de manifiesto que estos análisis y sus resultados debían tenerse en cuenta a la hora de diseñar los cursos y crear materiales adecuados.

La tarea de Carcedo impulsó proyectos dedicados a aprendientes de diferentes nacionalidades que estudiaban español en su país de origen o en un contexto de inmersión lingüística. Traemos a colación, a este propósito, la aportación de Samper Hernández (2002), quien publicó un trabajo realizado con 45 estudiantes de lenguas maternas diferentes procedentes de los *Cursos Internacionales de español* de la Universidad de Salamanca. La autora presentó nuevas modificaciones en lo referente a las variables sociolingüísticas: utilizó el sexo, la lengua materna, el conocimiento de otras lenguas (no simplemente románicas), el nivel de español. En segunda instancia, tomando conciencia de la falta de criterios específicos para la edición de las respuestas de los "nuevos" informantes, planteó algunos cambios significativos con respecto a las pautas del PPHDL (Samper Hernández 2002: 16). En particular, defiende tanto una mayor laxitud en la corrección ortográfica y en el tratamiento de los extranjerismos como una mayor amplitud en las relaciones asociativas de las palabras apor-

tadas en los listados. Estos trabajos sentaron las bases para el desarrollo de los estudios dedicados a discentes no nativos de español, que se consideran los pilares en los que se fundamenta la mayoría de los proyectos posteriores, los cuales introdujeron a su vez cambios y mejoras.[6]

En este sentido, cabe mencionar a Šifrar Kalan (2009, 2012, 2014, 2018) que se dedicó a aprendientes eslovenos estableciendo como objetivo privilegiado de su investigación la aplicación pedagógica de los resultados a la planificación léxica y extendió el campo investigativo hacia la sociolingüística proponiendo un análisis sobre la influencia que puede ejercer un periodo de estancia *Erasmus* en España sobre el léxico disponible, en particular con referencia a los estereotipos y a la cultura española.

El trabajo de Sánchez-Saus (2016) representa también una novedad en este panorama, que recordamos precisamente por la magnitud de las innovaciones: contribuyó al progreso de la metodología realizando un pormenorizado estudio estadístico del léxico disponible de su muestra compuesta por 322 sujetos a partir de nuevos CI. Si bien los dos trabajos de referencia para la disponibilidad léxica en ELE emplearon los dieciséis estímulos del PPHDL, la autora optó por cambiar algunos porque le parecían anticuados y de poco interés para informantes extranjeros que no dominan plenamente el español. De ahí que modificase los temas clásicos adaptándolos a la lista de contenidos que el MCER recomienda para el nivel, aun manteniendo las semejanzas con los tradicionales para permitir el cotejo con otros trabajos. Es más, propuso un análisis cualitativo-semántico de las respuestas mediante el uso del programa *Dispografo* gracias al cual pudo describir las unidades léxica en función del centro de interés al que pertenecen, relacionarlas con su núcleo temático, etiquetarlas según la categoría gramatical, examinando fenómenos diatópicos y diafásicos como los cambios semánticos, la importación (extranjerismos, préstamos, calcos) y la generación léxica (derivación, composición, acortamientos). Igualmente, destacamos la investigación de Hidalgo (2019) en la que se presenta otra innovación, ya que además de estudiar el léxico disponible de los 440 aprendices sinohablantes encuestados, realizó un análisis del manual de español más usado en China a fin de determinar su posible incidencia en el vocabulario de sus usuarios directos.

A estas alturas, resulta obligado comentar la situación de los estudios dedicados a estudiantes italianos ya que a pesar de la extensión geográfica de los trabajos de disponibilidad léxica en ELE, no se le había dedicado

[6] Véase a Hidalgo (2017) y Aabidi (2019) para una revisión más detallada de los estudios publicados sobre léxico disponible en ELE.

en Italia la misma atención. Contábamos con pocas contribuciones, teniendo en cuenta las altísimas cifras de aprendientes de español en este país, donde esta lengua es hoy la segunda más estudiada después del inglés, a todos los efectos y en todos los niveles, tanto en la enseñanza no reglada como en contextos institucionales: según el informe *El español: una lengua viva* del Instituto Cervantes, en 2021, los usuarios potenciales de español en Italia –no nativos– eran 897.624. Además del estudio de Caggiula (2013), de modestas proporciones, en el que participaron 50 informantes, disponemos de los trabajos de Rubio Sánchez (2015, 2017) sobre aprendientes preuniversitarios, cuya originalidad reside en la inclusión de cuatro centros de interés en italiano. Paolini (2017) en su trabajo de fin de máster estudió cuantitativa y cualitativamente la disponibilidad léxica de un grupo de estudiantes universitarios. Destacamos, además, el proyecto de Del Barrio y Mae Vann que se ultimó con la publicación de un diccionario de léxico disponible de aprendices italófonos en 2018.[7] Incluimos en esta breve reseña a Nalesso (2018a) donde realizamos un estudio piloto con una muestra restringida a 20 informantes con el objetivo de validar la metodología empírica que establecimos para este experimento. Asimismo, en Nalesso (2018b, 2022) aplicamos la disponibilidad léxica desde la perspectiva de la didáctica con el eje de averiguar la eficacia de algunas actividades explícitas para el aprendizaje del vocabulario analizando el campo nocional "alimentos y bebidas" y la situación ortográfica de los discentes. Tratamos, por último, el tema del léxico disponible desde un nuevo punto de vista estudiando la disponibilidad léxica terminológica (Nalesso 2020).

### 1.1.4 Otras aplicaciones de la disponibilidad léxica

La disponibilidad léxica no se limita a los estudios de lingüística y didáctica de lenguas, sino que hay otras disciplinas que aprovechan de los datos que cada investigación proporciona.

Una de las aplicaciones más relevantes se halla en el ámbito de la sociolingüística que estudia el influjo de las variables sociales en el léxico disponible, cuyo objeto es examinar la variación lingüística en función de los condicionantes extralingüísticos considerados en los análisis. En este marco, López Morales (1973) fue también iniciador en analizar los datos según el nivel socioeconómico de los participantes. Nos parece oportuno mencionar, asimismo, nuevas propuestas desde perspectivas inéditas: re-

[7] Tras algunos estudios preliminares de Del Barrio (2016, 2017a, 2017b, 2018).

cientemente se ha abordado, por ejemplo, un análisis socionomástico de los antropónimos (Fernández Juncal y Hernández Muñoz 2019).

Resulta de gran interés, además, la información manejada por la dialectología para realizar comparaciones interdialectales y estudios de regionalismos. Mackey (1971) abordó de forma pionera el estudio de la variedad diatópica para comprobar similitudes y diferencias entre los sujetos que hablan la misma lengua en diferentes partes del mundo, en su caso se trató del francés de Canadá y de Francia. En esta línea, se realizaron cotejos entre distintas variedades del español desde los trabajos pioneros de López Chávez (1992, 1995) que abarcaron Madrid, Gran Canaria, República Dominicana y Puerto Rico.

Las investigaciones de disponibilidad léxica también proporcionan información para la psicolingüística. Se realizan análisis de corte cognitivo debido a la naturaleza de las pruebas, de tipo asociativo, ya que la producción del léxico disponible es una tarea cognitiva compleja en la que intervienen diversos elementos psicológicos, más allá de los puramente lingüísticos. En este sentido, Hernández Muñoz (2005) describe esta metodología como «una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología». Se estudian en este marco: la organización del lexicón mental, el acceso y la selección del léxico en L1 y L2/LE, las estrategias de recuperación del léxico, las redes semánticas y, por último y sin ánimo de agotar todas las posibilidades, las asociaciones mentales activadas durante las pruebas (entre otros, Hernández Muñoz y López García 2014, Hernández Muñoz, Izura y Tomé Cornejo 2014, Del Barrio 2018, Tomé Cornejo 2018, Gómez Devís 2019).

Por su parte, la etnolingüística se centra en las palabras actualizadas en las encuestas que reflejan la cultura y las costumbres de una comunidad. De nuevo, Mackey (1971) fue pionero en tratar esta cuestión abordando las diferencias culturales entre Canadá y Francia. En el ámbito de ELE sobresale Carcedo (2000a, 2000b) quien llevó a cabo un análisis en función de las características culturales y geográficas de los encuestados. Todo esto es de gran utilidad en los estudios sobre aprendientes no nativos ya que permite observar la influencia de la cultura del país de origen y de la comunidad a la que el alumnado se enfrenta.

Recientemente, se han propuesto estudios que pasan de la disponibilidad léxica general a la disponibilidad léxica terminológica, esto es, que analizan el léxico disponible específico de un determinado sector profesional desde el punto de vista de la terminología, entendida como disciplina y campo de trabajo, porque aumenta cada vez el interés por la comunicación especializada. Hasta hoy no tenemos constancia de mu-

chas investigaciones realizadas con discentes extranjeros de español,[8] pero disponemos de estudios dedicados a hispanohablantes nativos que trabajaron con el lenguaje especializado en los ámbitos de: comunicación y prensa (Gómez Sánchez y Guerra Salas 2004, Guerra Salas y Gómez Sánchez 2004), matemática (Urzúa Sáez y Echeverría 2006, Salcedo y del Valle Leo 2013), fisioterapia (Navarro 2010), mecánica (Madrigal-Melchor, Rivera-Juárez, Enciso-Muñoz y López-Chávez 2012), informática (Luján García y Bolaños Medina 2014, Tomé 2016) y viticultura (Toniolo y Zurita 2019).

Señalamos por último que las posibilidades de la disponibilidad léxica en didáctica de las lenguas, materna y extranjera, abarcan una gran variedad de aplicaciones teóricas y prácticas. En particular, en el marco de la didáctica de ELE es posible lograr una amplia serie de objetivos que Paredes García (2015) sintetiza como sigue: analizar el léxico disponible de los aprendices en un determinado momento; definir las etapas de aprendizaje; estudiar las interferencias de la L1 y de otras LE conocidas; cotejar el vocabulario de los participantes con otros grupos de estudiantes y con los nativos; establecer cuáles son las unidades poco rentables o los déficits léxicos; proporcionar comparaciones culturales; examinar la adecuación de manuales y otros materiales pedagógicos; eliminar el léxico superfluo de los contenidos didácticos para llevar a cabo una selección de los contenidos apropiada y; por último, ampliar la cobertura de los análisis hacia la terminología (español para fines específicos).

## 1.2 La riqueza léxica

Los estudios sobre riqueza léxica surgieron a mediados del siglo XX en Francia a partir de textos producidos por hablantes nativos para medir el vocabulario y analizar su variedad. Se trata de una metodología que estudia la habilidad de manejar un repertorio de vocablos en un discurso,

[8] Llevamos a cabo un experimento preliminar con un grupo restringido de alumnos de la enseñanza secundaria en el que propusimos tres centros de interés finalizados a recoger el léxico disponible relativo al sector comercio y *marketing* en Nalesso (2020). Actualmente, estamos desarrollando el proyecto *DiLexTerM: studio della disponibilità lessicale terminologica multilingue in studenti universitari*, financiado por el *Dipartimento di Studi Linguistici e Letterari* (PPD 2022) de la Universidad de Padua, que propone el análisis de índices de disponibilidad léxica general y terminológica en italiano, español, inglés, francés y portugués de los estudiantes matriculados en cursos de idiomas en mencionado departamento.

escrito u oral, mediante la aplicación de indicadores cuantitativos (Reyes Díaz 2007: 147).

A la hora de examinar la competencia léxica de un sujeto, un grupo o una comunidad es indudable que los análisis de disponibilidad léxica resultan útiles, pero no aportan ninguna información sobre la capacidad de utilizar realmente las palabras, es decir la capacidad comunicativo-productiva (vocabulario activo). De ahí que para examinar pormenorizadamente el dominio del léxico nos parece oportuno añadir un ulterior análisis realizado a través de la riqueza léxica, que estudia el caudal léxico contenido en un texto: un buen uso del vocabulario es esencial en la determinación de la calidad de un mensaje a la vez que suscita una impresión positiva en el receptor (Laufer y Nation 1995: 307). De un buen nivel de riqueza léxica se desprende un buen conocimiento y uso de la lengua.

### 1.2.1 Estudios pioneros

Si bien a lo largo de los últimos setenta años se han ido planteando múltiples pautas de análisis, los primeros patrones de cálculo y fórmulas siguen aun vigentes: Guiraud (1954), teniendo en cuenta la distinción entre palabra/vocablo y palabras nocionales/funcionales del discurso, realizó la primera investigación sobre riqueza léxica en textos escritos.[9] Tras presentar en la primera parte de su obra, *Les caractères statistiques du vocabulaire*, los índices estadísticos aplicables a la medición del vocabulario, precisó que su objetivo no consistía en la evaluación del léxico extraído de los textos poéticos propuestos en la segunda parte, sino que el análisis servía como punto de partida para corroborar las hipótesis y los planteamientos definidos «teóricamente y en abstracto» (Guiraud 1954: 74-75).

[9] A este propósito, López Morales (2011: 17) afirma: «La léxico-estadística ha sido el primer peldaño en la constitución de la estadística lingüística; […] abarca el conjunto de operaciones, a veces sumamente complejas, que toman como unidades de trabajo las palabras y los vocablos; la palabra, unidad del texto, y el vocablo, unidad del léxico». Cabe señalar que en el lenguaje común los términos *palabra* y *vocablo* suelen utilizarse indistintamente para referirse a los lexemas de una lengua, pero en léxico-estadística designan dos conceptos distintos: las palabras son todas las unidades aportadas por los informantes en un test, repeticiones incluidas; los vocablos son todos los ítems diferentes arrojados, sin tener en cuenta las repeticiones. En el primer caso, nos referimos a los *tokens*, es decir a todas y cada una de las unidades léxicas, repetidas o no, que aparecen en un corpus (o en un texto); en el segundo caso, aludimos a los *types*, esto es, a cada una de las piezas léxicas distintas que se registran en una muestra, sin repeticiones (Müller 1968).

A raíz de esta investigación, se sentaron las bases metodológicas de la riqueza léxica que han ido refinándose en los trabajos posteriores a partir de las siguientes fórmulas:[10]

$$R = \frac{V}{N} \qquad R = \frac{V}{(2)N}$$

Cuando se consideran simultáneamente palabras nocionales y funcionales se utiliza la primera; en cambio si se tienen en cuenta exclusivamente las palabras semánticas se aplica la segunda, en la que se duplica la extensión el texto ya que parece que las palabas con significado pleno cubren su mitad. Más tarde, Müller (1968) confirmó que cuantificar el vocabulario de un texto supone dos procesos complementarios que pueden llevarse a cabo al mismo tiempo o en dos etapas seguidas: primero, el conteo de todas las palabras que forman el texto (*tokens*) y, segundo, el conteo de los diferentes vocablos (*types*).

Sucesivamente, Ham (1979) y Tesitelová (1992) ampliaron la noción inicial de riqueza léxica abogando por incluir el criterio de frecuencia en los análisis. Según ellos, es necesario indagar también las ocurrencias de los vocablos, además de los otros factores, porque no es posible establecer el nivel de riqueza sin tener en cuenta su dispersión o concentración en el corpus.

En lo que se refiere a los estudios en ámbito hispano, fue pionero, de nuevo, el proyecto de López Morales (1984) que marcó un punto de inflexión en este panorama, pues aportó notables cambios metodológicos. Propuso nuevos procedimientos, operaciones matemáticas e índices que midiesen la riqueza del vocabulario: la diversidad léxica y el intervalo de aparición de palabras de contenido nocional (IAT), aunque siguió basándose en los porcentajes de vocablos y palabras contenidos en un texto y distinguiendo las palabras nocionales de las funcionales.

Por último, cabe incluir entre los estudios pioneros la propuesta de Linnarud (1986), quien fijó un límite máximo de palabras en las muestras textuales porque se dio cuenta de que la diversidad léxica, a saber, la diferencia entre la cantidad de *tokens* y *types*, no se puede comparar si la extensión de los textos es distinta.

### 1.2.2 La riqueza léxica en el mundo hispánico

[10] Donde: R= riqueza del vocabulario; V= palabras (nocionales y funcionales) del texto; N= longitud del texto.

López Morales introdujo la riqueza léxica en el mundo hispánico, impulsando el desarrollo de proyectos tanto en Hispanoamérica como en España. Se abrió una corriente que, si de un lado sigue sus directrices principales, del otro, ha ido proponiendo cambios teóricos y procedimentales.

Haché (1991) destacó muy pronto que es fundamental incluir en los análisis textos del mismo tamaño ya que esto incide significativamente en los cómputos en términos de diversidad y densidad léxica. De estos planteamientos se corroboró lo que ya había demostrado Linnarud (1986), que después de las primeras cien palabras el cálculo de la riqueza léxica se distorsiona: «[…] sabemos que los datos se desvirtúan al pasar un texto de las 100 palabras. […] Necesitamos, pues, para tener metas claras, exámenes de textos escritos que se consideren ilustrativos, de los que se pueden tomar varias calas pero nunca superiores a las 100 palabras» (López Morales 2011: 24). Torres González (1999, 2003a, 2003b) trabajó con un corpus de escritos redactados por estudiantes preuniversitarios de diferentes niveles analizados según algunos de los índices que ya se habían empleado en investigaciones anteriores (número de vocablos, IAT, número de palabras nocionales), con la inclusión de uno nuevo: el índice de hápax, para detectar las unidades que solo aparecen una vez en un texto y contribuye a establecer el grado de variación léxica.

En este contexto, los proyectos se desarrollaron en gran medida según los patrones de la sociolingüística: analizaron cómo y cuánto la riqueza léxica se veía afectada por variables extralingüísticas como sexo, nivel sociocultural, tipo de escuela y zona geográfica de procedencia (entre las primeras investigaciones publicadas antes del comienzo del siglo XXI contamos con Ávila 1986, Echeverría *et al.* 1992, Portela 1992, Cintrón Serrano 1992, Andrés Pérez 1997). Además, los mismos autores observaron las diferencias causadas por el tipo de texto dándose cuenta de que la riqueza cambiaba según fuera narrativo, descriptivo, expositivo o argumentativo.

Todo esto hace patente que desde finales de los años ochenta, en distintos países hispanos aparecieron estudios que evaluaban el uso del vocabulario, cuyos objetivos, a nuestro parecer, se pueden asemejar a los del PPHDL desde otra perspectiva analítica.

### 1.2.3 La riqueza léxica en ELE

En el marco investigativo del español lengua extranjera se ha promovido hasta ahora un número limitado de proyectos dedicados a la riqueza

léxica si los comparamos con los de disponibilidad, pero parece que cada vez más se están planteando nuevos estudios en distintas zonas del mundo.

Pionera en este ámbito fue la memoria de máster defendida por García Rosas en 1996 que tenía el objetivo de analizar un corpus de redacciones escritas en función del tipo de texto y de factores sociolingüísticos (nivel de ELE, sexo, edad, lengua materna) de un grupo mixto de aprendices. Los resultados más relevantes demostraron que el modo discursivo que ofrece un rendimiento mejor es el narrativo, que la edad y la lengua materna parecen no influir en la riqueza léxica y, por último, que los informantes con un mayor dominio lingüístico aportan un número mayor de palabras y vocablos como es esperable.

Los demás trabajos sobre la producción escrita en ELE que emplean índices de riqueza aparecieron aproximadamente una década después y en muchos casos se trata de investigaciones llevadas a cabo con una muestra restringida y que todavía no presentan una fijación metodológica:

- Cuba Vega y Cuba (2004) y Baerlocher (2013) estudiaron el vocabulario de universitarios brasileños;
- Lucas Puerta (2006) trató la competencia léxica de aprendices de ELE francófonos de ascendencia hispanófona;
- Berton (2014, 2020) analizó un conjunto de textos escritos por discentes suecos;
- Castañeda-Jiménez y Jarvis (2014) examinaron la producción escrita de estudiantes estadounidenses;
- Wang (2016) trabajó con aprendientes sinohablantes;
- Lucha Cuadros y Díaz (2016) publicaron un estudio sobre la diversidad léxica en la producción escrita de alumnos de procedencia mixta;
- Tracy-Ventura (2017) se dedicó a la cuestión de la sofisticación léxica en estudiantes anglófonos de nivel universitario, que habían vivido nueve meses en un país hispanohablante;
- García Marcos (2019), por último, destinó parte de su trabajo a la riqueza léxica producida en un corpus de textos escritos por parte de un grupo de niños inmigrantes en la zona de Almería para compararlos con los escritos de sus compañeros nativos.

La situación de los estudios dedicados a italófonos cuenta con la tesis de máster de Basso (2017) donde se analiza la producción escrita de estudiantes universitarios, cuya finalidad es medir originalidad, variedad, densidad léxica y grado de incidencia de los errores en la variación del vocabulario. También con italianos son nuestros análisis realizados me-

diante distintos índices porque comprobamos un texto que presenta un buen índice de variedad no necesariamente tiene el mismo resultado en lo que se refiere al uso de las palabras nocionales. Dedujimos, por ende, que la aplicación de diferentes medidas de cálculo es esencial en este tipo de experimento (Nalesso 2018a, 2019b).

En resumidas cuentas, se abrió el camino hacia los estudiantes de ELE, pero aun son muchos los pasos por cumplir para que esta metodología se difunda como instrumento habitual de análisis. Sería deseable fomentar el planteamiento de estudios y grupos de investigación que profundicen los conocimientos adquiridos e incrementen el protocolo de las pruebas empíricas. Como asevera uno de los autores pioneros en este campo, López Morales (2011: 27): «Vale la pena rescatar aquella experiencia, actualizar esos materiales y proseguir con la tarea».

### 1.2.4 Otras aplicaciones de la riqueza léxica

La riqueza léxica puede ser explotada con provecho en ámbitos distintos de la investigación lingüística y didáctica, pese a que todavía no disponemos de muchos estudios que tratan desde otras perspectivas los datos obtenidos en los experimentos. Al igual que hemos visto en el apartado de disponibilidad, sería interesante poder aplicar esta metodología según los enfoques de otras disciplinas, como la dialectología, etnolingüística y psicolingüística para profundizar en calidad los resultados.

Lo que más sobresale hasta ahora es el interés de la sociolingüística. Como hemos apreciado en el marco de los estudios hispanos, se ocupa de investigar el influjo de los condicionantes sociales, culturales y económicos en la capacidad expresiva de los individuos. En este sentido, Roberto, Martí y Salamó Llorente (2012) pusieron de manifiesto la utilidad de la riqueza léxica «para predecir atributos demográficos latentes en textos de opinión del español» y, en la misma línea, Ávila Muñoz (2016) cotejó lo datos recogidos en diferentes investigaciones para averiguar qué patrones extralingüísticos condicionan la variación del vocabulario, y cuánto.

Sin embargo, cabe subrayar que los buenos resultados obtenidos hasta la fecha han hecho hincapié en su utilidad para la didáctica de lenguas ya que, además de medir la capacidad de uso del vocabulario, pueden emplearse en la creación de material didáctico encaminado a incrementar capacidad y calidad productiva de los aprendientes. Asimismo, estos análisis se revelan una herramienta aprovechable en los procesos de evaluación y programación didáctica, permiten valorar los conocimientos del alumnado y detectar lo que ya sabe antes de proponer otro *input*. En otras

palabras, ayuda a conocer el dominio del léxico activo de un sujeto o de un grupo teniendo constancia del vocabulario manejado para programar la enseñanza de las unidades léxicas desconocidas (Reyes Díaz 2007).

Las posibilidades y las ventajas ofrecidas son muchas y pueden llevar a resultados interesantes. Haché (1991) presentó una lista de aplicaciones utilizadas en estudios realizados con niños hispanohablantes nativos que podemos exportar a nuestro contexto: las pruebas de riqueza léxica permiten averiguar qué y cuánto vocabulario conocen los discentes para que se pueda adaptar la enseñanza en consecuencia; el léxico utilizado en los textos indica qué voces habrían de incluirse en los materiales didácticos para facilitar la comprensión del contenido y el gusto por la lectura en la lengua objeto; las listas de los lexemas activados pueden contrastarse con el vocabulario de los manuales y lecturas graduadas para comprobar si está en línea con las necesidades del alumnado; los resultados son útiles en la elaboración de diccionarios de aprendizaje para no nativos y en la evaluación del desarrollo del vocabulario activo de cada alumno y del estudiantado en general.

## Capítulo 2.
## Metodología de la investigación

Exponemos ahora la metodología empleada en la parte empírica de la investigación detallando la selección y la distribución de la muestra; las técnicas de elicitación de los datos y los procesos de recogida; los criterios de edición y el tratamiento del material. Es oportuno presentar nuestro *modus operandi* ya que nos basamos en los patrones expuestos en el marco teórico (Carcedo 2000c, Samper Hernández 2002, Sánchez-Saus 2016, Hidalgo 2019), pero algunas de las decisiones tomadas son distintas con respecto a otras investigaciones. Nuestro objetivo es analizar y evaluar la competencia léxica de aprendientes italianos de ELE: ofrecemos una imagen del procesamiento y del dominio del léxico que sea lo más fidedigna posible por lo que aportamos algunas modificaciones de lo ya practicado en otros estudios de la misma naturaleza.

Nos centramos en dos métodos empíricos, la disponibilidad y la riqueza léxica, encaminados a medir el vocabulario que un estudiante sabe y puede activar en una situación comunicativa ya que: «La riqueza, el alcance y el control del vocabulario son parámetros importantes en la adquisición de la lengua y, por ello, de la evaluación del dominio de la lengua que tiene el alumno, y de la planificación del aprendizaje y de la enseñanza de lenguas» (MCER 2002: 86). La aplicación de las dos metodologías permite comprender mejor la competencia léxica de los informantes a través de un estudio preciso y coherente mediante la correlación de los datos obtenidos en las dos pruebas. La disponibilidad se coloca en el plano paradigmático del léxico y analiza no solo de cuáles palabras dispone un aprendiente, sino cómo se organizan conceptualmente en su lexicón mental. Por su parte, la riqueza se sitúa en el plano sintagmático y examina cuántas palabras distintas es capaz un aprendiz de emplear en un texto y cómo lo hace.

## 2.1 La muestra

La investigación se basa en un experimento realizado a partir de una prueba de medición de la competencia léxica que diseñamos *ad hoc* para este proyecto suministrada a cien alumnos de la *Università degli Studi di Padova* durante el año académico 2017/2018. Se trata de aprendientes italianos de ELE procedentes de los cursos de *Lingua Spagnola II* y *III* del grado en lenguas extranjeras.[11] Optamos por trabajar con esta muestra, que tiene un nivel B1 y B2 de español, ya que nuestro propósito es indagar la evolución del conocimiento del alumnado de un nivel a otro y comprobar si dentro del mismo nivel se cumple un significativo desarrollo de la competencia tras la asistencia a un curso anual.

El paso previo a la suministración de las pruebas fue contactar con los docentes titulares de los cursos para presentarles el proyecto y pedirles que nos concediesen unas horas de clase. Una vez obtenido el permiso, proporcionamos la misma prueba dos veces a los mismos estudiantes, una a comienzo y una a final del año académico para capturar una imagen transversal y una longitudinal de la competencia léxica. En concreto, la primera administración, consistente en recopilar cien encuestas para el análisis transversal de los resultados, se llevó a cabo entre los meses de octubre y noviembre de 2017. En el mes de mayo de 2018, suministramos la segunda prueba al grupo más avanzado, recogiendo otras cincuenta encuestas para el análisis longitudinal. De esta forma recopilamos dos corpus: el *corpus general* está formado por los datos de la primera prueba realizada por cincuenta aprendices del segundo año (nivel B1) y cincuenta del tercero (nivel B2, que formarán el *corpus B2_a* en el análisis longitudinal); el *corpus B2_b* consta de los datos de la segunda prueba efectuada por los cincuenta estudiantes del tercer año a final del curso.

## 2.2 Las variables sociolingüísticas

Como es habitual en los estudios de esta índole, trabajamos con algunas variables sociolingüísticas con el propósito de comprobar si influyen o no en el desarrollo de la competencia léxica. De los muchos factores

[11] El sistema italiano prevé la división de los estudios universitarios en dos ciclos: un período de tres años que lleva a la licenciatura de *Laurea Triennale* y un segundo de dos años que permite la obtención de otro título, *Laurea Magistrale* (esto es, el título de máster o magister en España), que posibilita el acceso a los programas de posgrado. En este caso trabajamos con los alumnos de segundo y tercer año del primer ciclo.

considerables elegimos el sexo, el nivel de ELE y el número de lenguas extranjeras conocidas.[12]

Dividimos los informantes en función de tales factores solo tras empezar la fase de análisis de los datos ya que, inicialmente, seguimos un proceso aleatorio de selección de la muestra durante el cual consideramos solo la asistencia a los cursos académicos elegidos. Por esta razón, suministramos las pruebas durante una clase cualquiera sin avisar previamente al alumnado para eludir el absentismo.

### 2.2.1 Sexo

El género es el primer condicionante que utilizamos para la clasificación de los encuestados, se trata de un factor presente en todos los proyectos de corte sociolingüístico para comprobar si se detecta la predominancia de uno frente al otro.

| | ***Sexo*** | | |
|---|---|---|---|
| | **Mujer** | **Hombre** | **Total** |
| Corpus general | 87 | 13 | 100 |
| Corpus B2_a | 41 | 9 | 50 |
| Corpus B2_b | 45 | 5 | 50 |

Tabla 1. Distribución de los informantes en función de la variable sexo.

### 2.2.2 Nivel de español

El nivel de ELE es el segundo factor que permitió comprobar la hipótesis de que a mayor dominio lingüístico corresponde un mayor caudal de léxico activable. El profesorado nos proporcionó la correspondencia

[12] No consideramos la lengua materna, una de las variables más utilizadas en este tipo de investigación, porque asumimos que los alumnos fuesen italófonos de nacimiento o bien que, a pesar de no ser nativos, manejasen el italiano como LM al estudiar en contexto universitario italiano. Por otra parte, en el grupo participante solo había tres estudiantes extranjeros, con lo cual este condicionante no habría sido significativo. Tampoco tuvimos en consideración la edad, el nivel sociocultural y el tipo de centro porque no eran relevantes para nuestros objetivos.

entre los niveles en los que subdividía el estudiantado conforme a las indicaciones del MCER, por eso, trabajamos con los cursos de segundo y tercer año de licenciatura en virtud de nuestro propósito de analizar el nivel intermedio formado por los subgrupos B1 (umbral) y B2 (avanzado).

| | ***Nivel de ELE*** | | |
|---|---|---|---|
| | **B1** | **B2** | **Total** |
| Corpus general | 50 | 50 | 100 |
| Corpus B2_a | – | 50 | 50 |
| Corpus B2_b | – | 50 | 50 |

Tabla 2. Distribución de los informantes en función de la variable nivel de ELE.

En el cuestionario sociológico de la prueba (véase §2.3) preguntamos a los participantes que valorasen su nivel de español en una escala que recogía los niveles inicial, intermedio y avanzado para sondear la percepción de sus conocimientos y averiguar si había correlación con los estándares establecidos por el profesorado.

| | **Inicial** | **Intermedio** | **Avanzado** | **Total** |
|---|---|---|---|---|
| Corpus general | 15 | 74 | 11 | 100 |
| Corpus B2_a | 1 | 43 | 6 | 50 |
| Corpus B2_b | 1 | 39 | 10 | 50 |

Tabla 3. Percepción de los informantes de su nivel de ELE.

Los estudiantes se conforman con la opinión de los docentes y con las indicaciones curriculares de los cursos en los que están matriculados, los porcentajes más altos se detectan en el nivel intermedio. Solo el 15% de los componentes del corpus general contesta «inicial», mientras que el 20% de los del B2_b se siente más cercano a un nivel avanzado.

### 2.2.3 Conocimiento de otras lenguas extranjeras

Por último, analizamos la influencia del conocimiento de otras lenguas extranjeras a fin de comprobar la hipótesis de Cummins (1979), la cual plantea que un individuo que ya conoce una lengua extranjera em-

plea sus competencias lingüísticas durante el aprendizaje de un nuevo idioma como elemento facilitador.

| | *Otras LE conocidas* | | |
|---|---|---|---|
| | **LE =2** | **LE >2** | **Total** |
| Corpus general | 19 | 81 | 100 |
| Corpus B2_a | 9 | 41 | 50 |
| Corpus B2_b | 6 | 44 | 50 |

Tabla 4. Distribución de los informantes en función de la variable conocimiento de otras LE.

### 2.2.4 La codificación de las variables

Para trabajar con el programa *Dispogen* (véase §2.4.2) codificamos las variables y sus variantes como se indica a continuación. La tabla 5 proporciona la clave de identificación de cada encuestado, necesaria para dividir los grupos, los subgrupos y reconocer los informantes.

| **Variable** | **Variante** | **Código** |
|---|---|---|
| Sexo | Mujer | 1 |
| | Hombre | 2 |
| Nivel de ELE | Nivel B1 | 1 |
| | Nivel B2 | 2 |
| Conocimiento de otras LE | LE =2 | 1 |
| | LE >2 | 2 |

Tabla 5. Codificación de las variantes de las variables.

## 2.3 La encuesta

La encuesta que diseñamos para la recogida del material sigue los modelos tradicionales de las investigaciones en disponibilidad y riqueza léxica, empieza con unas disposiciones introductorias –que presentan los

objetivos y las motivaciones del proyecto– y cuenta con tres apartados: el cuestionario sociológico; la prueba de disponibilidad léxica; la prueba de riqueza léxica.[13] En el momento de la suministración siempre se siguió el mismo procedimiento, aplicando un protocolo estándar que garantizase la repetibilidad del experimento:

– presentación del investigador y del estudio,
– explicación de los contenidos y de las instrucciones,
– aclaración de la anonimidad,
– administración de las hojas,
– realización de la prueba.

El primer apartado, dos páginas, consta de dos partes cuyo propósito es la recogida de datos personales, a saber, informaciones sobre el bagaje lingüístico de los encuestados y su opinión sobre el aprendizaje de una LE. Teniendo en cuenta la preocupación constante de los estudiantes frente a las pruebas, se insistió en que esto servía para informes estadísticos y que solo nos interesaba su opinión por lo que no había respuestas correctas. No establecimos un límite de tiempo para rellenar esta parte, sino que nos aseguramos de que todos los participantes hubiesen tenido margen suficiente para completarla antes de pasar a la ejecución de las pruebas.

El cuestionario contenía más preguntas de las necesarias, ya que no todas las informaciones fueron codificadas, al haber fijado los factores sociolingüísticos *a posteriori* debido a que no sabíamos quién habría participado realmente en la encuesta, descartando, por tanto, algunos de los datos aportados.

En lo que se refiere a las preguntas finalizadas a recoger las opiniones de los informantes en torno al aprendizaje de una lengua extranjera, señalamos que servían para conocer sus actitudes y aportar información al profesorado de modo que pudiera diseñar una propuesta didáctica adecuada.[14] Además, permitieron comprobar si los estudiantes son conscientes de cómo funciona la adquisición de una lengua extranjera, ya que es importante concienciarlos para que su competencia se desarrolle con éxito.

[13] Se construyó conforme a criterios de fiabilidad, validez y practicabilidad ponderados en función de la muestra elegida. Tratándose de aprendientes no nativos, las instrucciones se escribieron en español, pero, con el fin de garantizar la plena comprensión y evitar malentendidos, las explicaciones orales se dieron en italiano en el momento de realización de la prueba.

[14] Cfr. la enseñanza centrada en el alumno (MCER 2002).

Las respuestas a la pregunta que interrogaba sobre los elementos más importantes de una lengua, donde se podía elegir entre gramática y léxico, destacan que los informantes saben que ambos son componentes fundamentales para el conocimiento de un idioma. Lógicamente, a la pregunta siguiente, «¿Cuánto cree que es importante conocer el vocabulario de una LE?», la mayoría contestó que aprender léxico es importante. Pero, la pregunta «¿Cómo cree que aprende más vocabulario?» pone de relieve que suponen que el léxico se adquiere de forma inconsciente, sobre todo mediante la lectura de textos, escucha de canciones o visión de películas. Les parece que las palabras se aprenden como resultado de varias actividades, no específicas, ya que no más del 20% indicó la respuesta «ejercicios de léxico». Quizá esto se deba al *modus operandi* al que están acostumbrados: las clases de español se centran más en la gramática, la contrastividad, el análisis de textos y la literatura, descuidando las actividades específicamente diseñadas para el conocimiento del vocabulario.[15]

El segundo apartado presenta la prueba de disponibilidad léxica. Se trata de una encuesta con los test estandarizados, compuesta por cuatro páginas de cuatro columnas cada una; en la parte superior de cada página constan los nombres, numerados de 1 a 16, de los centros de interés que el investigador ha ido presentando a medida que los informantes iban completando para evitar que pensasen antes de tiempo en las palabras vinculadas con el tema siguiente y que las respuestas fuesen el resultado de una reflexión (Samper Hernández 2002). Las fichas se rellenan con las unidades léxicas proporcionadas en relación con cada estímulo empezando por la primera línea en alto de modo que los ítems más disponibles son los que aparecen en las primeras posiciones de las listas. Cada columna permite un número máximo de aportaciones, pero para trabajar listas abiertas se puede seguir en el reverso del folio si la ficha no es suficiente. En lo que atañe al margen temporal, dejamos dos minutos para cada campo sin permitir volver atrás una vez terminado el tiempo establecido. Parece que este intervalo es suficiente para escribir bastantes unidades e incluso para rastrear en la memoria algunas más. Durante la ejecución, exhortamos a los informantes a escribir todas las palabras que se les ocurrían, aunque no estuviesen seguros de su grafía o de su significado.

[15] Merece la pena subrayar que si bien el Enfoque Léxico (Lewis 1993, 1997, 2000) ponga de relieve la importancia del aprendizaje incidental del vocabulario, este no es suficiente para el pleno desarrollo de la competencia léxica: hay que insistir en la utilidad de programar tareas que se dirijan a una práctica explícita y a un aprendizaje consciente del alumno.

Es bien sabido que los centros de interés son elementos nucleares de los estudios en disponibilidad léxica y que están sujetos a arbitrariedad debida a los intereses de cada investigador, que emplea en su estudio los que le resultan más provechosos para alcanzar sus objetivos y que sean adecuados a sus informantes. Sin embargo, somos conscientes también de la utilidad de aplicar una metodología empírica común para facilitar una visión comparativa de las investigaciones, respondiendo a la urgencia de revisar los datos desde una perspectiva sinóptica por lo que es importante poder cotejar distintos trabajos a partir de los campos empleados en las pruebas (Callealta y Gallego Gallego 2016, Jímenez Catalán 2017, Paredes García 2017). Para ello, aquí utilizamos los estímulos propuestos en el PPHDL, pese a que hayan sido cuestionados, reducidos, reformulados o sustituidos en muchas ocasiones. Además, como nos basamos en los patrones del MCER, algunos autores demostraron la correspondencia de estos campos con las indicaciones europeas.[16] Asimismo, mediante nuestro análisis proporcionamos una respuesta más a la cuestión crítica sobre la inadecuación de algunos temas en relación con necesidades e intereses del siglo XXI. Incluimos en la prueba los siguientes centros:

– CI01, "partes del cuerpo",
– CI02, "la ropa",
– CI03, "partes de la casa (sin muebles)",
– CI04, "los muebles de la casa",
– CI05, "alimentos y bebidas",
– CI06, "objetos colocados en la mesa para la comida",
– CI07, "la cocina y sus utensilios",
– CI08, "la escuela: muebles y materiales",
– CI09, "iluminación y calefacción",
– CI10, "la ciudad",
– CI11, "el campo",
– CI12, "medios de transporte",
– CI13, "trabajos del campo y del jardín",
– CI14, "los animales",
– CI15, "juegos y distracciones",
– CI16, "profesiones y oficios".[17]

[16] Bombarelli (2005) y Bartol Hernández (2010) detectaron cierta correlación entre la distribución de los CI tradicionales y los temas propuestos por el MCER. Paredes García (2015) y Ávila Muñoz (2016, 2017) cotejaron el léxico disponible extraído de sus trabajos (llevados a cabo, respectivamente, en la Comunidad de Madrid y en la provincia de Málaga) tanto con las indicaciones europeas como con las nociones específicas del PCIC para medir el grado de rentabilidad de las unidades léxicas propuestas por tales documentos.

[17] A los que hemos asignado las siguientes siglas: CI01, "partes del cuerpo" [CUE]; CI02,

El tercer apartado, una página, corresponde a la prueba de riqueza léxica en la que se requería que los informantes escribiesen un texto sobre un viaje. Elegimos este tema porque queríamos indagar la capacidad real de uso del léxico: uno más complejo o especializado no habría aportado los datos que necesitábamos, pues, intentamos maximizar la posibilidad de expresión de los aprendientes de este nivel.[18] El folio presenta una ficha con un máximo de líneas, aunque era posible recurrir al reverso al necesitar más espacio. Lo importante era que los relatos no excediesen las 120 palabras para permitir la recopilación de un corpus homogéneo y porque parece que después de las primeras 100 los cálculos se desvirtúan, como vimos. Era imprescindible, para el éxito de la prueba, que cada encuestado aportase una cantidad suficiente de unidades léxicas, dejando un cierto margen de error. Los encuestados disponían de 30 minutos para redactar su relato, un intervalo de tiempo suficiente considerados el nivel de ELE y la temática sencilla. Recopilamos las primeras 100 palabras de cada redacción para obtener precisión y estabilidad en los cómputos. Asimismo, trabajamos con textos constituidos por un número igual de *tokens* porque no nos interesaba comprobar la variación de los índices según el aumento del tamaño del texto. De todo el material originado durante esta tarea, desechamos las narraciones que no cumplían con los requisitos preestablecidos.

## 2.4 Análisis de la disponibilidad léxica

Para analizar cuantitativa y cualitativamente el léxico disponible nos basamos en los patrones del PPHDL y de otros trabajos dedicados a aprendientes no nativos de español. Seguimos estas directrices empíricas, ya que hemos considerado esencial poder cotejar nuestros resultados con los de otros proyectos.

"la ropa" [ROP]; CI03, "partes de la casa (sin muebles)" [CAS]; CI04, "los muebles de la casa" [MUE]; CI05, "alimentos y bebidas" [ALI]; CI06, "objetos colocados en la mesa para la comida" [MES]; CI07, "la cocina y sus utensilios" [COC]; CI08, "la escuela: muebles y materiales" [ESC]; CI09, "iluminación y calefacción" [ILU]; CI10, "la ciudad" [CIU]; CI11, "el campo" [CAM]; CI12, "medios de transporte" [TRA]; CI13, "trabajos del campo y del jardín" [TRC]; CI14, "los animales" [ANI]; CI15, "juegos y distracciones" [JUE]; CI16, "profesiones y oficios" [PRO].

[18] Laufer y Nation (1995: 308) destacaron la importancia de medir la competencia de un sujeto según la familiaridad que tiene desarrollada sobre un tema.

### 2.4.1 Los índices de disponibilidad léxica

Los principales indicadores que utilizamos para determinar el léxico disponible de nuestros informantes son el total de palabras y de vocablos.

Medimos también dos índices que informan sobre el grado de concreción semántica de un área léxica y su constitución interna:

– el índice de cohesión (Echeverría *et al.* 1987) estudia la coincidencia de las respuestas por CI, esto es, si el léxico es homogéneo dentro del corpus:[19]

$$\frac{\text{media de palabras por informante}}{\text{número total de vocablos}}$$

– la densidad léxica contabiliza la media de repeticiones de las respuestas:

$$\frac{\text{número de palabras}}{\text{número total de vocablos}}$$

Desarrollamos tres tipos de análisis: el transversal examina los resultados generales del corpus y los resultados por variable. El longitudinal está finalizado a averiguar si se detecta un desarrollo de la competencia léxica entre el comienzo y el final del año académico, basándonos en los corpus B2_a y B2_b. El tercer paso, el análisis comparativo, coteja nuestros datos con diferentes trabajos de la misma índole con el propósito de evaluar el bagaje léxico de distintos grupos de estudiantes de ELE de diferentes LM, entornos y contextos de aprendizaje. Además, el análisis descriptivo estadístico permite examinar el comportamiento intragrupal e intergrupal de los participantes: estudiamos la simetría o la dispersión mediante diagramas de caja y bigotes que muestran la varianza en el número de unidades aportadas.

Por su parte, el análisis cualitativo de los resultados ha arrojado nueva luz en torno a las respuestas. Examinamos el tipo de léxico activado observando cuáles son las temáticas predominantes y las categorías gramaticales más difundidas en cada campo nocional mediante la extracción

[19] Los valores del IC pueden oscilar entre 0 y 1: cuanto más se aproxima a 1, mayor cohesión se detecta porque las respuestas coinciden en gran medida, el centro se define más compacto o cerrado. Al contrario, cuanto más se aproxima a 0, resulta una menor coincidencia en las aportaciones (una menor cohesión), con lo cual el centro es más difuso o abierto.

de los vocablos que presentan un índice de disponibilidad (ID) igual o mayor a 0,1. Fijamos esta medida de corte sin aportar solamente las primeras unidades de los listados (Carcedo 2000c, Samper Hernández 2002) porque evita la interferencia de las unidades actualizadas individualmente o por pocos sujetos (Hidalgo 2019).[20] Para ello, compilamos otras listas que contienen los vocablos con ID ≥ 0,1 (que presentan una frecuencia acumulada ≥ 30% y de aparición ≤ 75%) a partir de las cuales calculamos la cardinalidad del conjunto, es decir el porcentaje de compatibilidad entre las distintas agrupaciones de informantes (Ávila Muñoz y Sánchez Sáez 2010, 2011) con el fin de averiguar qué correspondencia o similitud cualitativa existe entre las respuestas.[21]

Como colofón, cotejamos estos vocablos con el *Corpus de Referencia del Español Actual* (2008), el corpus de referencia para el español de la RAE, y el *Corpus de aprendices de español* (2018) del Instituto Cervantes, que recoge textos producidos por estudiantes de ELE de diferentes lenguas maternas y distintos niveles de competencia (de los cuales extrajimos los datos del intermedio). Nuestro propósito era averiguar si el léxico más disponible activado coincidía con los vocablos de estos listados para comprobar si los encuestados conocen, y tienen a su disposición, los lexemas de uso más común en español y que se encuentran frecuentemente en las producciones de otros discentes del mismo nivel.

### 2.4.2 Tratamiento informático de los datos

La digitalización del material constituye una de las fases más importantes de la investigación; para ello nos servimos del programa *Dispogen II* (Echeverría *et al.* 2005), que basándose en la fórmula de López Chávez y Strassburger Frías (1987), permite calcular el total de palabras y vocablos, el promedio por informante y por CI, así como el índice de cohesión. Además, extrae el ID de cada vocablo proporcionando información sobre su frecuencia, porcentaje de aparición y frecuencia acumulada. Puede

[20] Ruiz Basto (1987) fue el primero en plantear que el vocabulario activo está formado por aquellos vocablos que presentan un ID superior a 0,1 y actualizado por al menos un 20% de los informantes.

[21] Se calcula a partir del número de vocablos con ID ≥ 0,1 de cada grupo considerado (cardinalidad), sumando las unidades convergentes en los grupos (intersección), de aquellas que no están presentes en todos, pero sí en varios (unión) y de las exclusivas de cada uno (complemento). La suma de la unión y de los complementos se denomina suma disyuntiva. La compatibilidad coincide con el peso de la intersección con respecto al total de vocablos (intersección + suma disyuntiva).

examinar y cruzar los datos por variable sociolingüística, algo muy provechoso para este proyecto, pues es posible computar cinco variables a la vez, lo que nos permitió cotejar distintos datos con facilidad, ya que trabajamos con tres.

Como vimos, registramos las respuestas de cada informante con un código que se divide en tres segmentos a fin de permitir el correcto funcionamiento del *software*:

> **12200 001 01** pie, dedo, rodilla, brazo, pierna, cuello, barriga, mano, codos, espalda, cara, boca, nariz, cabeza, orejas, ojos
>
> **12200 002 01** mano, boca, brazo, barriga, hígado, dedo, uña, pierna, pie, pelo, mejilla, lengua, rodilla, ojos, estómago, corazón, cuello, cara
>
> **21200 003 01** mano, cabeza, pelo, ojos, boca, nariz, espalda, corazón, orejas, pie, rodilla, pierna, dedo, piel
>
> **21100 004 01** rodilla, pierna, cuerpo-humano, pie, pelo, ojos, mano, espalda, bigote
>
> **21200 005 01** cabeza, pie, ojos, lengua, mano
>
> **11200 006 01** barriga, ojos, nariz, pierna, cabeza, pelo, boca, mano, rodilla, dedo, pecho, cintura, garganta
>
> **11200 007 01** naso, nariz, boca, ojos, mano, espalda, pelo
>
> **11200 008 01** ojos, pierna, espalda, nariz, boca, brazo, cintura, mano, cabeza, estómago, frente
>
> **11200 009 01** cabeza, boca, pelo, rubio, pelirrojo, oír, ojos, marrón, azul, labios
>
> **21100 010 01** cabeza, mano, pie, rodilla, tobillo, ojos, cara, pelo, nariz, boca, espalda

Los cinco dígitos iniciales coinciden con las variables contempladas en el estudio:[22] los tres primeros indican las variantes del género, del nivel de ELE y del conocimiento de otras LE; la serie siguiente indica el número asignado al informante; las dos cifras finales se refieren al centro de interés. Para poner un ejemplo, si tenemos un sujeto identificado con **12200**

[22] Es necesario poner cinco dígitos porque el programa no funciona si se ponen menos, por ello introdujimos los requeridos, pese a que los últimos dos (00) no estuviesen relacionados con ninguno de los condicionantes sociolingüísticos.

**001 01**, se trata de las respuestas relativas al CI01, "las partes del cuerpo", de una mujer de nivel B2 que conoce más de dos lenguas extranjeras.

### 2.4.3 La edición de los datos

Un paso previo al procesamiento informático de los datos es la estandarización de las respuestas para que el corpus resulte homogéneo y consienta análisis fiables.

El establecimiento de las normas de despojo es una búsqueda de soluciones que no siempre los investigadores comparten, pero es necesaria ya que de estas dependen los resultados de la investigación. En este estudio, las primeras fases de revisión y corrección fueron relativamente sencillas,[23] mientras que la lematización ha implicado un mayor esfuerzo debido a las tomas de decisiones que conlleva. Procedimos a una fase de edición mediante la aplicación de criterios generales comúnmente utilizados en todo tipo de proyecto de disponibilidad léxica, junto a criterios específicos utilizados en trabajos dedicados a aprendientes no nativos de español y a criterios particulares de la producción de nuestros informantes por centro de interés.

#### Criterios comunes de la investigación en disponibilidad léxica

En primera instancia, aplicamos, "con cierta laxitud" debido al tipo de informante, pautas comunes de la investigación en disponibilidad léxica fijados por Samper Padilla (1998) y planteados inicialmente para el PPHDL:

- Eliminación de los términos repetidos: al encontrar la misma palabra más de una vez dentro de un centro de interés descartamos la que estaba en la posición más baja.
- Corrección de la ortografía: corregimos los errores ortográficos incluyendo en el corpus todas las entradas que pudimos documentar.
- Unificación ortográfica: al detectar variaciones ortográficas de una palabra registramos la forma más repetida en el corpus.
- Unificación de variantes meramente flexivas: entre las respuestas aparecen palabras con el género y el número que acude más instintivamente a la mente de los informantes, por lo que, a fin de man-

[23] Para solucionar eventuales dificultades de comprensión tomamos como referencia las siguientes obras de las Academias de la Lengua Española (RAE y ASALE): *Diccionario de la lengua española* (2017); *Diccionario panhispánico de dudas* (2005); *Nueva gramática de la lengua española* (2009); *Ortografía de la lengua española* (2010); *Corpus de referencia del español actual* (2008, 2015).

tener homogeneidad, se reduce cada entrada a la forma no marcada del paradigma (el masculino singular de sustantivos y adjetivos, el infinitivo de verbos). Sin embargo, es preciso señalar un par de excepciones relativas a (i) el número: los *pluralia tantum* (*tijeras*); las palabras compuestas cuyo segundo elemento es en plural (*rascacielos*, *videojuegos*); las lexías que se usan mayoritariamente en plural y designan una realidad múltiple (*vacaciones*, *orejas*) pese a que tengan su entrada singular en el diccionario (*vacación*, *oreja*); los lexemas que implican una acepción diferente según el número (*padres*, *deberes*, *apuntes*); (ii) el género: los significantes que se refieren a dos entidades distintas según el sexo como *ternero*-animal y *ternera*-alimento o los heterónimos *hombre*/*mujer*, *gallo*/*gallina*, *actor*/*actriz*.

- Unificación de derivados regulares: unificamos las variantes morfológicas bajo una sola forma que no supone alteración de significado, esto es, aparecen los lexemas primitivos sin derivaciones (*casita* > *casa*).
- Unificación de formas plenas y acortamientos: utilizamos los paréntesis[24] para registrar formas plenas, acortamientos o abreviamientos bajo una sola entrada (*bici* o *bicicleta* > *bici(cleta)*, *tele* o *televisión* > *tele(visión)*, *bus* o *autobús* > *(auto)bús*).
- Tratamiento de los sintagmas: las unidades multipalabra se trataron como un único ítem, de hecho, utilizamos guiones para la recopilación de los sintagmas (*ama-de-casa*) y de las construcciones idiomáticas introducidas por *ir de* (*ir-de-compras*).
- Tratamiento de las marcas comerciales: aceptamos exclusivamente las marcas comerciales consideradas apelativos comunes por la mayoría de los hispanohablantes (*Coca-Cola*, *PlayStation*, *Facebook*).

### Criterios propios de la investigación en ELE

Empleamos los siguientes patrones (Samper Hernández 2002, Sánchez-Saus 2016):

- Tratamiento de los errores: se registraron todas las entradas que se entendían con facilidad, salvo que se tratase de palabras españolas.
- Tratamiento de los préstamos: aceptamos los "neologismos universalizados" presentes en el DLE, tanto en la grafía adaptada fonológica y ortográficamente al español (*hámster*) cuanto en la grafía ori-

[24] Los paréntesis suponen que aparecieron ambas variantes de la misma unidad.

ginal (*pasta*); los extranjerismos utilizados con frecuencia por los hispanohablantes, aunque no se encuentren en el diccionario, por no haber otra forma de referirse a tales objetos o realidades. Excluimos, obviamente, las palabras de la LM o de otras LE conocidas.
– Tratamiento de las interferencias de otras lenguas: al encontrar entre las respuestas extranjerismos adaptados semántica o estructuralmente al español, se eliminaron.
– Unificación en la presentación de las unidades: se descartaron todos los artículos, adjetivos y adverbios que acompañaban las actualizaciones, a excepción de los que formaban unidades multipalabra.
– Amplitud de las relaciones asociativas: aceptamos las palabras activadas mediante relaciones secundarias, sobre todo en los campos léxicos considerados más complejos para un estudiante extranjero eliminando solo las asociaciones inexplicables.

### Criterios particulares por centros de interés

Reseñamos ahora las decisiones específicas adoptadas durante la edición de cada uno de los centros de interés a través de algunos ejemplos representativos:

– CI01, "partes del cuerpo": corregimos errores de acentuación o de elección falsa de consonante (**oido*, **musculo*, **naríz*, **tovillo*); registramos las formas plurales más frecuentes (*ojos*, *codos*, *nalgas*, *uñas*, *labios*, *dientes*, *muelas*); descartamos todos los extranjerismos (*mustache*), las palabras inventadas (**cavíl*, **piedes*) o adaptadas al español (**orillas*, posiblemente del francés *oreille*, **gamba*, **unghia*, del italiano).
– CI02, "la ropa": la corrección se concentró en los pocos casos de errores ortográficos (**cinturon*, **calzetines*, **camiceta*, **jersei*, **falta*). Eliminamos algunas formas inventadas (**jaqueta*, **escarfa*, posiblemente del inglés *scarf*) y los extranjerismos *boxer*, *t-shirt*, *slip*, *collant*, *sweater* (habríamos mantenido el vocablo españolizado *suéter*) a excepción de *bikini* (la forma es admitida en español, pese a la posibilidad de utilizar *biquini*).
– CI03, "partes de la casa (sin muebles)": suprimimos muchas respuestas, ya que detectamos un gran número de muebles o electrodomésticos que preferimos no admitir al ser objetos del campo siguiente. Además, quitamos de las listas los falsos amigos *tienda* y *taberna*, pero mantenimos *elevador* tratándose de una variante diatópica uti-

lizada en Cuba, El Salvador, Guatemala y México, como sinónimo de *ascensor*, palabra peninsular.

– CI04, "los muebles de la casa": corregimos algunas faltas de acentuación (**estanteria*, **lampara*, **frigorifico*); unificamos las entradas *mesa de noche*, *mesita de noche* y *mesilla de noche* bajo la forma *mesilla de noche*, la más utilizada y utilizamos los paréntesis para indicar la elisión de partes de una unidad: *(horno)microondas*, *tele(-visión)*.
– CI05, "alimentos y bebidas": mantuvimos los plurales de los vocablos que, aunque tienen su forma singular, aparecieron más frecuentemente en plural (*tapas*, *mejillones*, *churros*, *lentejas*, *cereales*). Los errores más frecuentes son ortográficos y no perjudican el conocimiento del vocabulario (**lemón*, **gaspacho*, **jamon*, **salsicha*) por lo que se corrigieron, mientras que eliminamos muchos extranjerismos (*riz*, *sushi*, *hamburger*, *cocktail*, *hot-dog*, *gin*, *spaghetti*), en particular porque en algunos casos tendrían su adaptación al español (*hamburguesa*, *cóctel*, *perrito-caliente*, *ginebra*, *espagueti*). En lo que se refiere a los préstamos italianos *pizza* y *pasta*, descartamos el primero por aparecer en cursiva en el DLE –lo que denota su naturaleza de barbarismo– mientras que aceptamos el segundo al estar totalmente incorporado en español.
– CI06, "objetos colocados en la mesa para la comida": se trata de un centro de escasa productividad de palabras, por lo que decidimos ampliar las posibilidades asociativas, aceptando acciones secundarias (*aceite*, *sal*, *pimienta*, *vinagre*, *desayuno*, *cambiar*, *poner*, *tomar*, *pulir*).
– CI07, "la cocina y sus utensilios": optamos por registrar la forma masculina de aquellos vocablos que se pueden utilizar en masculino o femenino al ser la forma más utilizada por los encuestados (*batidor*, *refrigerador*, *tostador*) a excepción de *licuadora*; quitamos de las listas los extranjerismos (*freezer*, *moka*, *pentola*, *mixer*) y las unidades inventadas (**cucino*).
– CI08, "la escuela: muebles y materiales": empleamos los paréntesis para indicar la elisión de partes de una unidad (*bolí(grafo)*, *profe(-sor)*); registramos los ítems que repetidamente aparecieron en plural (*tijeras*, *apuntes*, *deberes*, *sacapuntas*) y la forma masculina de los utensilios que figuraban tanto en masculino como en femenino (*computador*, *calculador*); descartamos los extranjerismos y las interferencias (*post-it*, **cantina* y **lavaña*, posiblemente del francés *cantine* y del italiano *lavagna*).

– CI09, "iluminación y calefacción": de nuevo, estamos ante un centro poco productivo que nos llevó a mantener las asociaciones secundarias. Recurrimos a los paréntesis para agrupar bajo un solo lema la palabra *termo(sifón)*, debido a la presencia de ambas formas.
– CI10, "la ciudad": mantuvimos las palabras plurales que contaban con más apariciones en esta forma (*correos* y *rascacielos*); utilizamos los paréntesis en *disco(teca)*, *auto(móvil)*, *bici(cleta)*, *(auto)bús*; suprimimos las formas inventadas y los extranjerismos. No fue registrada la multitud de topónimos, monumentos o calles, tanto si estaban escritos en español o en otras lenguas, por tratarse de nombres propios.
– CI11, "el campo": registramos casi todos los vocablos aportados, inclusas las palabras activadas por conexiones secundaría al ser un campo nocional que revela un carácter asociativo muy disperso. Descartamos los extranjerismos (*contadino*, *prato* y *arbre*).
– CI12, "medios de transporte": recurrimos frecuentemente a los paréntesis (*bici(cleta)*, *moto(cicleta)*, *(auto)bús*, *auto(móvil)*); corregimos los errores ortográficos (**pasejero*, **caravela*, **tandem*, **pié*) y suprimimos los extranjerismos *skateboard*, *camper*, *roulotte*, *scooter* (tendría su versión adaptada, *escúter*) o *jeep* (sería *todoterreno*) y el nombre propio *Über*.
– CI13, "trabajos del campo y del jardín": admitimos casi todas las respuestas, a excepción de los falsos amigos procedentes del italiano (**potar*, **potación*, **bonificar* o las formas inventadas **anafiar*, **seminar*), tratándose de un área poco productiva.
– CI14, "los animales": tras una simple revisión ortográfica eliminamos los extranjerismos (*aquila*, *pesce*, *elephant*) y respetamos los heterónimos *cabrón/cabra*, *toro/vaca*, *gallo/gallina*, *caballo/yegua*.
– CI15, "juegos y distracciones": quitamos los numerosos extranjerismos (*baseball*, *football*, *volleyball*, *ping pong*, *ballet*, *rugby*, *computer*, *sport*, *social network*, *sortir*), las palabras inventadas y las transferencias (**escondido*, **mosca ciega*, **luna parque*), las marcas comerciales poco presentes como *Cluedo* y *Taboo*, pero mantuvimos *Facebook*, *Play Station* y *Monopoly*. Aceptamos los vocablos adaptados al español *beisbol* y *voleibol* (en la variante sin acento siendo lo más frecuente en el corpus) y registramos la forma plural de algunos vocablos: *videojuegos*, *rompecabezas*, *cartas*, *damas*, *construcciones*, *dibujos animados*. De nuevo utilizamos los paréntesis al registrar *disco(teca)*, *pelí(cula)*, *tele(visión)*, *bici(cleta)*. Antes de terminar, conviene reseñar la edición de los sintagmas en vista

de la gran variedad aportada y, aún más, porque muchos aparecen por duplicado, la locución por una parte y el sustantivo aislado por otra: *leer un libro* y *libro*, *tocar un instrumento* e *instrumento musical*, *jugar al fútbol* y *fútbol*. Decidimos registrar sustantivos y verbos como dos entradas distintas: *leer*, *libro*, *tocar*, *instrumento musical*, *jugar*, *fútbol*. No es el caso de *escuchar música*, ya que encontramos esta forma fija muchas veces, así que la mantuvimos.

– CI16, "profesiones y trabajos": registramos la forma masculina de las profesiones, salvo en aquellos casos donde solo se haya actualizado solamente la femenina (*ama de casa*, *señora de la limpieza*) manteniendo los heterónimos (*actor*/*actriz*, *azafata*, *preste*, *cura*, *monja*); corregimos las faltas ortográficas (**idráulico*, **hefe*, **avocado*) y suprimimos los extranjerismos, falsos amigos y palabras inventadas (*hostess*, *influencer*, *autista*, **judice*, *regista*, **impiegado*, **infermer*).

## 2.5 Análisis de la riqueza léxica

Como en el apartado anterior, tras la elicitación, edición y tratamiento informático de los datos, llevamos a cabo el análisis en dos momentos distintos y complementarios. Primero, el estudio cuantitativo establece el grado de riqueza léxica producido en los relatos según los indicadores de variedad y densidad. Segundo, para profundizar en el estudio de los datos, proponemos un acercamiento a un análisis de tipo cualitativo del corpus general.

### 2.5.1 Los índices de riqueza léxica

La riqueza léxica permite estudiar las palabras en contexto midiendo el número de palabras y vocablos en un texto y su densidad, esto es, el empleo de unidades funcionales del discurso frente a las nocionales. Para ello, contamos con un conjunto de indicadores que concurren a su medición: la variación léxica, la sofisticación léxica, la densidad léxica y el número de errores (Read 2000). Hay que aplicar diferentes índices para obtener datos completos (Jarvis 2013) siguiendo la propuesta de Capsada y Torruella (2017: 397) según la cual es necesaria la aplicación de un «eclecticismo como una forma de síntesis»:

> no escoger un único índice como el paladín de la riqueza léxica, sino, por el contrario, tratar de utilizar la información que nos pueden aportar todos estos cinco índices que han demostrado que presentan un buen comportamiento en sus medidas. De esta manera trabajaremos con más

información y, además, el posible mal comportamiento de los índices en algunos textos, como ya hemos indicado que sucede, podrá quedar compensado.

La elección de los índices se suele establecer en función de los participantes en un estudio, Laufer y Nation (1995) propusieron el empleo de medidas distintas según se trabaje con informantes pertenecientes a grupos heterogéneos, especialmente de un contexto educacional diferente, pero no es nuestro caso. El proceso ha sido sencillo, ya que trabajamos con alumnos procedentes del mismo entorno educativo, con lo cual aplicamos sin dificultad las medidas más útiles para la ejecución del estudio. Descartamos los indicadores que implican errores por ser demasiado subjetivos y dependientes del juicio del investigador (Laufer y Nation 1995: 310-313) como los que calculan la variación semántica o la calidad léxica: habría sido difícil encontrar palabras sofisticadas dado el nivel de ELE de los informantes y el tema del relato. Igualmente, no empleamos el índice de la originalidad léxica porque se basa en la cantidad de *tokens* de los textos y no habría tenido sentido en nuestro análisis, ya que tienen el mismo tamaño. A este propósito, tampoco aplicamos técnicas de normalización de las fórmulas como las de Zipf (1935) o Yule (1944). Con todo eso y en base a la bibliografía revisada, calculamos:

– la variación léxica contabiliza el grado de diversidad del vocabulario indicando en tantos por ciento la capacidad de expresarse del autor de una redacción:

$$\frac{\text{número de } \textit{types} \text{ x } 100}{\text{número total de } \textit{tokens}}$$

– la *Type/Token Ratio* (TTR) es otra forma de computar la variación léxica describiendo la relación entre *types* y *tokens* de un texto:[25]

$$\frac{\text{número de } \textit{types}}{\text{número } \textit{tokens}}$$

[25] Este índice puede variar de 0 a 1. El valor máximo se obtiene en el caso de que no se repita ninguna palabra, pero es algo imposible en un texto que no sea una oración. De igual manera, es improbable alcanzar el 0 ya que siempre se puede contar con un *type* como mínimo en una muestra (al contrario, no habría ningún texto). De ahí que el valor ideal de TTR resulte el que más se acerca a 1.

– el índice de hápax detecta las palabras empleadas solo una vez en un texto (los denominados *hápax*) y concurre a establecer la variación léxica:[26]

$$\frac{\text{número total de } \textit{types}}{\text{suma de hápax}}$$

– la densidad léxica calcula el porcentaje de unidades nocionales en proporción al número total de *tokens* que componen un texto:

$$\frac{\text{número de } \textit{tokens} \text{ léxicos x } 100}{\text{número total de } \textit{tokens}}$$

– el intervalo de aparición de palabras nocionales (IAT) determina el intervalo en el que aparecen las palabras de contenido semántico con respecto a las unidades funcionales:[27]

$$\frac{\text{número total de } \textit{tokens}}{\text{número de } \textit{tokens} \text{ léxicos}}$$

Incluimos, además, el estudio de los descriptivos estadísticos para examinar el comportamiento intragrupal e intergrupal de los participantes en función de cada variante de las variables analizadas y del momento de suministración de la prueba.

Es preciso tener en cuenta que, si bien estos datos sean numéricos, el análisis cuantitativo aporta también informaciones cualitativas sobre la competencia léxica (López Morales 2011: 15):

> La llamada 'calidad de la escritura' está integrada por una serie de factores, entre los que destacan sin duda: la riqueza léxica, la madurez sintáctica, los esquemas de cohesión y la coherencia discursiva. La amplitud y variedad del vocabulario está muy apoyado en la disponibilidad léxica del hablante, la madurez sintáctica, en su grado de entrenamiento combinatorial de oraciones simples en el discurso, los esquemas de cohesión y la coherencia discusiva dependen esencialmente del 'orden' que quiera dársele conscientemente a los elementos constitutivos del discurso.

[26] Cuanto más pequeño es el índice, más rico es el vocabulario ya que la riqueza léxica es inversamente proporcional.

[27] Cuanto más se acerca a 1, mayor resulta el uso de palabras temáticas y, por consiguiente, la riqueza léxica (el 1 no es alcanzable porque significaría que todos los *tokens* de un texto fuesen palabras nocionales).

### 2.5.2 Tratamiento informático de los datos

El análisis cuantitativo se desarrolla con *AntConc* (Anthony 2018), un programa que permite extraer datos de amplias muestras textuales. Los textos, en formato *.txt*, mantienen el mismo código identificativo para cada informante que hemos explicado antes. Extrajimos del corpus las listas de frecuencia y el conteo de *types* y *tokens* para poder calcular los índices. Asimismo, con el fin de observar la densidad léxica utilizamos la función *Stoplist* para que el programa filtrase una lista de elementos que queríamos eliminar de los cómputos (las palabras funcionales) para averiguar el porcentaje de unidades nocionales sobre el total de palabras. A continuación, el análisis cualitativo se ejecuta con *Sketch Engine*, un recurso informático que combina estadísticas con criterios lingüísticos y permite realizar ulteriores búsquedas con respecto a *AntConc*. Tras haber cargado los textos en el *software* utilizamos las herramientas *Wordlist* para la recopilación de las listas de frecuencia; *Keywords* para la extracción de las palabras clave del corpus; *Word Sketch* para observar el uso de algunas de estas palabras clave mediante un sistema de interrogación de corpus que averigua el comportamiento colocacional de un lexema.

### 2.5.3 La edición de los datos

Antes de la informatización dedicamos una importante parte del proceso analítico a la edición de los datos que supone el tratamiento de los errores y la normalización de las variantes flexivas. De nuevo, establecimos algunos criterios siguiendo las prácticas convencionales que se emplean en la bibliografía sobre el análisis de la producción escrita en lengua extranjera.

#### Tratamiento de los errores

Rastreamos dos tipologías de error: las faltas (ortográficas, fonológicas, morfosintácticas), que no perjudican el conocimiento del vocabulario español, que se toleran y se corrigen,[28] y los errores léxicos propiamente dichos, que descartamos. De entre las faltas, considerado el nivel de los aprendientes, ignoramos los errores de ortografía: corregimos la

[28] Corregimos los errores, en la medida de lo posible, dejando para futuras investigaciones el análisis de su incidencia en los valores de riqueza léxica entre el texto original y la versión editada ya que se podría comprobar si la inclusión o exclusión en los cómputos contribuye a mejorar los datos de manera significativa o no.

mala colocación de tilde (**azúl*); la incorrecta duplicación de consonantes (**professor*); la selección falsa de consonantes (**pizo*) con la excepción de las unidades que, alejándose demasiado de la forma correcta, denotan un claro desconocimiento de la palabra.

**Tratamiento de la morfología derivativa, formas plenas y acortamientos**

Unificamos las variantes flexivas de los *tokens* bajo su forma no marcada, mediante la recopilación de un solo lema por cada palabra. En lo que atañe a los verbos, los participios se tratan como lemas independientes de manera que su uso con función de adjetivo no se incluya entre los predicados. Asimismo, dejamos invariadas tanto las formas léxicas derivadas sustantivas y adjetivas –apreciativas, diminutivas, aumentativas, superlativas, despectivas (*famosísimo*, *muchísimo*, *mercadillo*, *poquito*, *pueblecito*)– como los acortamientos léxicos (*bici*, *moto*), puesto que contribuyen a cuantificar el caudal activo del vocabulario, además de testimoniar la capacidad de formación de las palabras.

**Tratamiento de préstamos e interferencias de otras lenguas**

Eliminamos los extranjerismos y los préstamos de la LM y de otras LE (*band*, *b&b*, *camper*, *check-in*, *cous-cous*, *jeep*, *pullman*, *quad*, *souvenir*, *tour*); las interferencias, formas híbridas o inventadas, aunque basadas en vocablos existentes en otras lenguas, al no constituir léxico útil para la investigación (**gita*, **dipinto*, **posto*, **luego*[29]).

**Tratamiento de los nombres propios, palabras nocionales y palabras funcionales**

Descartamos los nombres propios por no formar parte del vocabulario de una lengua, mientras que optamos por incluir los topónimos en español, ya que muchos de los que estaban presentes en el muestreo se suelen enseñar, al contrario, excluimos aquellos escritos en italiano o en otro idioma (entre otros, *Firenze*, *Ginevra*, *New York*).

En cuanto a la distinción entre palabras nocionales y funcionales, elaboramos una lista de parada (*Stoplist*) *ad hoc* basada en la propuesta de Hallebeek (1986: 210-215) para una mayor libertad de respuesta y una

[29] Sería en este caso un italianismo, de *luogo*.

mayor aceptación de palabras, considerado nivel de los informantes. La *Stoplist* está formada por las siguientes categorías: preposiciones, conjunciones, artículos, pronombres personales y posesivos, demostrativos, indefinidos, relativos, interrogativos y exclamativos, numerales e interjecciones.

# Capítulo 3.
# Análisis de la disponibilidad léxica

## 3.1 Estudio cuantitativo

La parte inicial del capítulo presenta el estudio cuantitativo de los datos procedentes de las pruebas de disponibilidad léxica suministradas a los alumnos-informantes. Abre el análisis transversal, realizado a partir de los resultados generales de la primera suministración de la prueba en función del número de palabras y vocablos actualizados, los promedios de respuestas, los índices de cohesión y de densidad léxica. Seguidamente, nos centramos en la incidencia de las variables según el mismo proceso analítico.

A continuación, introducimos una novedad con respecto a otros estudios publicados hasta la fecha: el análisis longitudinal que hemos llevado a cabo gracias a la segunda suministración de las encuestas a los participantes de nivel B2, para examinar la evolución de los conocimientos tras la asistencia a un curso académico de un año.

En la última sección, cotejamos los resultados con los de otras investigaciones realizadas en el mismo ámbito para averiguar si hay puntos comunes y cuáles son las diferencias en el desarrollo del bagaje léxico de los aprendientes de ELE.

### 3.1.1 Análisis transversal: resultados generales

Tras el proceso de edición de las respuestas, la digitalización mediante *Dispogen* presenta un total de 12.579 palabras y 1.490 vocablos aportados por 100 informantes en la primera prueba realizada a comienzo del año académico.

### Número de palabras

La tabla muestra la repartición de los datos generales por CI, incluyendo el número total de palabras que los componen, el promedio por informante y el rango de cada uno, es decir, los campos ordenados según productividad.

| Rango | CI | Palabras | Promedio/ informante |
|---|---|---|---|
| 2 | CI01.CUE | 1.341 | 13,41 |
| 8 | CI02.ROP | 791 | 7,91 |
| 12 | CI03.CAS | 610 | 6,10 |
| 11 | CI04.MUE | 660 | 6,60 |
| 1 | CI05.ALI | 1.463 | 14,63 |
| 13 | CI06.MES | 530 | 5,30 |
| 14 | CI07.COC | 365 | 3,65 |
| 7 | CI08.ESC | 803 | 8,03 |
| 15 | CI09.ILU | 342 | 3,42 |
| 3 | CI10.CIU | 1.171 | 11,71 |
| 10 | CI11.CAM | 685 | 6,85 |
| 6 | CI12.TRA | 808 | 8,08 |
| 16 | CI13.TRC | 252 | 2,52 |
| 4 | CI14.ANI | 1.080 | 10,80 |
| 9 | CI15.JUE | 784 | 7,84 |
| 5 | CI16.PRO | 894 | 8,94 |
| Total | | 12.579 | 7,86 |

Tabla 6. Número total de palabras y media por informante.

Destacan tres franjas en las cuales dividimos los CI según productividad:

– un grupo no alcanza las 400 unidades léxicas (CI07, “la cocina y sus utensilios”; CI09, “iluminación y calefacción”; CI13, “trabajos del campo y del jardín”);
– un conjunto no llega al promedio de 786 ítems (CI15, “juegos y distracciones”; CI11, “el campo”; CI04, “los muebles de la casa”; CI03, “partes de la casa (sin muebles)”; CI06, “objetos colocados en la mesa para la comida”);
– otros sobrepasan el promedio (CI16, “profesiones y oficios”; CI12, “medios de transporte”; CI08, “la escuela: muebles y materiales”; CI02, “la ropa”), entre los que hay cuatro que superan los 1.000 lexemas (CI05, “alimentos y bebidas”; CI01, “partes del cuerpo”; CI10, “la ciudad”; CI14, “los animales”).

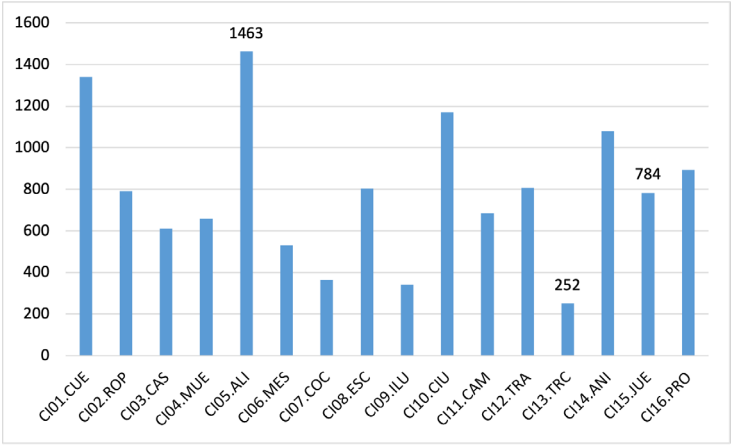


Gráfico 1. Número total de palabras por CI.

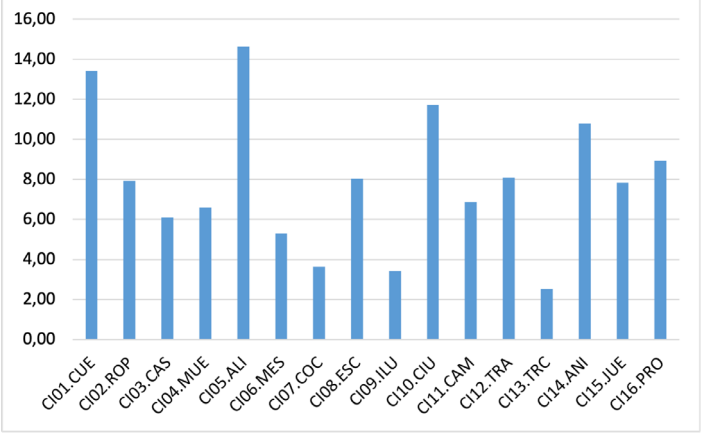


Gráfico 2. Promedio de palabras por informante.

En lo que atañe al promedio de palabras por encuestado, los centros se ordenan como enseña la tabla 7 en función de su mayor o menor productividad (por rango). Si tomamos como referencia la media general de 7,86 palabras por informante, vemos cuáles se sitúan por encima y por debajo de este valor.

| Rango | CI | Palabras | Promedio/ informante |
|---|---|---|---|
| 1 | CI05.ALI | 1.463 | 14,63 |
| 2 | CI01.CUE | 1.341 | 13,41 |
| 3 | CI10.CIU | 1.171 | 11,71 |
| 4 | CI14.ANI | 1.080 | 10,80 |
| 5 | CI16.PRO | 894 | 8,94 |
| 6 | CI12.TRA | 808 | 8,08 |
| 7 | CI08.ESC | 803 | 8,03 |
| 8 | CI02.ROP | 791 | 7,91 |
| 9 | CI15.JUE | 784 | 7,84 |
| 10 | CI11.CAM | 685 | 6,85 |
| 11 | CI04.MUE | 660 | 6,60 |
| 12 | CI03.CAS | 610 | 6,10 |
| 13 | CI06.MES | 530 | 5,30 |
| 14 | CI07.COC | 365 | 3,65 |
| 15 | CI09.ILU | 342 | 3,42 |
| 16 | CI13.TRC | 252 | 2,52 |

Tabla 7. CI ordenados según productividad.

Ocho centros de interés rebasan la media, casi duplicándola, mientras que los que se sitúan por debajo presentan un rendimiento muy inferior, en línea con lo observado por Samper Padilla, Bellón Fernández y Samper Hernández (2003) que insisten en la poca rentabilidad de ciertas temáticas lejanas de los intereses de los jóvenes y a menudo no trabajadas ni en aula ni en los materiales de ELE.

El centro más productivo es el CI05, "alimentos y bebidas", que ofrece una media de 14,63 palabras por informante, seguido por el CI01, "partes del cuerpo"; el CI10, "la ciudad"; el CI14, "los animales", que superan las diez unidades. Al contrario, los campos menos rentables, que no consiguen superar las cuatro palabras por encuestado, son el CI07, "la cocina y sus utensilios"; el CI09 "iluminación y calefacción; el CI13, "trabajos del campo y del jardín".

### Número de vocablos

Analizamos ahora la variedad léxica de los campos semánticos, esto es, el número de vocablos que los informantes han activado en los listados, sin contar las repeticiones, para averiguar si existe una distinción entre el número de palabras y el número de vocablos aportados y qué tipo de relación se establece entre estos dos índices.[30]

En total disponemos de 1.490 palabras diferentes que se reparten como sigue en los CI:

| Rango | CI | Vocablos | Promedio/ informante |
|---|---|---|---|
| 7 | CI01.CUE | 91 | 0,91 |
| 8 | CI02.ROP | 72 | 0,72 |
| 15 | CI03.CAS | 47 | 0,47 |
| 14 | CI04.MUE | 49 | 0,49 |
| 1 | CI05.ALI | 175 | 1,75 |
| 16 | CI06.MES | 43 | 0,43 |
| 11 | CI07.COC | 66 | 0,66 |
| 9 | CI08.ESC | 72 | 0,72 |
| 13 | CI09.ILU | 57 | 0,57 |
| 3 | CI10.CIU | 157 | 1,57 |

[30] Las cifras no son completamente reales puesto que habría habido que quitar los vocablos compartidos por distintos centros de interés, pero *Dispogen* no permite realizar esta distinción, con lo cual trabajamos con todos los datos.

| | | | |
|---|---|---|---|
| 4 | CI11.CAM | 141 | 1,41 |
| 12 | CI12.TRA | 58 | 0,58 |
| 10 | CI13.TRC | 67 | 0,67 |
| 6 | CI14.ANI | 102 | 1,02 |
| 5 | CI15.JUE | 131 | 1,31 |
| 2 | CI16.PRO | 162 | 1,62 |
| Total | | 1.490 | 0,93 |

Tabla 8. Número total de vocablos.

Seis campos presentan cantidades mayores de vocablos, superando el promedio de 93 (CI05, "alimentos y bebidas"; CI16, "profesiones y oficios", CI10, "la ciudad"; CI11, "el campo"; CI15, "juegos y distracciones"; CI14, "los animales"). Al observar este conjunto sobresale una gran diferencia con respecto al tema más pobre, el CI06, "objetos colocados en la mesa para la comida", que revela una media de 43. La reproducción gráfica desglosa la partición de los vocablos mostrando que es menos equilibrada con respecto a la del total de palabras.

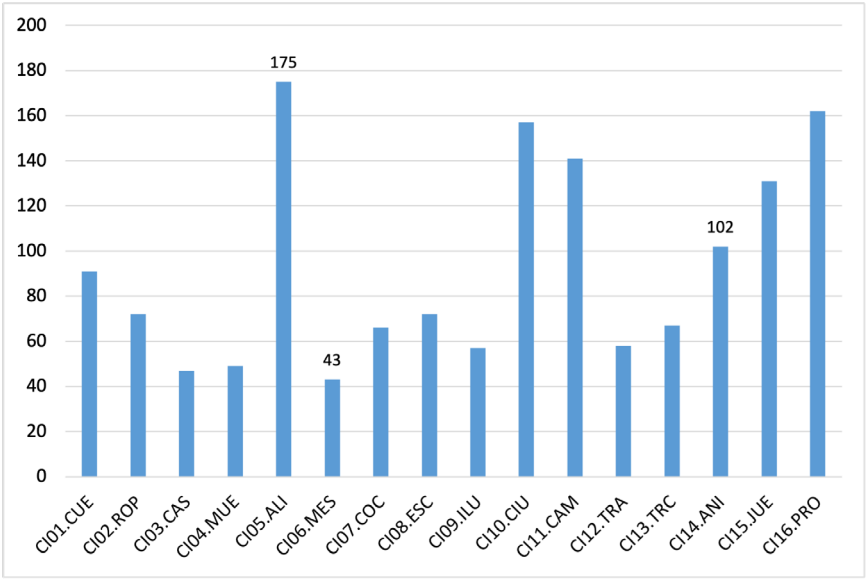


Gráfico 3. Número total de vocablos por CI.

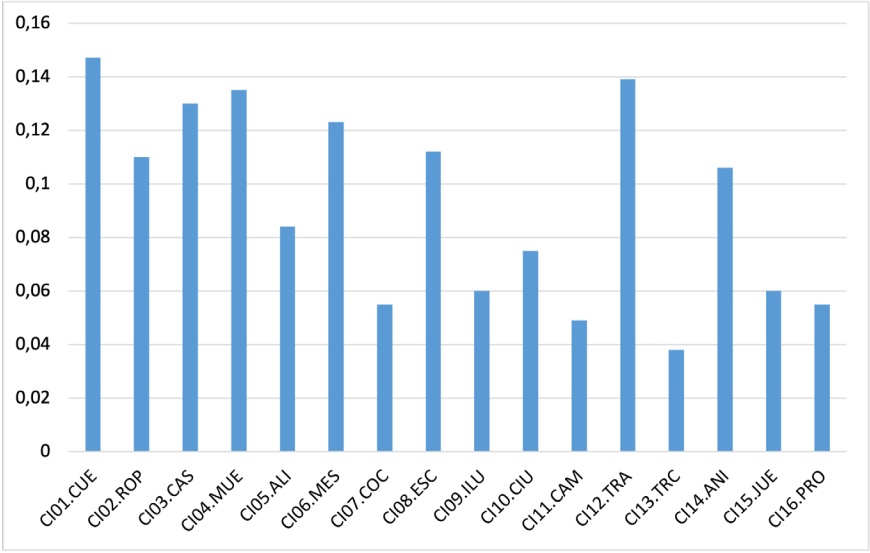


Gráfico 4. Promedio de vocablos por informante.

Los campos se ordenan según el rango de variedad léxica como se muestra a continuación:

| Rango | CI | Vocablos | Promedio/ informante |
|---|---|---|---|
| 1 | CI05.ALI | 175 | 1,75 |
| 2 | CI16.PRO | 162 | 1,62 |
| 3 | CI10.CIU | 157 | 1,57 |
| 4 | CI11.CAM | 141 | 1,41 |
| 5 | CI15.JUE | 131 | 1,31 |
| 6 | CI14.ANI | 102 | 1,02 |
| 7 | CI01.CUE | 91 | 0,91 |
| 8 | CI02.ROP | 72 | 0,72 |
| 9 | CI08.ESC | 72 | 0,72 |
| 10 | CI13.TRC | 67 | 0,67 |
| 11 | CI07.COC | 66 | 0,66 |
| 12 | CI12.TRA | 58 | 0,58 |

| | | | |
|---|---|---|---|
| 13 | CI09.ILU | 57 | 0,57 |
| 14 | CI04.MUE | 49 | 0,49 |
| 15 | CI03.CAS | 47 | 0,47 |
| 16 | CI06.MES | 43 | 0,43 |

Tabla 9. CI ordenados según variedad.

Parece no existir una relación entre productividad y riqueza léxica, ya que los centros más productivos no son necesariamente lo más ricos.

### Índice de cohesión y densidad léxica

El estudio de la relación entre palabras y vocablos permite examinar el grado de cohesión semántica de los centros de interés y obtener ulterior información sobre la variedad de las respuestas.

| CI | Índice de cohesión | Densidad léxica |
|---|---|---|
| CI01.CUE | 0,147 | 14,74 |
| CI02.ROP | 0,110 | 10,99 |
| CI03.CAS | 0,130 | 12,98 |
| CI04.MUE | 0,135 | 13,47 |
| CI05.ALI | 0,084 | 8,36 |
| CI06.MES | 0,123 | 12,33 |
| CI07.COC | 0,055 | 5,53 |
| CI08.ESC | 0,112 | 11,15 |
| CI09.ILU | 0,060 | 6 |
| CI10.CIU | 0,075 | 7,46 |
| CI11.CAM | 0,049 | 4,86 |
| CI12.TRA | 0,139 | 13,93 |
| CI13.TRC | 0,038 | 3,76 |
| CI14.ANI | 0,106 | 10,59 |

| | | |
|---|---|---|
| CI15.JUE | 0,060 | 5,98 |
| CI16.PRO | 0,055 | 5,52 |
| Promedio | 0,092 | 9,23 |

Tabla 10. Índice de cohesión y densidad léxica por CI.

Ocho áreas superan el promedio y resultan las más cerradas y compactas, en particular el CI01, "partes del cuerpo", alcanza la mayor concreción semántica del corpus y, a la vez, es uno de los temas más rentables en absoluto, lo cual destaca la relación que a menudo se establece entre productividad, riqueza y sus respectivas cohesión y densidad (0,147 y 14,74). Al contrario, el CI11, "el campo", y el CI13, "trabajos del campo y del jardín", son los más abiertos (respectivamente 0,049 y 4,86; 0,038 y 3,76).

| Rango | CI | Índice de cohesión | Densidad léxica |
|---|---|---|---|
| 1 | CI01.CUE | 0,147 | 14,74 |
| 2 | CI12.TRA | 0,139 | 13,93 |
| 3 | CI04.MUE | 0,135 | 13,47 |
| 4 | CI03.CAS | 0,130 | 12,98 |
| 5 | CI06.MES | 0,123 | 12,33 |
| 6 | CI08.ESC | 0,112 | 11,15 |
| 7 | CI02.ROP | 0,110 | 10,99 |
| 8 | CI14.ANI | 0,106 | 10,59 |
| 9 | CI05.ALI | 0,084 | 8,36 |
| 10 | CI10.CIU | 0,075 | 7,46 |
| 11 | CI09.ILU | 0,060 | 6 |
| 12 | CI15.JUE | 0,060 | 5,98 |
| 13 | CI07.COC | 0,055 | 5,53 |
| 14 | CI16.PRO | 0,055 | 5,52 |
| 15 | CI11.CAM | 0,049 | 4,86 |
| 16 | CI13.TRC | 0,038 | 3,76 |

Tabla 11. CI ordenados según cohesión y densidad.

Gráficamente:

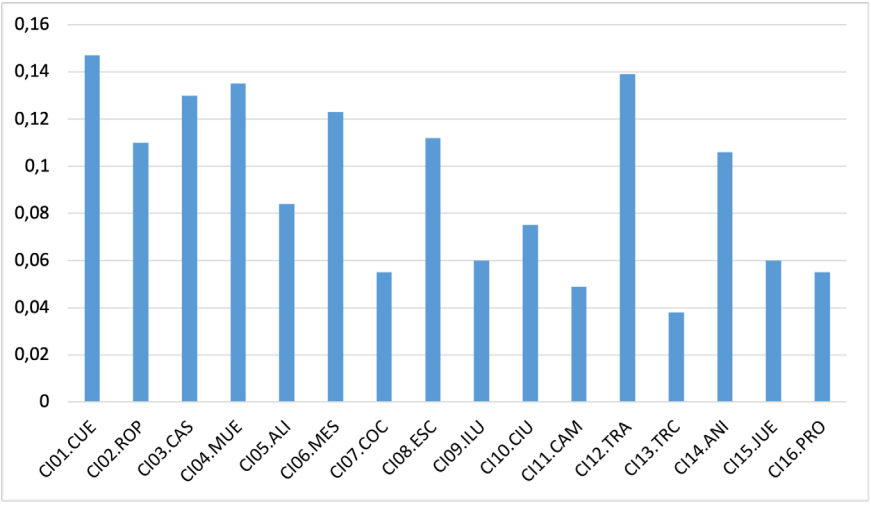


Gráfico 5. Índice de cohesión por CI.

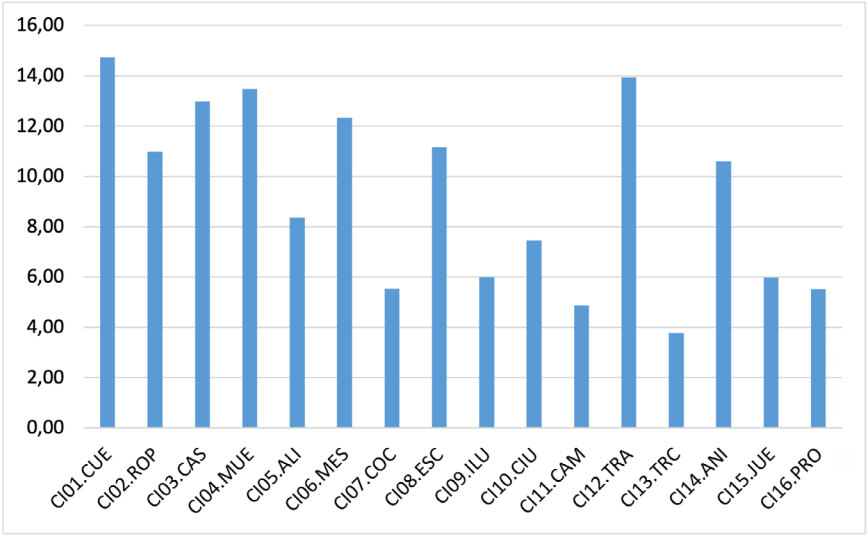


Gráfico 6. Densidad léxica por CI.

Según la clasificación de Gómez Molina y Gómez Devís (2004: 83) los CI que tienen un IC igual o superior a 0,06 son semánticamente muy compactos y tienen una buena asociación conceptual; los que se acercan

al 0,05 presentan un grado medio de coincidencia de las respuestas; los que tienen un índice menor de 0,04 son más difusos.

Antes de terminar, es interesante comparar estos datos con los demás índices analizados hasta ahora. La tabla 12 muestra el modo en el que la cohesión y la densidad léxica influyen en la relación entre productividad y variedad poniendo de relieve la coincidencia o la divergencia de las respuestas de los informantes.[31]

| CI | P | R | V | R | D | IC | R |
|---|---|---|---|---|---|---|---|
| CI01.CUE | 91 | 7 | 1.341 | 2 | 14,74 | 0,147 | 1 |
| CI02.ROP | 72 | 8 | 791 | 8 | 10,99 | 0,110 | 7 |
| CI03.CAS | 47 | 15 | 610 | 12 | 12,98 | 0,130 | 4 |
| CI04.MUE | 49 | 14 | 660 | 11 | 13,47 | 0,135 | 3 |
| CI05.ALI | 175 | 1 | 1.463 | 1 | 8,36 | 0,084 | 9 |
| CI06.MES | 43 | 16 | 530 | 13 | 12,33 | 0,123 | 5 |
| CI07.COC | 66 | 11 | 365 | 14 | 5,53 | 0,055 | 13 |
| CI08.ESC | 72 | 9 | 803 | 7 | 11,15 | 0,112 | 6 |
| CI09.ILU | 57 | 13 | 342 | 15 | 6 | 0,060 | 11 |
| CI10.CIU | 157 | 3 | 1.171 | 3 | 7,46 | 0,075 | 10 |
| CI11.CAM | 141 | 4 | 685 | 10 | 4,86 | 0,049 | 15 |
| CI12.TRA | 58 | 12 | 808 | 6 | 13,93 | 0,139 | 2 |
| CI13.TRC | 67 | 10 | 252 | 16 | 3,76 | 0,038 | 16 |
| CI14.ANI | 102 | 6 | 1.080 | 4 | 10,59 | 0,106 | 8 |
| CI15.JUE | 131 | 5 | 784 | 9 | 5,98 | 0,060 | 12 |
| CI16.PRO | 162 | 2 | 894 | 5 | 5,52 | 0,055 | 14 |

Tabla 12. Recapitulación datos de DL.

El CI11, "el campo", es uno de los más ricos y no resulta muy productivo (rango 10) al situarse en el rango 15 de cohesión, el cual indica que hay pocas repeticiones y, en consecuencia, es abierto. El CI16, "profesiones y oficios", pese a ofrecer una gran cantidad de palabras, es otro de los centros

[31] Donde: P= palabras; R= rango; V= vocablos; D= densidad léxica; IC= índice de cohesión.

más variados (rango 2) que alcanza un bajo grado de cohesión. En el caso opuesto, el CI01, "partes del cuerpo", es uno de los más productivos (rango 2) y compactos a la vez, ya que su densidad presenta el valor más alto de todo el corpus (rango 1), pero no arroja una gran cantidad de vocablos (rango 7). Lo mismo ocurre en el CI12, "los medios de transporte", que pasa del rango 6 de productividad al 2 de riqueza: cohesión y densidad son elevados (rango 2).

Es esencial tener en cuenta el hecho de que estos datos dependen de la productividad y de la riqueza de los centros. El CI05, "alimentos y bebidas", y el CI10, "la ciudad", son simultáneamente muy productivos y ricos: el total de vocablos no incide en los índices de cohesión y densidad. De manera parecida, el CI13, "trabajos del campo y del jardín", se coloca en posiciones bajas en todos los índices: parece que los informantes aportan pocas palabras que difieren notablemente y no hay una media alta de repeticiones, posiblemente porque no es un tema trabajado en clase. Además, algunos centros (CI03, "partes de la casa (sin muebles)"; CI04, "los muebles de la casa"; CI06, "objetos colocados en la mesa para la comida") no entran ni entre los más productivos ni entre los más ricos, pero la escasez de vocablos permite un alto grado de densidad, quizá debido a un aprendizaje sistemático, que lleva a los estudiantes a escribir las mismas lexías (Gómez Molina y Gómez Devís 2004: 76-77):

> Se corrobora así lo que han observado otras investigaciones, sea en ELE sea en ELM, esto es, que '[…] la mayor o menor riqueza de vocablos no se corresponde con la productividad de unidades léxicas en las diferentes áreas temáticas; es decir, que no se cumple la máxima: a mayor número de respuestas, mayor número de palabras diferentes.'

## Índice de disponibilidad, frecuencia de aparición y frecuencia acumulada

Este ulterior recuento de los datos sirve para llevar a cabo el estudio cualitativo que ocupará la próxima parte del capítulo. Mediante *Dispogen* hemos extraído del corpus los vocablos más disponibles en cada centro de interés, esto es, los que alcanzan un ID igual o superior a 0,1 (ID ≥ 0,1). Complementan los datos los porcentajes de la frecuencia de aparición ≥ 30% y de la frecuencia acumulada ≤ 75%. Sin embargo, dejamos fuera las unidades que, aunque hayan sido actualizadas, no se revelan igualmente disponibles. Los vocablos totales con ID ≥ 0,1 son 176, cuyo número por cada campo semántico oscila de un mínimo de 3 a un máximo de 25:

| CI | ID ≥ 0,1 | Frecuencia de aparición ≥ 30% | Frecuencia acumulada ≤ 75 % |
|---|---|---|---|
| CI01.CUE | 19 | 16 | 17 |
| CI02.ROP | 12 | 9 | 15 |
| CI03.CAS | 7 | 7 | 11 |
| CI04.MUE | 10 | 5 | 12 |
| CI05.ALI | 25 | 13 | 39 |
| CI06.MES | 6 | 6 | 6 |
| CI07.COC | 8 | 1 | 15 |
| CI08.ESC | 11 | 9 | 13 |
| CI09.ILU | 3 | 3 | 13 |
| CI10.CIU | 17 | 11 | 36 |
| CI11.CAM | 9 | 3 | 35 |
| CI12.TRA | 10 | 9 | 8 |
| CI13.TRC | 4 | 1 | 18 |
| CI14.ANI | 12 | 8 | 22 |
| CI15.JUE | 13 | 2 | 31 |
| CI16.PRO | 10 | 3 | 41 |
| Total | 176 | 106 | 332 |

Tabla 13. Número de vocablos más disponibles.

No siempre hay una relación proporcional entre la riqueza léxica y la cantidad de vocablos más disponibles, ya que, por ejemplo, el CI09, "iluminación y calefacción", tiene el menor número de vocablos más disponibles, pero no es el más pobre en absoluto. El CI01, "partes del cuerpo", aporta 19 vocablos que tienen un ID ≥ 0,1 (rango 2) pero no es uno de los centros más rico (rango 7).

Al observar los datos relacionados con la frecuencia de aparición, la tendencia cambia: el CI que llega a valores más altos es el CI01, "partes del cuerpo", mientras que el CI07, "la cocina y sus utensilios", y el CI13, "trabajos del campo y del jardín", aportan las cantidades menores de palabras diferentes. Los vocablos más frecuentes disminuyen del 66% con respecto a los más disponibles pasando a ser 106 en total.

La frecuencia acumulada modifica, otra vez, la distribución de los resultados: el CI16, "profesiones y oficios", ofrece el número más elevado de vocablos y el CI06, "objetos colocados en la mesa para la comida", presenta el menor. Por lo general, este parámetro registra un incremento de los vocablos en todos los centros, hasta llegar al +89% en relación con el total de los ID y triplicar el total de la frecuencia de aparición.

### 3.1.2 Análisis transversal: resultados por variable

En este apartado indagamos cómo varían los índices en función de las tres variables sociolingüística contempladas. De nuevo, nos fijamos en el total de palabras y vocablos, la media por informante, el índice de cohesión y la densidad léxica.

#### Variable sexo

Contamos con una mayor presencia de mujeres, que eran 87 y han aportado un total de 11.070 palabras y 1.389 vocablos (media: 127,24 y 15,97), y una menor participación de hombres que componían un grupo más restringido de 13 sujetos que han escrito 1.509 palabras y 617 vocablos (media: 116,08 y 47,46).

Como tenemos en cuenta la gran diferencia en la constitución de las agrupaciones, trabajamos con los datos promediales para realizar un análisis fiable. De todos modos, las tablas 14 y 15 muestran los datos en bruto por cada variante, aunque no los utilizamos en la fase de cotejo, y las medias que sí, permiten verificar la influencia del factor.[32]

| ***Mujer*** | | | | | | |
|---|---|---|---|---|---|---|
| **CI** | **P** | **R** | **Pp/I** | **V** | **R** | **Pv/I** |
| CI01.CUE | 1.167 | 2 | 13,41 | 83 | 7 | 0,95 |
| CI02.ROP | 699 | 8 | 8,03 | 67 | 9 | 0,77 |
| CI03.CAS | 532 | 12 | 6,11 | 46 | 14 | 0,53 |
| CI04.MUE | 590 | 11 | 6,78 | 46 | 15 | 0,53 |
| CI05.ALI | 1.273 | 1 | 14,63 | 162 | 1 | 1,86 |

[32] Donde: Pp/I: promedio de palabras por informante; Pv/I: promedio de vocablos por informante.

| | | | | | | |
|---|---|---|---|---|---|---|
| CI06.MES | 456 | 13 | 5,24 | 39 | 16 | 0,45 |
| CI07.COC | 317 | 14 | 3,64 | 65 | 10 | 0,75 |
| CI08.ESC | 722 | 6 | 8,30 | 71 | 8 | 0,82 |
| CI09.ILU | 307 | 15 | 3,53 | 56 | 12 | 0,64 |
| CI10.CIU | 1.056 | 3 | 12,14 | 152 | 2 | 1,75 |
| CI11.CAM | 600 | 10 | 6,90 | 123 | 4 | 1,41 |
| CI12.TRA | 700 | 7 | 8,05 | 53 | 13 | 0,61 |
| CI13.TRC | 229 | 16 | 2,63 | 65 | 11 | 0,75 |
| CI14.ANI | 954 | 4 | 10,97 | 92 | 6 | 1,06 |
| CI15.JUE | 685 | 9 | 7,87 | 120 | 5 | 1,38 |
| CI16.PRO | 783 | 5 | 9 | 149 | 3 | 1,71 |
| Total | 11.07 | | 7,95 | 1.389 | | 1 |

Tabla 14. Índices de DL según la variante mujer.

| ***Hombre*** | | | | | | |
|---|---|---|---|---|---|---|
| **CI** | **P** | **R** | **Pp/I** | **V** | **R** | **Pv/I** |
| CI01.CUE | 174 | 2 | 13,38 | 58 | 3 | 4,46 |
| CI02.ROP | 92 | 8 | 7,08 | 33 | 8 | 2,54 |
| CI03.CAS | 78 | 11 | 6 | 26 | 10 | 2 |
| CI04.MUE | 70 | 13 | 5,38 | 27 | 9 | 2,08 |
| CI05.ALI | 190 | 1 | 14,62 | 86 | 1 | 6,62 |
| CI06.MES | 74 | 12 | 5,69 | 25 | 11 | 1,92 |
| CI07.COC | 48 | 14 | 3,69 | 20 | 14 | 1,54 |
| CI08.ESC | 81 | 10 | 6,23 | 24 | 12 | 1,85 |
| CI09.ILU | 35 | 15 | 2,69 | 17 | 15 | 1,31 |
| CI10.CIU | 115 | 4 | 8,85 | 53 | 4 | 4,08 |
| CI11.CAM | 85 | 9 | 6,54 | 49 | 7 | 3,77 |
| CI12.TRA | 108 | 6 | 8,31 | 21 | 13 | 1,62 |

| | | | | | | |
|---|---|---|---|---|---|---|
| CI13.TRC | 23 | 16 | 1,77 | 11 | 16 | 0,85 |
| CI14.ANI | 126 | 3 | 9,69 | 50 | 6 | 3,85 |
| CI15.JUE | 99 | 7 | 7,62 | 51 | 5 | 3,92 |
| CI16.PRO | 111 | 5 | 8,54 | 66 | 2 | 5,08 |
| Total | 1.509 | | 7,25 | 617 | | 2,97 |

Tabla 15. Índices de DL según la variante hombre.

### Número de palabras

Las informantes actualizan una cantidad mayor de palabras (+9,61%), pero los compañeros presentan un número evidentemente superior de vocablos (+66,35%). Si tomamos como referencia la productividad de los centros de interés coinciden las posiciones de siete campos en ambos conjuntos (CI05, "alimentos y bebidas"; CI01, "partes del cuerpo"; CI16, "profesiones y oficios"; CI02, "la ropa"; CI07, "la cocina y sus utensilios"; CI09, "iluminación y calefacción"; CI13, "trabajos del campo y del jardín"), entre los cuales los más productivos son el CI05, "alimentos y bebidas", y el CI01, "partes del cuerpo". Por el contrario, los que presentan un menor número de actualizaciones son el CI07, "la cocina y sus utensilios"; el CI09, "iluminación y calefacción"; el CI13, "trabajos del campo y del jardín". El único campo que ocupa rangos claramente diferentes según el género de los encuestados es el CI08, "la escuela: muebles y materiales", donde los hombres son menos productivos (rango 10 frente al 6).

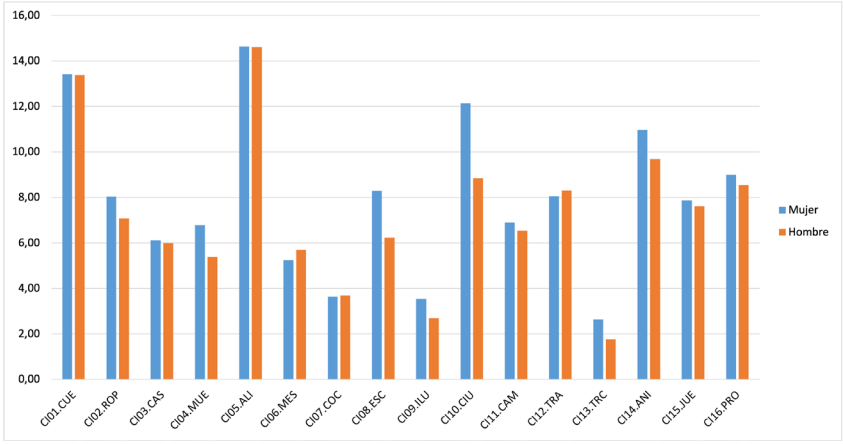


Gráfico 7. Promedio de palabras según la variable sexo.

No se notan valores muy distintos: en general, el desnivel es del 9,62%, pero si desglosamos los promedios, averiguamos dónde la desigualdad es más alta: los campos más homogéneos son el CI05, "alimentos y bebidas", y el CI01, "partes del cuerpo". El rendimiento cambia a favor del conjunto femenino, entre otros, en CI10, "la ciudad" (+37,21%); CI08, "la escuela: muebles y materiales" (+33,19%); CI04, "los muebles de la casa" (+25,94%); CI14, "los animales" (+13,14%). En cambio, los hombres resultan ligeramente más productivos en CI07, "la cocina y sus utensilios" (+1,37%); CI12, "los medios de transporte" (+3,23%); CI06, "objetos colocados en la mesa para la comida" (+8,59%).

Para profundizar el cotejo, la tabla contiene los descriptivos estadísticos relativos a la capacidad productiva de los CI, representada en el diagrama de cajas a continuación.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | Mujer | Hombre |
| Media | 7,95 | 7,25 |
| Mediana | 7,95 | 6,81 |
| Mínimo | 2,63 | 1,77 |
| Máximo | 14,63 | 14,62 |

Tabla 16. Estadísticos descriptivos para la variable sexo.

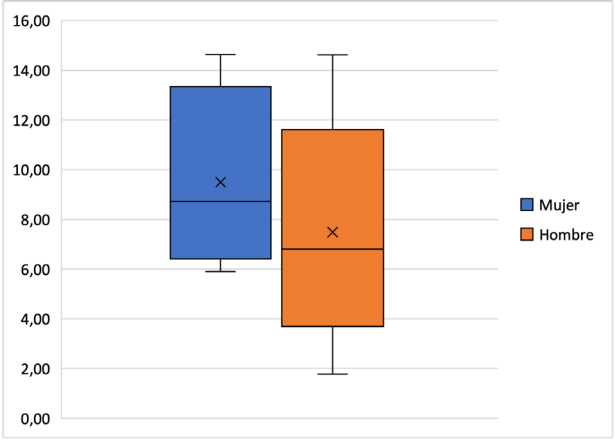


Gráfico 8. Diagrama de cajas para la variable sexo.

Se pone de relieve la superioridad de las mujeres y una desviación superior en las respuestas de los hombres: la mediana (6,81) se coloca por debajo de la media matemática (7,25) y el intervalo de datos entre Q2 y Q3 es más amplio (de ahí que la caja sea más alargada en la parte superior), conllevando una mayor variedad en el número de unidades aportadas.

### Número de vocablos

Si analizamos los promedios de palabras diferentes se hace patente un cambio notable de la tendencia: los hombres sobrepasan la media de las mujeres, casi triplicándolas, ofreciendo una riqueza mayor. Las diferencias más evidentes atañen a C101, "partes del cuerpo"; C108, "la escuela: muebles y materiales"; C107, "la cocina y sus utensilios"; C113, "trabajos del campo y del jardín"; C103, "partes de la casa (sin muebles)", que cambian su colocación de 4 a 6 puestos según cada opción de la variable. Por otra parte, el tema más rico en ambos grupos es el C105, "alimentación y bebidas" (rango 1), siguen el C115, "juegos y distracciones" (rango 5); el C114, "los animales" (rango 6); el C112, "los medios de transporte" (rango 13).

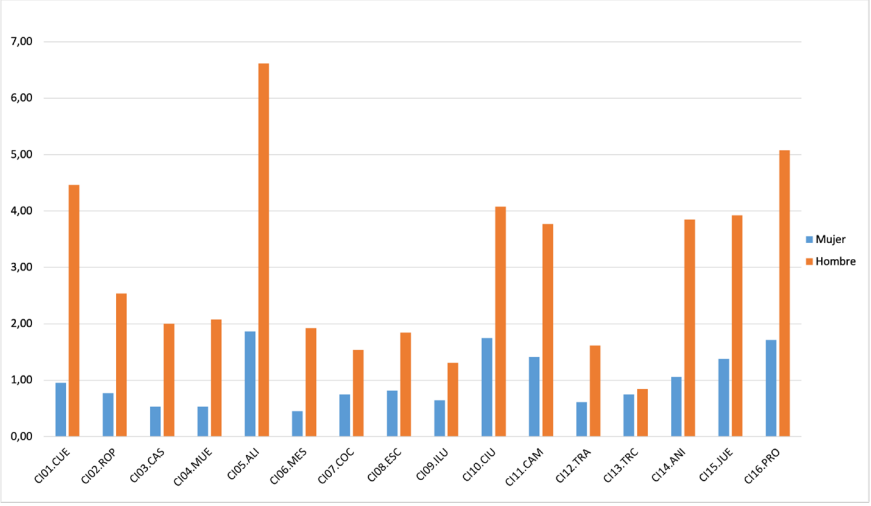


Gráfico 9. Promedio de vocablos según la variable sexo.

Los histogramas naranjas corresponden a las respuestas de los hombres más variadas en todas las áreas y, sobre todo, en el C105, "alimentos y bebidas"; el C101, "partes del cuerpo"; el C116, "profesiones y oficios". Las diferencias se reducen de manera notable en C107, "la cocina y sus

utensilios"; CI09, "iluminación y calefacción"; CI13, "trabajos de campo y del jardín", ya que los cómputos revelan cifras similares.

### Índice de cohesión y densidad léxica

En lo que atañe a la cohesión, los valores alcanzados por las mujeres son bajos, por lo que sus respuestas resultan más heterogéneas. El léxico disponible de este subgrupo es más disperso con respecto al de los hombres: ofrecen un muestreo más compacto que supone una mayor asociación conceptual.

| CI | Índice de cohesión | |
|---|---|---|
| | Mujer | Hombre |
| CI01.CUE | 0,162 | 0,231 |
| CI02.ROP | 0,120 | 0,214 |
| CI03.CAS | 0,133 | 0,231 |
| CI04.MUE | 0,147 | 0,199 |
| CI05.ALI | 0,090 | 0,170 |
| CI06.MES | 0,134 | 0,228 |
| CI07.COC | 0,056 | 0,185 |
| CI08.ESC | 0,117 | 0,260 |
| CI09.ILU | 0,063 | 0,158 |
| CI10.CIU | 0,080 | 0,167 |
| CI11.CAM | 0,056 | 0,133 |
| CI12.TRA | 0,152 | 0,396 |
| CI13.TRC | 0,040 | 0,161 |
| CI14.ANI | 0,119 | 0,194 |
| CI15.JUE | 0,066 | 0,149 |
| CI16.PRO | 0,060 | 0,129 |
| Promedio | 0,100 | 0,200 |

Tabla 17. Índice de cohesión por CI según la variable sexo.

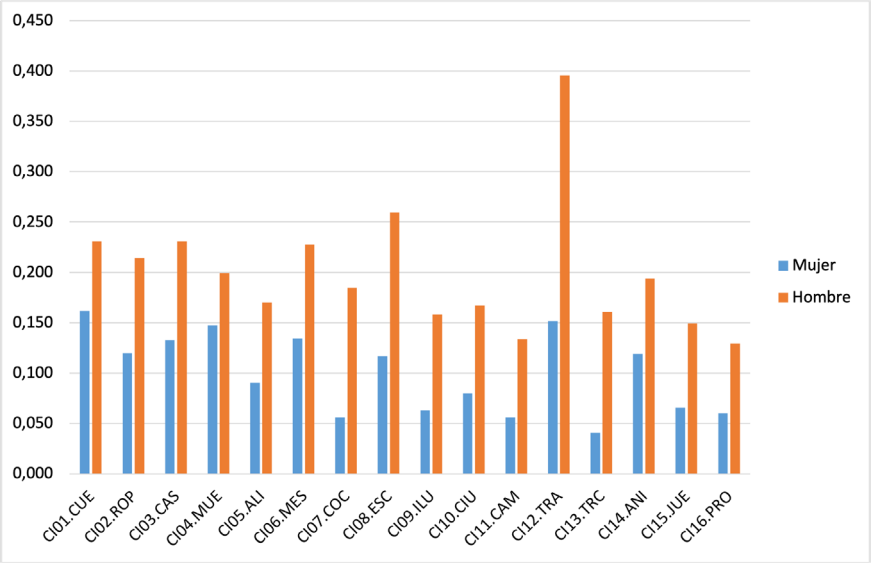


Gráfico 10. Índice de cohesión por CI según la variable sexo.

Antes de analizar la densidad léxica, cabe señalar que se trata de un índice que depende del número de informantes, por tanto, no podemos emplearlo directamente en el cotejo entre los dos subgrupos, pero nos permite establecer en cuáles campos las mujeres o los hombres tienen una media más alta de repeticiones según su ordenación por rango, alejándonos de los valores en sí mismos.

Las mujeres presentan una mayor densidad en el CI01, "las partes del cuerpo", que es uno de los más productivos, seguido por el CI12, "los medios de transporte", que se ubica en el rango 13 debido a la poca cantidad de vocablos actualizados. Lo mismo ocurre en el muestreo de hombres, donde este centro alcanza su densidad máxima. Por otra parte, las mujeres presentan un valor bajo en CI13, "los trabajos del campo y del jardín", posicionado en el rango 16 según el total de palabras, a continuación, hallamos con el CI11, "el campo", en virtud de la gran aportación de vocablos que limitan la media de repeticiones. Los hombres escriben una gran cantidad de palabras diferentes en el CI16, "profesiones y oficios", que aparece como el menos denso.

| CI | Densidad léxica | | | |
|---|---|---|---|---|
| | Mujer | Rango | Hombre | Rango |
| CI01.CUE | 14,06 | 1 | 3 | 3 |
| CI02.ROP | 10,43 | 6 | 2,79 | 6 |

| | | | | |
|---|---|---|---|---|
| CI03.CAS | 11,57 | 5 | 3 | 4 |
| CI04.MUE | 12,83 | 3 | 2,59 | 7 |
| CI05.ALI | 7,86 | 9 | 2,21 | 10 |
| CI06.MES | 11,69 | 4 | 2,96 | 5 |
| CI07.COC | 4,88 | 15 | 2,40 | 9 |
| CI08.ESC | 10,17 | 8 | 3,38 | 2 |
| CI09.ILU | 5,48 | 12 | 2,06 | 13 |
| CI10.CIU | 6,95 | 10 | 2,17 | 11 |
| CI11.CAM | 4,88 | 14 | 1,73 | 15 |
| CI12.TRA | 13,21 | 2 | 5,14 | 1 |
| CI13.TRC | 3,52 | 16 | 2,09 | 12 |
| CI14.ANI | 10,37 | 7 | 2,52 | 8 |
| CI15.JUE | 5,71 | 11 | 1,94 | 14 |
| CI16.PRO | 5,26 | 13 | 1,68 | 16 |
| Promedio | 8,68 | | 2,60 | |

Tabla 18. Densidad léxica por CI según la variable sexo.

### Variable nivel de ELE

Los encuestados se dividen homogéneamente en dos grupos de 50 alumnos.[33] Como era esperable, se nota un incremento del rendimiento del grupo más avanzado (B2), que ofrece un total de 6.517 palabras y de 1.161 vocablos (media: 121,24 y 20,36), cifras superiores con respecto a los aprendientes de nivel B1 que aportan 6.062 palabras y 1.018 vocablos (media: 130,34 y 23,22).

En detalle, las cifras que los informantes han conseguido en las pruebas se exponen en las tablas y nos llevarán a comprobar el peso del factor.

[33] Por eso no ha sido necesario normalizar los datos como en el caso del factor sexo y, como veremos, en el apartado dedicado al conocimiento de otras lenguas extranjeras, donde el desnivel entre una variante y otra es alto.

| *Nivel B1* | | | | | | |
|---|---|---|---|---|---|---|
| **CI** | **P** | **R** | **Pp/I** | **V** | **R** | **Pv/I** |
| CI01.CUE | 594 | 3 | 11,88 | 58 | 7 | 1,16 |
| CI02.ROP | 364 | 9 | 7,28 | 54 | 9 | 1,08 |
| CI03.CAS | 300 | 12 | 6 | 35 | 13 | 0,70 |
| CI04.MUE | 316 | 11 | 6,32 | 38 | 10 | 0,76 |
| CI05.ALI | 669 | 1 | 13,38 | 121 | 1 | 2,42 |
| CI06.MES | 249 | 13 | 4,98 | 30 | 16 | 0,60 |
| CI07.COC | 161 | 15 | 3,22 | 35 | 14 | 0,70 |
| CI08.ESC | 392 | 8 | 7,84 | 57 | 8 | 1,14 |
| CI09.ILU | 173 | 14 | 3,46 | 36 | 12 | 0,72 |
| CI10.CIU | 596 | 2 | 11,92 | 118 | 2 | 2,36 |
| CI11.CAM | 349 | 10 | 6,98 | 92 | 4 | 1,84 |
| CI12.TRA | 396 | 7 | 7,92 | 38 | 11 | 0,76 |
| CI13.TRC | 107 | 16 | 2,14 | 35 | 15 | 0,70 |
| CI14.ANI | 518 | 4 | 10,36 | 71 | 6 | 1,42 |
| CI15.JUE | 417 | 6 | 8,34 | 90 | 5 | 1,80 |
| CI16.PRO | 461 | 5 | 9,22 | 110 | 3 | 2,20 |
| Total | 6.062 | | 7,58 | 1.01 | | 1,27 |

Tabla 19. Índices de DL según la variante nivel B1.

| *Nivel B2* | | | | | | |
|---|---|---|---|---|---|---|
| **CI** | **P** | **R** | **Pp/I** | **V** | **R** | **Pv/I** |
| CI01.CUE | 747 | 2 | 14,94 | 72 | 7 | 1,44 |
| CI02.ROP | 427 | 6 | 8,54 | 60 | 8 | 1,20 |
| CI03.CAS | 310 | 12 | 6,20 | 40 | 14 | 0,80 |
| CI04.MUE | 344 | 10 | 6,88 | 41 | 13 | 0,82 |

| | | | | | | |
|---|---|---|---|---|---|---|
| CI05.ALI | 794 | 1 | 15,88 | 148 | 1 | 2,96 |
| CI06.MES | 281 | 13 | 5,62 | 33 | 16 | 0,66 |
| CI07.COC | 204 | 14 | 4,08 | 51 | 11 | 1,02 |
| CI08.ESC | 411 | 8 | 8,22 | 54 | 10 | 1,08 |
| CI09.ILU | 169 | 15 | 3,38 | 46 | 12 | 0,92 |
| CI10.CIU | 575 | 3 | 11,50 | 116 | 3 | 2,32 |
| CI11.CAM | 336 | 11 | 6,72 | 102 | 4 | 2,04 |
| CI12.TRA | 412 | 7 | 8,24 | 40 | 15 | 0,80 |
| CI13.TRC | 145 | 16 | 2,90 | 55 | 9 | 1,10 |
| CI14.ANI | 562 | 4 | 11,24 | 90 | 6 | 1,80 |
| CI15.JUE | 367 | 9 | 7,34 | 96 | 5 | 1,92 |
| CI16.PRO | 433 | 5 | 8,66 | 117 | 2 | 2,34 |
| Total | 6.517 | | 8,15 | 1.161 | | 1,45 |

Tabla 20. Índices de DL según la variante nivel B2.

**Número de palabras**

Los estudiantes que se encuentran en una etapa de estudio más avanzada aportan, en general, un caudal mayor de respuestas. En porcentaje se registra un incremento del 7,51% de palabras, a pesar de que los alumnos de nivel inferior sean más productivos en CI09, "iluminación y calefacción" (+2,37%); CI11, "el campo" (+3,87%); CI10, "la ciudad" (+3,65%); CI16, "profesiones y oficios" (+6,47%); CI15, "juegos y distracciones" (+13,62%).

Si observamos el orden de los CI según su productividad, siete se encuentran en las mismas posiciones sin distinción de variante (CI03, "partes de la casa (sin muebles)"; CI05, "alimentos y bebidas"; CI08, "la escuela: muebles y materiales"; CI12, "medios de transporte"; CI13, "trabajos del campo y del jardín"; CI14, "los animales"; CI16, "profesiones y oficios"). El más productivo es el CI05, "alimentos y bebidas", que se aleja del último en más de 500 palabras, se trata del CI13, "trabajos del campo y del jardín", que es el menos rentable en ambos muestreos.

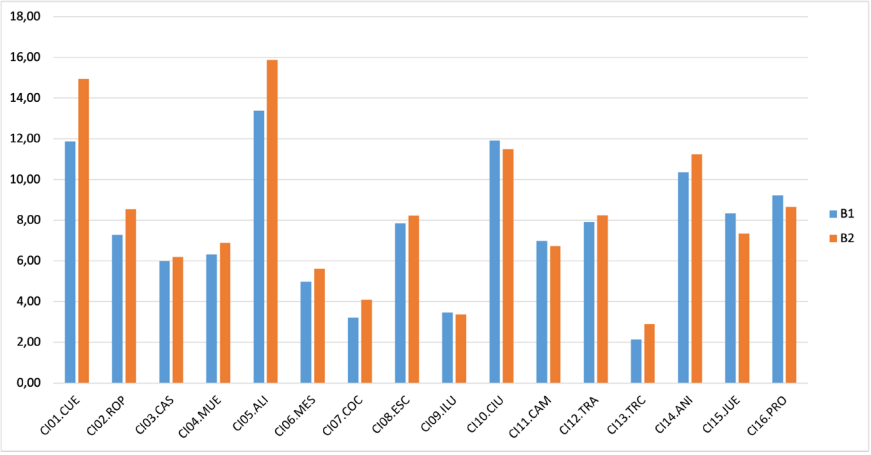


Gráfico 11. Promedio de palabras según la variable nivel de ELE.

No hay una significativa oscilación entre las dos agrupaciones, pero los promedios favorecen los informantes de nivel B2. En general, la disparidad va de 7,58 palabras en la variante B1 a 8,15 palabras en la B2. Detectamos como máximo un aumento del 25,76% en el CI01, "partes del cuerpo".

Los datos estadísticos ayudan a organizar la información, pues, se nota la capacidad del grupo avanzado para actualizar más palabras: los valores máximo y mínimo (15,88 y 2,90) son más altos que los del nivel umbral (13,38 y 2,14).

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B1 | B2 |
| Media | 7,58 | 8,15 |
| Mediana | 7,56 | 7,78 |
| Mínimo | 2,14 | 2,90 |
| Máximo | 13,38 | 15,88 |

Tabla 21. Estadísticos descriptivos para la variable nivel de ELE.

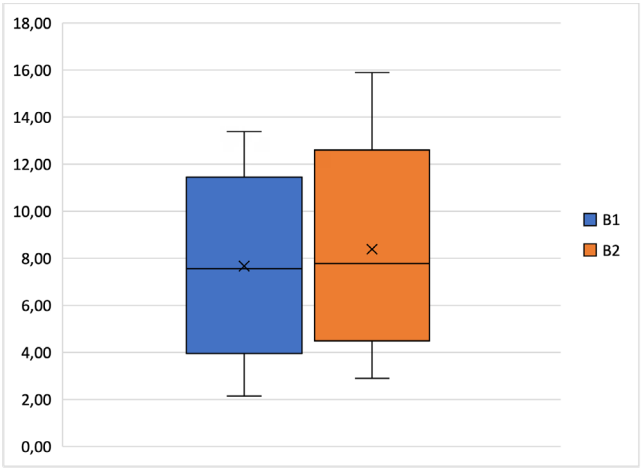


Gráfico 12. Diagrama de cajas para la variable nivel de ELE.

Los alumnos de nivel B2 presentan una desviación mayor de las respuestas entre el segundo y el tercer cuartil, que incluye la media matemática (8,15), que se coloca por encima de la mediana (7,78). En cambio, media y mediana de los B1 se solapan: su producción intragrupal es más simétrica.

### Número de vocablos

El grupo avanzado prevalece en casi todos los centros de interés, llegando a responder con una media de vocablos mayor del 57% en CI13, "trabajos del campo y del jardín", y del 46% en CI07, "la cocina y sus utensilios", (en media el aumento registrado es del 14,05%) a excepción de los CI08, "la escuela: muebles y materiales", (-5%) y CI10, "la ciudad", (-2%), en los que domina, ligeramente, la otra variante.

Los campos que coinciden en ambas agrupaciones según el rango de variedad léxica son CI01, "partes del cuerpo"; CI05, "alimentos y bebidas"; CI06, "objetos colocados en la mesa para la comida"; CI09, "iluminación y calefacción"; CI11, "el campo"; CI14, "los animales"; CI15, "juegos y distracciones". De nuevo, el CI05, "alimentos y bebidas", ocupa la primera posición y resulta el más rico en los dos niveles. En cambio, el CI06, "objetos colocados en la mesa para la comida", es el menos variado.

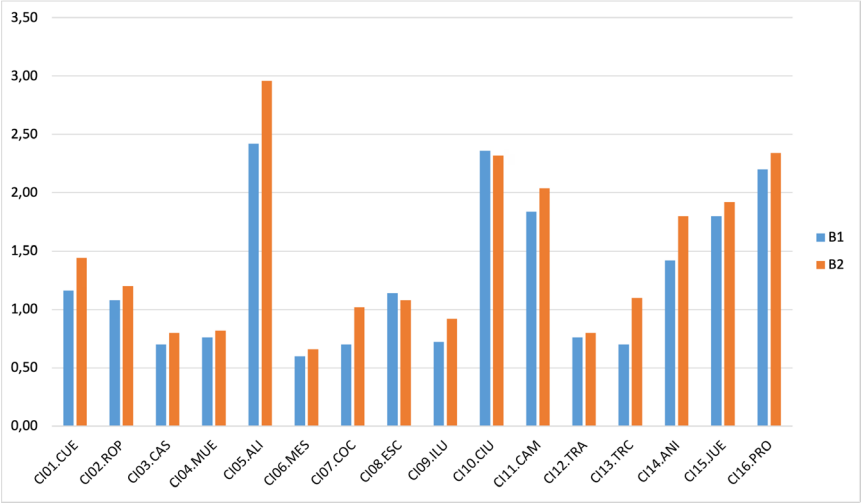


Gráfico 13. Promedio de vocablos según la variable nivel de ELE.

### Índice de cohesión y densidad léxica

En lo que atañe a los índices de cohesión y densidad, la tabla muestra los valores desglosados por centro de interés. En este caso unimos los dos indicadores dado que las dos agrupaciones tenían el mismo número de sujetos.

| CI | Índice de cohesión | | Densidad léxica | |
|---|---|---|---|---|
| | B1 | B2 | B1 | B2 |
| CI01.CUE | 0,205 | 0,208 | 10,24 | 10,38 |
| CI02.ROP | 0,135 | 0,137 | 6,74 | 7,12 |
| CI03.CAS | 0,171 | 0,149 | 8,57 | 7,75 |
| CI04.MUE | 0,166 | 0,161 | 8,32 | 8,39 |
| CI05.ALI | 0,111 | 0,103 | 5,53 | 5,36 |
| CI06.MES | 0,166 | 0,164 | 8,30 | 8,52 |
| CI07.COC | 0,092 | 0,077 | 4,60 | 4 |
| CI08.ESC | 0,138 | 0,146 | 6,88 | 7,61 |
| CI09.ILU | 0,096 | 0,071 | 4,81 | 3,67 |

| | | | | |
|---|---|---|---|---|
| CI10.CIU | 0,101 | 0,095 | 5,05 | 4,96 |
| CI11.CAM | 0,076 | 0,063 | 3,79 | 3,29 |
| CI12.TRA | 0,208 | 0,198 | 10,42 | 10,30 |
| CI13.TRC | 0,061 | 0,051 | 3,06 | 2,64 |
| CI14.ANI | 0,146 | 0,120 | 7,30 | 6,24 |
| CI15.JUE | 0,093 | 0,074 | 4,63 | 3,82 |
| CI16.PRO | 0,084 | 0,071 | 4,19 | 3,70 |
| Promedio | 0,128 | 0,118 | 6,40 | 6,11 |

Tabla 22. Índice de cohesión y densidad léxica por CI según la variable nivel de ELE.

Aunque el léxico disponible del grupo B1 parece más compacto y denso de vocablos, el desnivel no es sobresaliente, ya que computamos un aumento de 8,47% en la media del IC y un 4,25% en la densidad léxica, con lo cual parece que este factor no tiene mucha influencia al pasar de un nivel a otro.

Si miramos los valores de cada centro destaca que el B2 alcanza una cohesión más alta en tres (CI08, "la escuela: muebles y materiales"; CI01, "partes del cuerpo"; CI02, "la ropa"), las respuestas del otro grupo son ligeramente más compactas en los demás campos. Lógicamente, se detecta la misma tendencia en lo que se refiere a la densidad.

Con respecto a la incidencia de estos índices sobre la relación entre productividad y riqueza léxica, ponemos de relieve lo que ocurre en ambas variantes en el CI11, "el campo", y en el CI12, "los medios de transporte": en el primero calculamos un bajo nivel de cohesión y densidad que corresponde a una variedad léxica apreciable mientras que en el segundo los índices se colocan entre los más altos, por lo que el centro es más productivo que rico.

Gráficamente:

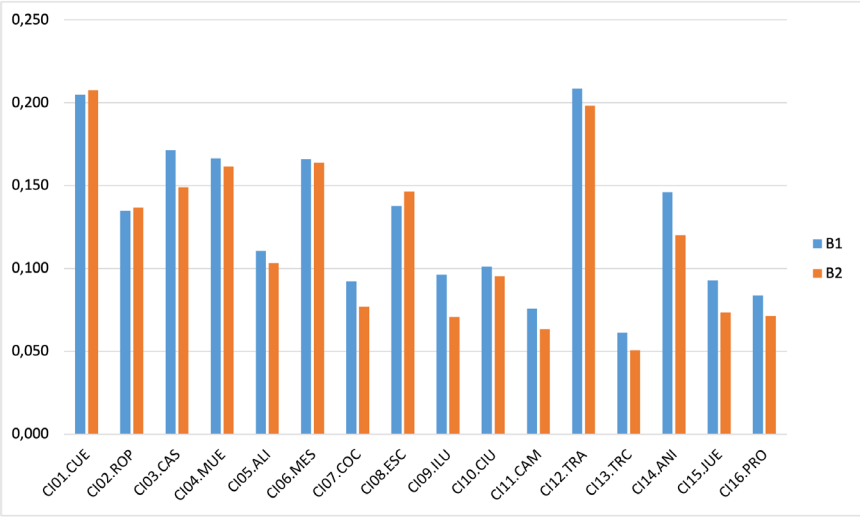


Gráfico 14. Índice de cohesión según la variable nivel de ELE.

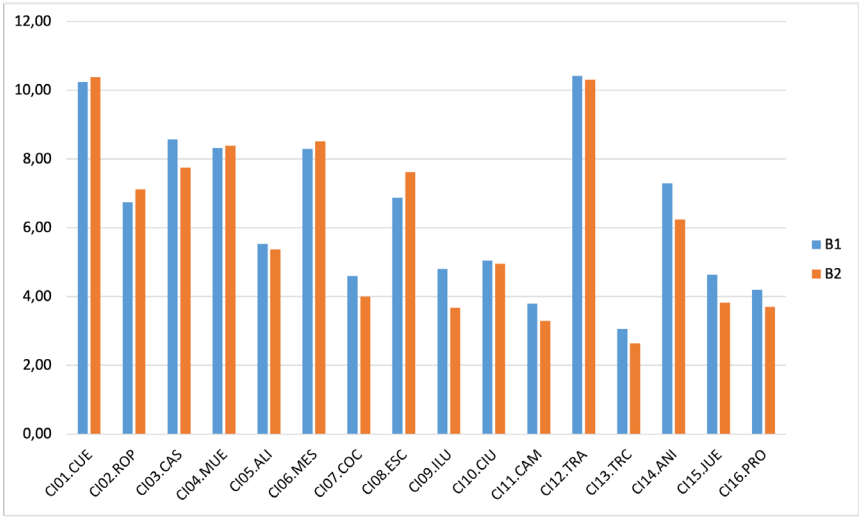


Gráfico 15. Densidad léxica según la variable nivel de ELE.

### Variable conocimiento de otras LE

Las dos agrupaciones son dispares, los informantes que conocen dos lenguas extranjeras eran 19 y los que conocen más de dos constituyeron

un conjunto más numeroso formado por 81 sujetos. Contabilizamos en el primer grupo 2.202 palabras y 659 vocablos (media: 115,89 y 34,68) y en el segundo 10.377 palabras y 1.381 vocablos (media: 128,11 y 17,05). Debido a este desnivel tenemos en cuenta, de nuevo, los datos promediales durante el análisis.

| ***Conoce 2 LE*** | | | | | | |
|---|---|---|---|---|---|---|
| **CI** | **P** | **R** | **Pp/I** | **V** | **R** | **Pv/I** |
| CI01.CUE | 242 | 2 | 12,74 | 53 | 6 | 2,79 |
| CI02.ROP | 135 | 8 | 7,11 | 34 | 9 | 1,79 |
| CI03.CAS | 111 | 11 | 5,84 | 25 | 10 | 1,32 |
| CI04.MUE | 100 | 13 | 5,26 | 24 | 11 | 1,26 |
| CI05.ALI | 264 | 1 | 13,89 | 92 | 1 | 4,84 |
| CI06.MES | 102 | 12 | 5,37 | 23 | 12 | 1,21 |
| CI07.COC | 60 | 14 | 3,16 | 22 | 13 | 1,16 |
| CI08.ESC | 131 | 9 | 6,89 | 35 | 8 | 1,84 |
| CI09.ILU | 50 | 15 | 2,63 | 16 | 15 | 0,84 |
| CI10.CIU | 193 | 4 | 10,16 | 62 | 4 | 3,26 |
| CI11.CAM | 119 | 10 | 6,26 | 53 | 7 | 2,79 |
| CI12.TRA | 158 | 5 | 8,32 | 18 | 14 | 0,95 |
| CI13.TRC | 36 | 16 | 1,89 | 15 | 16 | 0,79 |
| CI14.ANI | 208 | 3 | 10,95 | 66 | 3 | 3,47 |
| CI15.JUE | 150 | 6 | 7,89 | 54 | 5 | 2,84 |
| CI16.PRO | 143 | 7 | 7,53 | 67 | 2 | 3,53 |
| Total | 2.202 | | 7,24 | 659 | | 2,17 |

Tabla 23. Índices de DL según la variante LE =2.

| ***Conoce más de 2 LE*** | | | | | | |
|---|---|---|---|---|---|---|
| **CI** | **P** | **R** | **Pp/I** | **V** | **R** | **Pv/I** |
| CI01.CUE | 1.099 | 2 | 13,57 | 82 | 7 | 1,01 |
| CI02.ROP | 656 | 7 | 8,10 | 70 | 8 | 0,86 |

| | | | | | | |
|---|---|---|---|---|---|---|
| CI03.CAS | 499 | 12 | 6,16 | 47 | 15 | 0,58 |
| CI04.MUE | 560 | 11 | 6,91 | 48 | 14 | 0,59 |
| CI05.ALI | 1.199 | 1 | 14,80 | 160 | 1 | 1,98 |
| CI06.MES | 428 | 13 | 5,28 | 39 | 16 | 0,48 |
| CI07.COC | 305 | 14 | 3,77 | 63 | 10 | 0,78 |
| CI08.ESC | 672 | 6 | 8,30 | 67 | 9 | 0,83 |
| CI09.ILU | 292 | 15 | 3,60 | 52 | 13 | 0,64 |
| CI10.CIU | 978 | 3 | 12,07 | 149 | 3 | 1,84 |
| CI11.CAM | 566 | 10 | 6,99 | 126 | 4 | 1,56 |
| CI12.TRA | 650 | 8 | 8,02 | 56 | 12 | 0,69 |
| CI13.TRC | 216 | 16 | 2,67 | 62 | 11 | 0,77 |
| CI14.ANI | 872 | 4 | 10,77 | 88 | 6 | 1,09 |
| CI15.JUE | 634 | 9 | 7,83 | 121 | 5 | 1,49 |
| CI16.PRO | 751 | 5 | 9,27 | 151 | 2 | 1,86 |
| Total | 10.377 | | 8,01 | 1.381 | | 1,07 |

Tabla 24. Índices de DL según la variante LE >2.

### Número de palabras

Los alumnos que conocen más de dos LE aportan el 10,54% más de palabras, si bien no logran la misma supremacía en el promedio de vocablos, ya que los compañeros que conocen dos los doblan aportando el +103,40% de variedad léxica.

Los dos grupos coinciden en el grado de productividad en seis centros de interés (CI05, "alimentos y bebidas"; CI01, "partes del cuerpo"; CI07, "la cocina y sus utensilios"; CI09, "iluminación y calefacción"; CI11, "el campo"; CI13, "trabajos del campo y del jardín"), de los que los más rentables son el CI05, "alimentos y bebidas" (rango 1), y el CI01, "partes del cuerpo" (rango 2). Por otro lado, el CI13, "trabajos del campo y del jardín", presenta menos palabras en absoluto (rango 16).

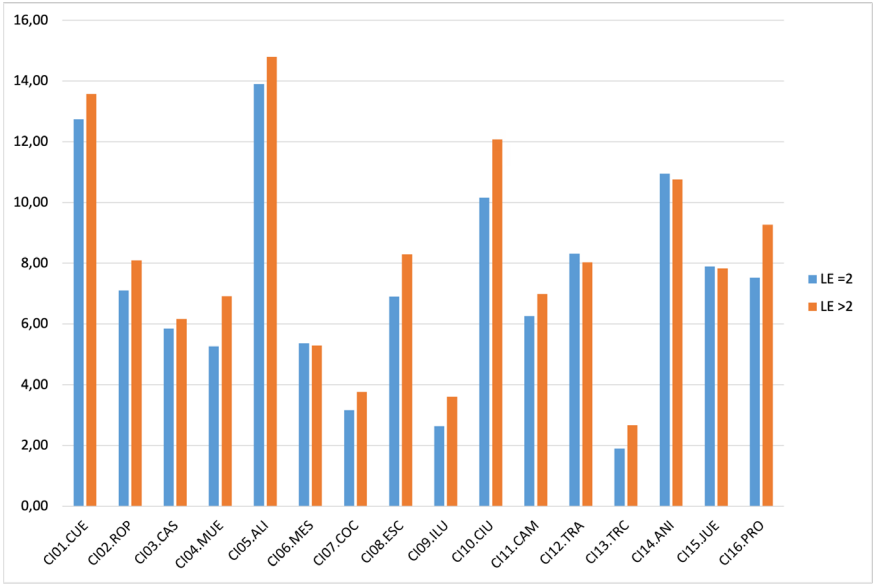


Gráfico 16. Promedio de palabras según la variable conocimiento de otras LE.

Al desglosar los valores según centro de interés, se revela un desnivel de casi dos palabras en el CI10, "la ciudad" (+18,86%), seguido por el CI08, "la escuela: muebles y materiales" (+20,33%); el CI16, "profesiones y oficios" (+23,19%); el CI04, "los muebles de la casa" (+31,36%). En los demás temas las diferencias son menores, pero siempre a favor del conjunto LE >2. El rendimiento aumenta ligeramente en CI12, "los medios de transporte" (+3,61%); CI14, "los animales" (+1,67%); CI06, "objetos colocados en la mesa para la comida" (+1,51%); CI15, "juegos y distracciones" (+0,89%) en el otro grupo. La tabla y el gráfico ilustran la simetría o la dispersión detectadas en la producción de palabras:

| **Descriptivos** | **Variante** | |
|---|---|---|
| | LE =2 | LE >2 |
| Media | 7,24 | 8,01 |
| Mediana | 7 | 7,93 |
| Mínimo | 1,89 | 2,67 |
| Máximo | 13,89 | 14,80 |

Tabla 25. Estadísticos descriptivos para la variable conocimiento de otras LE.

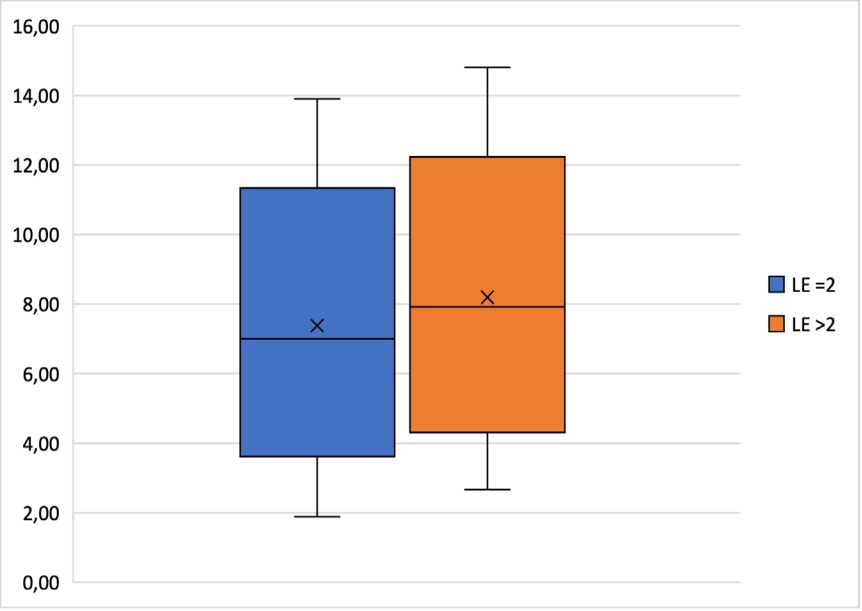


Gráfico 17. Diagrama de cajas para la variable conocimiento de otras LE.

La superioridad de los encuestados que conocen más de dos LE se hace manifiesta, pero pese a esta divergencia cuantitativa, se rastrea una correspondencia entre las desviaciones de los dos muestreos, ya que el rango intercuartílico (la diferencia entre Q3 y Q1) es homogéneo: los intervalos se revelan equilibrados en lo que atañe a la distribución de respuestas.

### Número de vocablos

El orden de los CI según la variedad léxica pone de manifiesto una situación parecida: los más ricos en sendos grupos son el CI05, "alimentos y bebidas" (rango 1); el CI16, "profesiones y oficios" (rango 2); el CI15, "juegos y distracciones" (rango 5). Los campos con una menor cantidad de vocablos son: CI03, "partes de la casa (sin muebles)"; CI04, "los muebles de la casa"; CI06, "objetos colocados en la mesa para la comida"; CI07, "la cocina y sus utensilios"; CI11, "el campo".

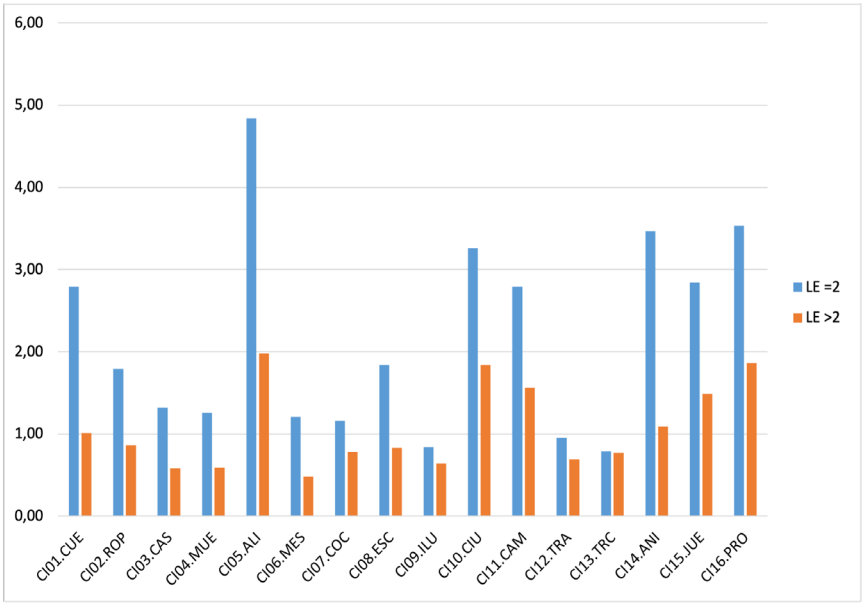


Gráfico 18. Promedio de vocablos según la variable conocimiento de otras LE.

De manera evidente, la disparidad se advierte en el CI05, "alimentos y bebidas"; el CI14, "los animales"; el CI01, "partes del cuerpo", donde la cantidad de vocablos activadas por el grupo LE =2 se multiplica por más del doble. El desnivel baja ligeramente, aunque siguiendo alto, en CI09, "iluminación y calefacción" (+31%); CI12, "los medios de transporte" (+38%); CI07, "la cocina y sus utensilios" (+49%). La diferencia se reduce como máximo del +3% en el CI13, "trabajos del campo y del jardín".

### Índice de cohesión y densidad léxica

La cohesión destaca que los índices del grupo LE =2 son más altos y, por tanto, el léxico disponible es más cerrado y tiene una mayor asociación conceptual con respecto al otro que presenta más palabras distintas, en particular en el CI12, "medios de transporte"; el CI03, "partes de la casa (sin muebles)"; el CI06, "objetos colocados en la mesa para la comida".

| CI | Índice de cohesión | |
|---|---|---|
| | LE =2 | LE >2 |
| CI01.CUE | 0,240 | 0,165 |
| CI02.ROP | 0,209 | 0,116 |

| | | |
|---|---|---|
| CI03.CAS | 0,234 | 0,131 |
| CI04.MUE | 0,219 | 0,144 |
| CI05.ALI | 0,151 | 0,093 |
| CI06.MES | 0,233 | 0,135 |
| CI07.COC | 0,144 | 0,060 |
| CI08.ESC | 0,197 | 0,124 |
| CI09.ILU | 0,164 | 0,069 |
| CI10.CIU | 0,164 | 0,081 |
| CI11.CAM | 0,118 | 0,055 |
| CI12.TRA | 0,462 | 0,143 |
| CI13.TRC | 0,126 | 0,043 |
| CI14.ANI | 0,166 | 0,122 |
| CI15.JUE | 0,146 | 0,065 |
| CI16.PRO | 0,112 | 0,061 |
| Promedio | 0,193 | 0,101 |

Tabla 26. Índice de cohesión por CI según la variable conocimiento de otras LE.

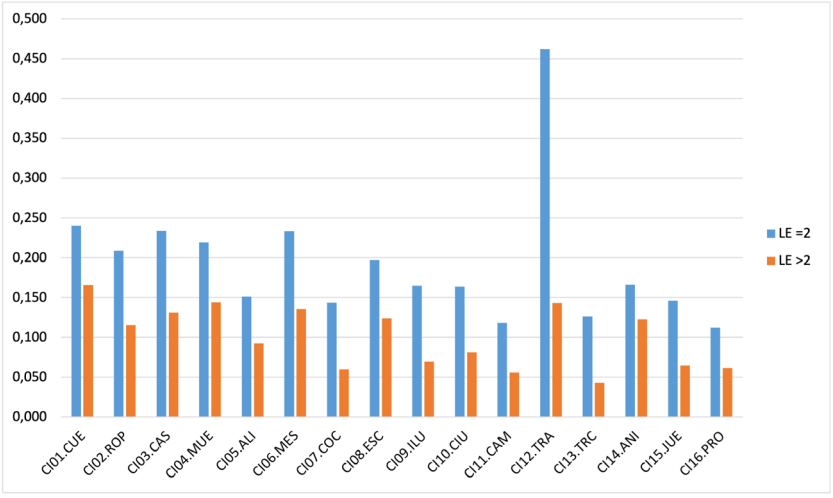


Gráfico 19. Índice de cohesión según la variable conocimiento de otras LE.

En conclusión, la variante LE =2 presenta la mayor densidad léxica en CI12, "los medios de transporte" (8,78), y la menor en CI16, "profesiones y oficios" (2,13). Por su parte, la variante LE >2 llega a su valor máximo en CI01, "las partes del cuerpo" (13,40) y al mínimo en CI13, "trabajos del campo y del jardín" (3,48). La correspondencia por rango se observa en cuatro centros: CI06, "objetos colocados en la mesa para la comida" (rango 4); CI10, "la ciudad" (rango 10); CI15, "juegos y distracciones" (rango 12); CI11, "el campo" (rango 15).

| CI | Densidad léxica | | | |
|---|---|---|---|---|
| | LE =2 | Rango | LE >2 | Rango |
| CI01.CUE | 4,57 | 2 | 13,40 | 1 |
| CI02.ROP | 3,97 | 6 | 9,37 | 8 |
| CI03.CAS | 4,44 | 3 | 10,62 | 5 |
| CI04.MUE | 4,17 | 5 | 11,67 | 2 |
| CI05.ALI | 2,87 | 11 | 7,49 | 9 |
| CI06.MES | 4,43 | 4 | 10,97 | 4 |
| CI07.COC | 2,73 | 13 | 4,84 | 14 |
| CI08.ESC | 3,74 | 7 | 10,03 | 6 |
| CI09.ILU | 3,13 | 9 | 5,62 | 11 |
| CI10.CIU | 3,11 | 10 | 6,56 | 10 |
| CI11.CAM | 2,25 | 15 | 4,49 | 15 |
| CI12.TRA | 8,78 | 1 | 11,61 | 3 |
| CI13.TRC | 2,40 | 14 | 3,48 | 16 |
| CI14.ANI | 3,15 | 8 | 9,91 | 7 |
| CI15.JUE | 2,78 | 12 | 5,24 | 12 |
| CI16.PRO | 2,13 | 16 | 4,97 | 13 |
| Promedio | 3,67 | | 8,14 | |

Tabla 27. Densidad léxica por CI según la variable conocimiento de otras LE.

### 3.1.3 Análisis longitudinal

En esta sección cotejamos los datos procedentes de la primera suministración de la prueba, realizada a comienzo del año académico, y de la segunda, realizada a final del curso por los 50 estudiantes de nivel B2 con el propósito de medir la evolución en el tiempo de su léxico disponible. Apreciamos un total de 15.008 palabras y 2.405 vocablos repartidos como sigue entre los dos grupos: B2_a 6.517 palabras y 1.161 vocablos (media: 130,34 y 23,22); B2_b 8.491 palabras y 1.244 vocablos (media: 169,82 y 24,88).

#### Número de palabras

A comienzo del año los informantes activan un total de 6.517 ítems y a final 8.491, con un incremento del 30,29%. En la segunda prueba el número de palabras es mayor en todos los campos, desde un +5% en el CI01, "partes del cuerpo", hasta reduplicarlo en el CI09, "iluminación y calefacción".

| *B2_a* | | | CI | *B2_b* | | |
|---|---|---|---|---|---|---|
| **R** | **P** | **Pp/I** | | **R** | **P** | **Pp/I** |
| 2 | 747 | 14,94 | CI01.CUE | 2 | 785 | 15,70 |
| 6 | 427 | 8,54 | CI02.ROP | 7 | 594 | 11,88 |
| 12 | 310 | 6,20 | CI03.CAS | 11 | 421 | 8,42 |
| 10 | 344 | 6,88 | CI04.MUE | 10 | 444 | 8,88 |
| 1 | 794 | 15,88 | CI05.ALI | 1 | 850 | 17 |
| 13 | 281 | 5,62 | CI06.MES | 13 | 399 | 7,98 |
| 14 | 204 | 4,08 | CI07.COC | 15 | 288 | 5,76 |
| 8 | 411 | 8,22 | CI08.ESC | 5 | 689 | 13,78 |
| 15 | 169 | 3,38 | CI09.ILU | 14 | 338 | 6,76 |
| 3 | 575 | 11,50 | CI10.CIU | 4 | 710 | 14,20 |
| 11 | 336 | 6,72 | CI11.CAM | 12 | 402 | 8,04 |
| 7 | 412 | 8,24 | CI12.TRA | 8 | 522 | 10,44 |
| 16 | 145 | 2,90 | CI13.TRC | 16 | 208 | 4,16 |
| 4 | 562 | 11,24 | CI14.ANI | 3 | 729 | 14,58 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | 367 | 7,34 | CI15.JUE | 9 | 484 | 9,68 |
| 5 | 433 | 8,66 | CI16.PRO | 6 | 628 | 12,56 |
| / | 407,31 | 8,15 | Promedio | / | 530,69 | 10,61 |

Tabla 28. Número de palabras por CI según la fecha de la prueba.

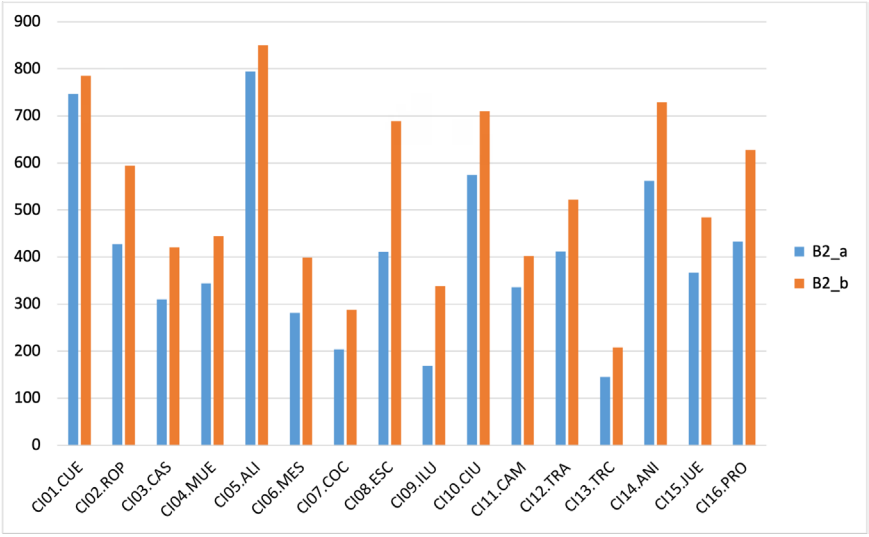


Gráfico 20. Número total de palabras según la fecha de la prueba.

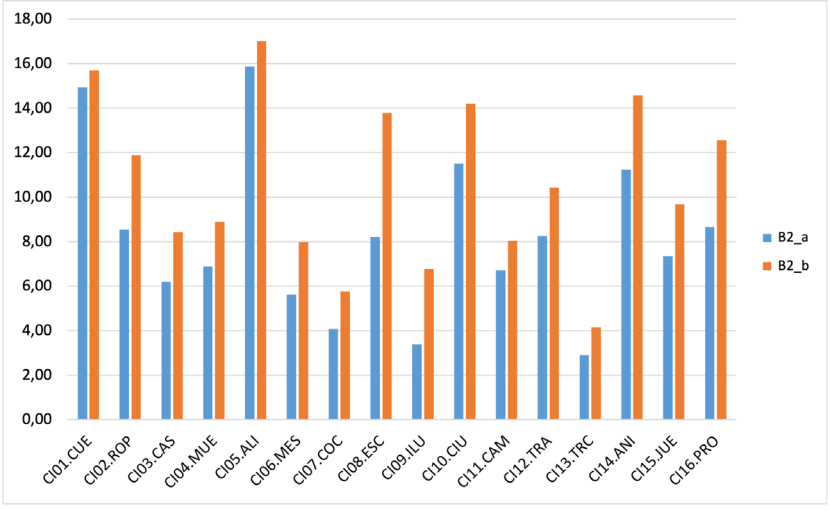


Gráfico 21. Promedio de palabras por informante según la fecha de la prueba.

Al considerar el orden de los centros de interés por rango, se pueden clasificar en función de si se encuentran por encima o por debajo de la media.

| *B2_a* | | | CI | *B2_b* | | |
|---|---|---|---|---|---|---|
| **R** | **P** | **Pp/I** | | **R** | **P** | **Pp/I** |
| 1 | 794 | 15,88 | CI01.CUE | 1 | 850 | 17 |
| 2 | 747 | 14,94 | CI02.ROP | 2 | 785 | 15,70 |
| 3 | 575 | 11,50 | CI03.CAS | 3 | 729 | 14,58 |
| 4 | 562 | 11,24 | CI04.MUE | 4 | 710 | 14,20 |
| 5 | 433 | 8,66 | CI05.ALI | 5 | 689 | 13,78 |
| 6 | 427 | 8,54 | CI06.MES | 6 | 628 | 12,56 |
| 7 | 412 | 8,24 | CI07.COC | 7 | 594 | 11,88 |
| 8 | 411 | 8,22 | CI08.ESC | 8 | 522 | 10,44 |
| 9 | 367 | 7,34 | CI09.ILU | 9 | 484 | 9,68 |
| 10 | 344 | 6,88 | CI10.CIU | 10 | 444 | 8,88 |
| 11 | 336 | 6,72 | CI11.CAM | 11 | 421 | 8,42 |
| 12 | 310 | 6,20 | CI12.TRA | 12 | 402 | 8,04 |
| 13 | 281 | 5,62 | CI13.TRC | 13 | 399 | 7,98 |
| 14 | 204 | 4,08 | CI14.ANI | 14 | 338 | 6,76 |
| 15 | 169 | 3,38 | CI15.JUE | 15 | 288 | 5,76 |
| 16 | 145 | 2,90 | CI16.PRO | 16 | 208 | 4,16 |

Tabla 29. CI ordenados en función de su productividad léxica según la fecha de la prueba.

En la primera prueba los centros que la sobrepasan son ocho, mientras que en la segunda son siete. Se trata de los mismos CI, con lo cual la capacidad productiva es idéntica en función de estos estímulos. He aquí seis que se colocan exactamente en los mismos rangos (CI05, "alimentos y bebidas"; CI01, "partes del cuerpo"; CI15, "juegos y distracciones"; CI04, "los muebles de la casa"; CI06, "objetos colocados en la mesa para la comida"; CI13, "trabajos del campo y del jardín"). El CI05, "alimentos y bebidas", ocupa la primera posición con un incremento del 7% en B2_b. El segundo centro más productivo es el CI01, "partes del cuerpo", que

crece del 5%. El campo menos rentable, el C113, "trabajos del campo y del jardín", también goza de un aumento del 43%. Otros estímulos se han posicionado en rangos superiores, resultando más productivos (C114, "los animales"; C108, "la escuela: muebles y materiales"; C103, "partes de la casa (sin muebles)"; C109, "iluminación y calefacción") en detrimento de otros que han retrocedido, pero siempre ofreciendo un enriquecimiento cuantitativo a final del curso.

La tabla proporciona los datos estadísticos que corroboran lo expuesto.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B2_a | B2_b |
| Media | 8,15 | 10,61 |
| Mediana | 7,78 | 10,06 |
| Mínimo | 2,90 | 4,16 |
| Máximo | 15,88 | 17 |

Tabla 30. Estadísticos descriptivos según la fecha de la prueba.

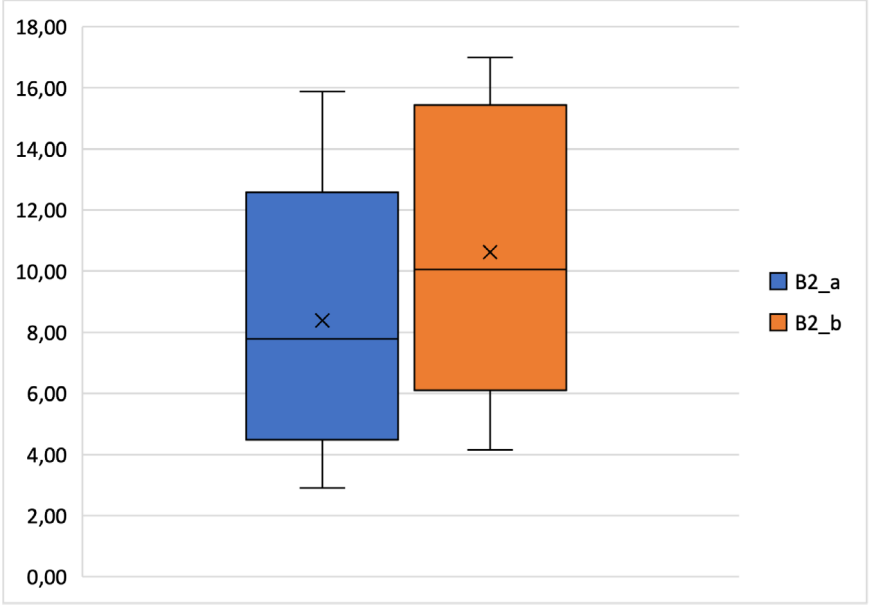


Gráfico 22. Diagrama de cajas según la fecha de la prueba.

En lo referente a la oscilación intergrupal, se confirma la mejora de los resultados en la segunda aplicación de la prueba, ya que la caja naranja se extiende más y no baja hasta el punto inferior de la azul (2,90) quedándose más arriba (4,16). Asimismo, la posición de las cajas en el gráfico permite observar dónde la variación intragrupal es mayor, a saber, en los valores más altos que se posicionan entre la mediana y el tercer cuartil del grupo B2_b. Igualmente, el B2_a presenta una asimetría más grande en el mismo intervalo, pero simultáneamente disminuye hasta tocar el punto mínimo.

**Número de vocablos**

El análisis de la cantidad de vocablos no revela valores más altos en el grupo B2_b, ya que resulta menos rico en seis campos semánticos (CI05, "alimentos y bebidas"; CI01, "partes del cuerpo"; CI15, "juegos y distracciones"; CI11, "el campo"; CI13, "trabajos del campo y del jardín"; CI16, "profesiones y oficios"), hasta disminuir del 5,70% en el CI05, "alimentos y bebidas". En todo caso, hay un aumento general del 7,50% (83 vocablos más, en media 1,66).

| ***B2_a*** | | | **CI** | ***B2_b*** | | |
|---|---|---|---|---|---|---|
| **R** | **V** | **Pv/I** | | **R** | **V** | **Pv/I** |
| 7 | 72 | 1,44 | CI01.CUE | 9 | 65 | 1,30 |
| 8 | 60 | 1,20 | CI02.ROP | 8 | 67 | 1,34 |
| 14 | 40 | 0,80 | CI03.CAS | 14 | 45 | 0,90 |
| 13 | 41 | 0,82 | CI04.MUE | 15 | 41 | 0,82 |
| 1 | 148 | 2,96 | CI05.ALI | 1 | 140 | 2,80 |
| 16 | 33 | 0,66 | CI06.MES | 16 | 35 | 0,70 |
| 11 | 51 | 1,02 | CI07.COC | 11 | 61 | 1,22 |
| 10 | 54 | 1,08 | CI08.ESC | 4 | 99 | 1,98 |
| 12 | 46 | 0,92 | CI09.ILU | 12 | 58 | 1,16 |
| 3 | 116 | 2,32 | CI10.CIU | 2 | 124 | 2,48 |
| 4 | 102 | 2,04 | CI11.CAM | 6 | 96 | 1,92 |
| 15 | 40 | 0,80 | CI12.TRA | 10 | 65 | 1,30 |
| 9 | 55 | 1,10 | CI13.TRC | 13 | 49 | 0,98 |
| 6 | 90 | 1,80 | CI14.ANI | 5 | 97 | 1,94 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5 | 96 | 1,92 | CI15.JUE | 7 | 89 | 1,78 |
| 2 | 117 | 2,34 | CI16.PRO | 3 | 113 | 2,26 |
| / | 23,22 | 1,45 | Promedio | / | 24,88 | 1,56 |

Tabla 31. Número de vocablos por CI según la fecha de la prueba.

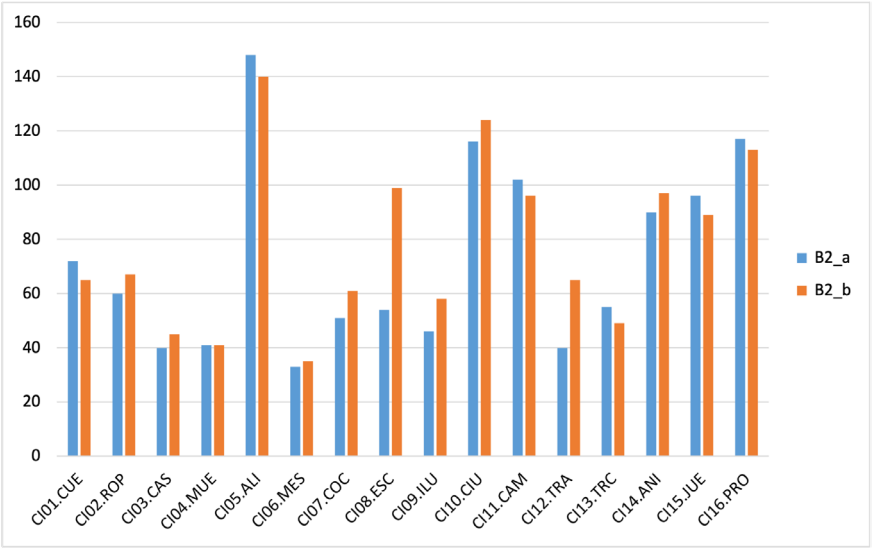


Gráfico 23. Número total de vocablos según la fecha de la prueba.

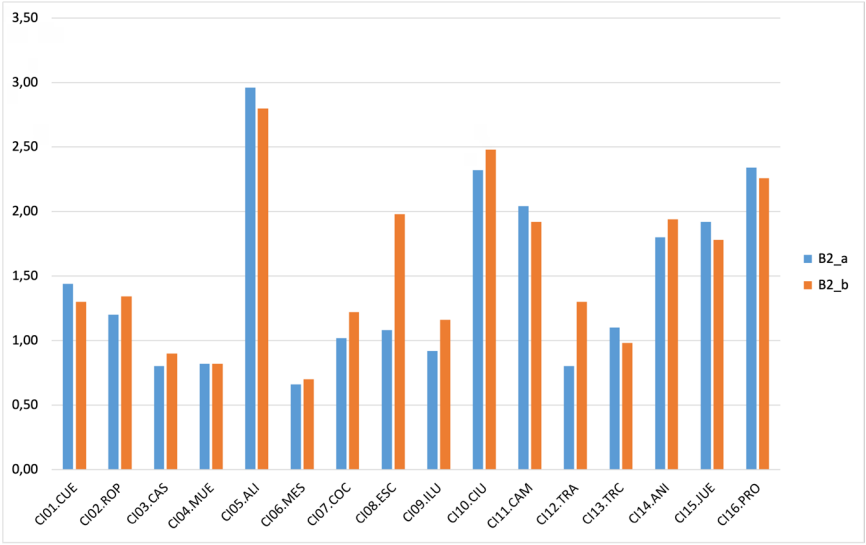


Gráfico 24. Promedio de vocablos por informante según la fecha de la prueba.

En la primera suministración los centros que superan la media son seis y en la segunda siete, se trata de los mismos si excluimos el CI08, "la escuela: muebles y materiales", que en B2_a se coloca en la posición 10.

| *B2_a* | | | CI | *B2_b* | | |
|---|---|---|---|---|---|---|
| **R** | **V** | **Pv/I** | | **R** | **V** | **Pv/I** |
| 1 | 148 | 2,96 | CI01.CUE | 1 | 140 | 2,80 |
| 2 | 117 | 2,34 | CI02.ROP | 2 | 124 | 2,48 |
| 3 | 116 | 2,32 | CI03.CAS | 3 | 113 | 2,26 |
| 4 | 102 | 2,04 | CI04.MUE | 4 | 99 | 1,98 |
| 5 | 96 | 1,92 | CI05.ALI | 5 | 97 | 1,94 |
| 6 | 90 | 1,80 | CI06.MES | 6 | 96 | 1,92 |
| 7 | 72 | 1,44 | CI07.COC | 7 | 89 | 1,78 |
| 8 | 60 | 1,20 | CI08.ESC | 8 | 67 | 1,34 |
| 9 | 55 | 1,10 | CI09.ILU | 9 | 65 | 1,30 |
| 10 | 54 | 1,08 | CI10.CIU | 10 | 65 | 1,30 |
| 11 | 51 | 1,02 | CI11.CAM | 11 | 61 | 1,22 |
| 12 | 46 | 0,92 | CI12.TRA | 12 | 58 | 1,16 |
| 13 | 41 | 0,82 | CI13.TRC | 13 | 49 | 0,98 |
| 14 | 40 | 0,80 | CI14.ANI | 14 | 45 | 0,90 |
| 15 | 40 | 0,80 | CI15.JUE | 15 | 41 | 0,82 |
| 16 | 33 | 0,66 | CI16.PRO | 16 | 35 | 0,70 |

Tabla 32. CI ordenados en función de su variedad léxica según la fecha de la prueba.

Seis CI se posicionan exactamente en los mismos rangos (CI05, "alimentos y bebidas"; CI02, "la ropa"; CI07, "la cocina y sus utensilios"; CI09, "iluminación y calefacción"; CI03, "partes de la casa (sin muebles)"; CI06, "objetos colocados en la mesa para la comida"). Los cambios más evidentes afectan a los centros CI08, "la escuela: muebles y materiales", que pasa del rango 10 al 4 –aumentando su riqueza del 83%– y CI12, "los medios de transporte" –que incrementa del 63% su variedad léxica– desplazándose del 15 al 10. Por otro lado, el CI13, "trabajos del campo y del jardín", baja de la posición 9 a la 13, disminuyendo del 12%.

### Índice de cohesión y densidad léxica

El corpus B2_b presenta un conjunto de palabras más cerrado con respecto al B2_a: el IC crece del 23,75%, al aumentar el nivel lingüístico aumentan las asociaciones mentales que permiten el alcance de un vocabulario más articulado.

| *B2_a* | | | CI | *B2_b* | | |
|---|---|---|---|---|---|---|
| **R** | **IC** | **D** | | **R** | **IC** | **D** |
| 1 | 0,208 | 10,38 | CI01.CUE | 1 | 0,242 | 12,08 |
| 7 | 0,137 | 7,12 | CI02.ROP | 5 | 0,177 | 8,87 |
| 5 | 0,149 | 7,75 | CI03.CAS | 4 | 0,187 | 9,36 |
| 4 | 0,161 | 8,39 | CI04.MUE | 3 | 0,217 | 10,83 |
| 9 | 0,103 | 5,36 | CI05.ALI | 9 | 0,121 | 6,07 |
| 3 | 0,164 | 8,52 | CI06.MES | 2 | 0,228 | 11,40 |
| 11 | 0,077 | 4 | CI07.COC | 14 | 0,094 | 4,72 |
| 6 | 0,146 | 7,61 | CI08.ESC | 8 | 0,139 | 6,96 |
| 14 | 0,071 | 3,67 | CI09.ILU | 10 | 0,117 | 5,83 |
| 10 | 0,095 | 4,96 | CI10.CIU | 11 | 0,115 | 5,73 |
| 15 | 0,063 | 3,29 | CI11.CAM | 16 | 0,084 | 4,19 |
| 2 | 0,198 | 10,30 | CI12.TRA | 6 | 0,161 | 8,03 |
| 16 | 0,051 | 2,64 | CI13.TRC | 15 | 0,085 | 4,24 |
| 8 | 0,120 | 6,24 | CI14.ANI | 7 | 0,150 | 7,52 |
| 12 | 0,074 | 3,82 | CI15.JUE | 13 | 0,109 | 5,44 |
| 13 | 0,071 | 3,70 | CI16.PRO | 12 | 0,111 | 5,56 |
| / | 0,118 | 6,11 | Promedio | / | 0,146 | 7,30 |

Tabla 33. Índice de cohesión y densidad léxica por CI según la fecha de la prueba.

En la segunda aplicación la cohesión aumenta en todos los campos, hasta alcanzar un +67% en el CI13, "trabajos del campo y del jardín". Las únicas excepciones son el CI12, "los medios de transporte", y el CI08, "la escuela: muebles y materiales", que bajan del 19% y 5%. Los informantes tienden a actualizar más unidades léxicas distintas a final del año académico que a comienzo. En lo que se refiere a la densidad léxica destaca la misma tendencia, es decir que en la segunda prueba los encuestados

repiten más palabras, hasta un +61% (en promedio +19,49%). De nuevo, encontramos dos casos aislados, el C112, "los medios de transporte", y el C108, "la escuela: muebles y materiales", donde hay más palabras diferentes: los valores bajan del 22% y del 9%.

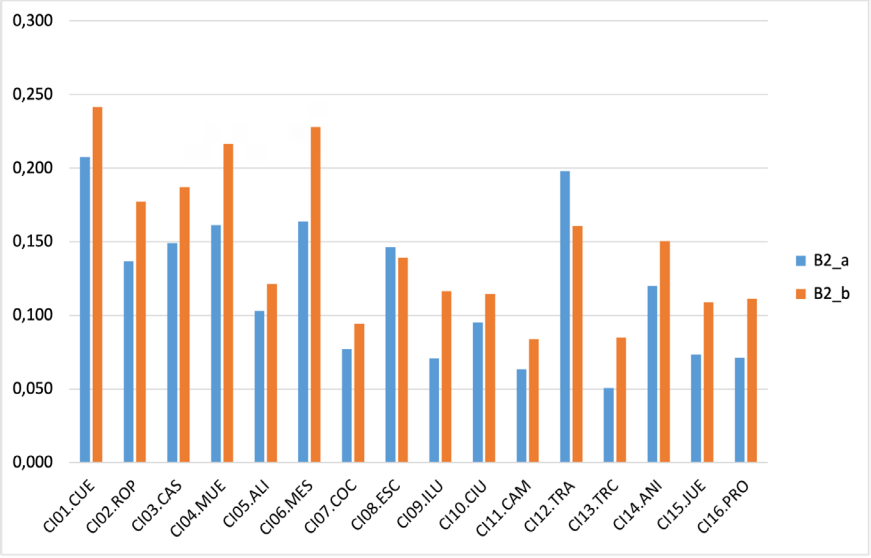


Gráfico 25. Índice de cohesión según la fecha de la prueba.

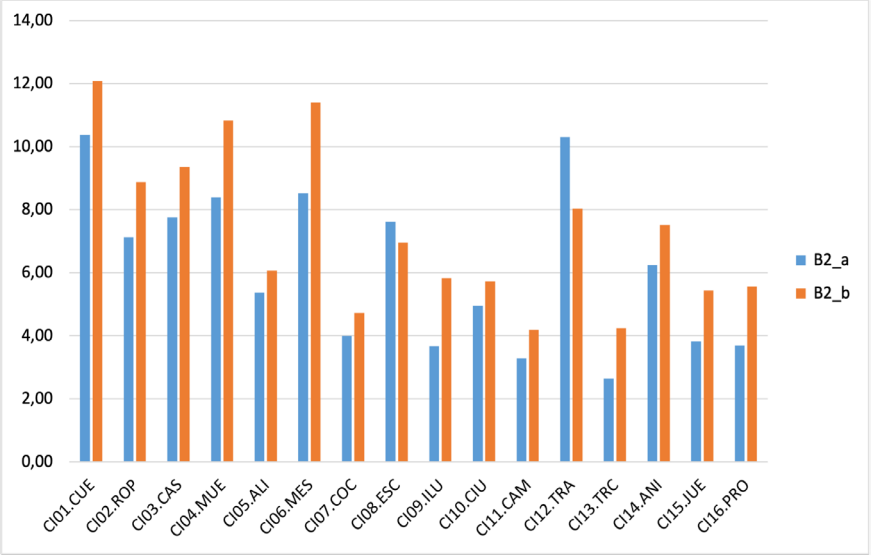


Gráfico 26. Densidad léxica según la fecha de la prueba.

### 3.1.4 Análisis comparativo

Pese a que «la disponibilidad léxica es una medida de las diferencias culturales» (Mackey 1971: 26-27), para terminar el análisis cuantitativo, cotejamos en esta última sección los resultados de nuestros informantes con los de otros grupos de aprendientes de ELE. Averiguamos si estudiantes que comparten el nivel lingüístico tienen la misma competencia, aunque tengan lenguas maternas diferentes y vivan en entornos culturales y escolares distintos (aprendizaje en el país de origen o en contexto de inmersión). Durante la fase de selección hemos tenido en cuenta exclusivamente trabajos realizados con estudiantes universitarios de nivel intermedio según nuestra tipología de informante.[34] Por eso incluimos las siguientes investigaciones, aun a sabiendas de que no todas se centran en los mismos CI, sin embargo, hemos podidos realizar la comparación entre nueve[35]:

- Carcedo (2000c), 50 informantes finlandeses, hablantes de finés y sueco.
- Samper Hernández (2002), 14 aprendientes en inmersión en la Universidad de Salamanca, cuyas lenguas maternas son inglés, italiano y japonés.
- Šifrar Kalan (2014), 100 informantes eslovenos.
- Sánchez-Saus (2016), 125 estudiantes en las Universidades de Andalucía, nativos de alemán, finés, francés, inglés, italiano y polaco.
- Del Barrio y Vann (2018), 68 estudiantes universitarios italianos.
- Hidalgo (2019), 294 encuestados sinohablantes, hablantes de chino mandarín y cantonés.

[34] Algunos autores trabajan también con alumnos de la educación secundaria, pero a la hora de confrontar los resultados recopilamos solo los datos relativos a los universitarios.

[35] Nuestros centros coinciden con los de Carcedo (2000c), Samper Hernández (2002), Del Barrio y Vann (2018). Por otro lado, Šifrar Kalan (2014) analiza once áreas, de las que descartamos "acciones que se realizan todos los días" y "la casa" que separamos en dos. Sánchez-Saus (2016) trabaja con dieciocho campos semánticos, entre los cuales detectamos doce que corresponden con los nuestros, pese a la denominación distinta ("el cuerpo humano"; "la casa"; "escuela y universidad"; "ocio y tiempo libre"; "profesiones y trabajos"). Hidalgo (2019) estudia dieciocho temas, diez coinciden con los nuestros, no obstante la designación sea diferente ("el cuerpo humano"; "la ropa y complementos"; "la casa"; "comidas y bebidas"; "escuela y universidad"; "ocio y tiempo libre"; "profesiones y trabajos").

Ahora bien, sobresale enseguida que los tamaños muéstrales cambian entre una y otra, por lo que utilizamos los valores promediales.

| | Investigación | Informantes | País | CI | Pp | Pv |
|---|---|---|---|---|---|---|
| 1 | Carcedo (2000c) | 50 | Finlandia | 16 | 162,12 | 44,36 |
| 2 | Samper Hernández (2002) | 14 | España | 16 | 174,92 | 65,07 |
| 3 | Šifrar Kalan (2014) | 100 | Eslovenia | 11 | 174,24 | 28,22 |
| 4 | Sánchez-Saus (2016) | 125 | España | 18 | 285,66 | 47,17 |
| 5 | Del Barrio y Vann (2018) | 68 | Italia | 16 | 145,93 | 26,41 |
| 6 | Hidalgo (2019) | 294 | China | 18 | 284,99 | 42,60 |
| 7 | Nalesso (2019a) | 100 | Italia | 16 | 125,79 | 14,90 |

Tabla 34. Datos de DL en las investigaciones cotejadas.

### Número de palabras

En general, todos los encuestados aportan la cantidad mayor de palabras en el CI05, "alimentos y bebidas", que se confirma el estímulo más rentable; el número menor de aportaciones se registra en CI12, "los medios de transporte", y CI02, "la ropa".

En orden, los encuestados que escriben más palabras son los de Sánchez-Saus (2016) con una media de 16,57; Hidalgo (2019) con 16,01; Šifrar Kalan (2014) con 15,73; Samper Hernández (2002) con 13,99; Carcedo (2000c) con 12,80; Del Barrio y Vann (2018) con 11,16; Nalesso (2019a) con 9,94.

| CI | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| CI01.CUE | 16,22 | 16 | 19,07 | 18,34 | 14,51 | 17,29 | 13,41 |
| CI02.ROP | 12,18 | 11,85 | 11,96 | 12,57 | 10,44 | 11,90 | 7,91 |
| CI03.CAS | 9,66 | 8,85 | 18,55 | 18,12 | 8,12 | 16,74 | 6,10 |
| CI05.ALI | 18,14 | 21,21 | 20,11 | 22,65 | 14,94 | 18,06 | 14,63 |
| CI10.CIU | 12,46 | 19,92 | 16,44 | 19,18 | 12,54 | 16,73 | 11,71 |
| CI12.TRA | 10,26 | 11,35 | 12,23 | 13,08 | 9,28 | 13,38 | 8,08 |
| CI14.ANI | 15,02 | 12,35 | 15,19 | 13,38 | 11,51 | 15,58 | 10,80 |
| CI15.JUE | 8,76 | 12,14 | 14,06 | 18,27 | 8,66 | 17,88 | 7,84 |
| CI16.PRO | 12,54 | 12,21 | 13,96 | 13,58 | 10,44 | 16,54 | 8,94 |
| Promedio | 12,80 | 13,99 | 15,73 | 16,57 | 11,16 | 16,01 | 9,94 |

Tabla 35. Promedios de palabras en las investigaciones cotejadas.

Los informantes de Sánchez-Saus (2016) predominan en tres centros, quizá se deba al contexto de aprendizaje en inmersión que podría conllevar una ventaja con respecto a los otros aprendices, aunque no hay esta supremacía en Samper Hernández (2002), cuyos alumnos se encontraban también en España. Los encuestados por Hidalgo (2019) aportan el número más alto de palabras en tres campos. No sorprenden los resultados alcanzados en Šifrar Kalan (2014), ya que sus estudiantes tienen un nivel B2+, ligeramente mayor con respecto a los demás. A continuación, se colocan los finlandeses de Carcedo (2000c), los italianos de Del Barrio y Vann (2018) y los nuestros que no sobresalen en ningún CI.

Es interesante observar el promedio de palabras por informante y compararlo con el léxico disponible de hispanohablantes nativos, que suelen superar las 20 respuestas por sujeto (Samper Hernández 2002). No extrañan los valores más bajos, derivados de la menor competencia de los participantes al ser estudiantes extranjeros de nivel intermedio. De todos modos, merece la pena subrayar que tres grupos dominan en el CI05, "alimentos y bebidas", aportando un promedio de más de 20 palabras (Samper Hernández 2002, Sánchez-Saus, 2016, Šifrar Kalan 2014). Los informantes de Carcedo (2000c) e Hidalgo (2019) también arrojan cantidades apreciables, si bien se queden por debajo de las 20 unidades.

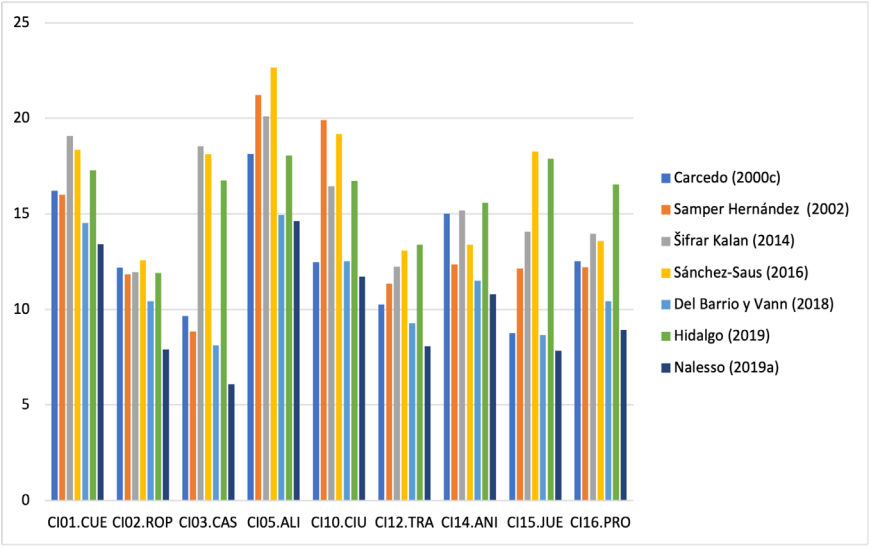


Gráfico 27. Promedios de palabras por informante en las investigaciones cotejadas.

A continuación, vemos los descriptores estadísticos que presentan la oscilación intra e intergrupal.

| Investigación | | Media | Mediana | Mínimo | Máximo |
|---|---|---|---|---|---|
| 1 | Carcedo (2000c) | 11,52 | 12,46 | 8,76 | 18,14 |
| 2 | Samper Hernández (2002) | 12,59 | 12,21 | 8,85 | 21,21 |
| 3 | Šifrar Kalan (2014) | 14,16 | 15,19 | 11,96 | 20,11 |
| 4 | Sánchez-Saus (2016) | 14,92 | 18,12 | 12,57 | 22,65 |
| 5 | Del Barrio y Vann (2018) | 10,04 | 10,44 | 8,12 | 14,94 |
| 6 | Hidalgo (2019) | 14,41 | 16,73 | 11,90 | 18,06 |
| 7 | Nalesso (2019a) | 8,94 | 8,94 | 6,10 | 14,63 |

Tabla 36. Estadísticos descriptivos en las investigaciones cotejadas.

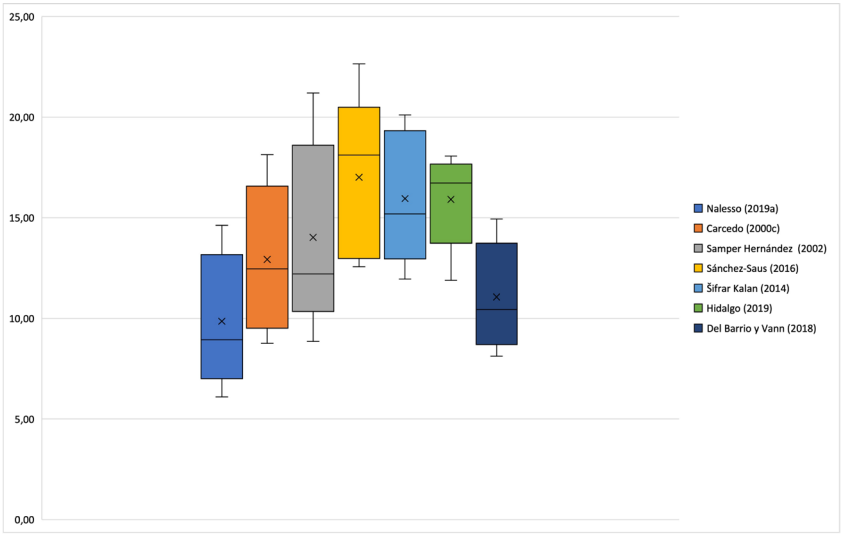


Gráfico 28. Diagrama de cajas para las investigaciones cotejadas.

Samper Hernández (2002) revela el intervalo intercuartil más elevado, se pasa de un valor mínimo de 8,85 a un máximo de 21,21 y la desviación mayor se computa entre la mediana (12,21) y el Q3 (16), donde se coloca también el promedio de palabras por informante (12,59).

Se detecta una gran diferencia entre los cuartiles segundo y tercero en Carcedo (2000c), Šifrar Kalan (2014), Del Barrio y Vann (2018), Nalesso (2019a) ya que las cajas son más amplias con respecto a las que se posicionan por debajo. Además, en Carcedo (2000c), Del Barrio y Vann (2018) y Nalesso (2019a) contamos con una simetría difusa en las cuatro partes del diagrama que no se encuentra en otros estudios. Al contrario, en Sánchez-Saus (2016) e Hidalgo (2019) la desviación máxima está entre los intervalos Q2-Q1 y Q1-mínimo, indicado por el brazo inferior (respectivamente 11,96 y 11,90).

### Número de vocablos

Los campos más ricos de vocablos son el CI05, "alimentos y bebidas"; el CI10, "la ciudad"; el CI16, "profesiones y oficios"; el CI15, "juegos y distracciones". Los que tienen menor variación léxica son el CI01, "partes del cuerpo"; el CI02, "la ropa", y el CI12, "medios de transporte". Los aprendientes que aportan el mayor número de vocablos son en orden decreciente los de Samper Hernández (2002) con una media de 4,83; Carcedo (2000c) con 3,39; Šifrar Kalan (2014) con 2,50; Sánchez-Saus (2016) con

2,48; Hidalgo (2019) con 1,95; Del Barrio y Vann (2018) con 1,91; Nalesso (2019a) con 1,11.

| CI | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| CI01.CUE | 2,40 | 3,35 | 1,45 | 1,71 | 1,07 | 1,21 | 0,91 |
| CI02.ROP | 2,26 | 3,07 | 1,93 | 1,55 | 1,50 | 1,09 | 0,72 |
| CI03.CAS | 2,06 | 3,14 | 2,85 | 2,22 | 0,93 | 1,90 | 0,47 |
| CI05.ALI | 5,44 | 7,92 | 3,28 | 2,90 | 2,94 | 2 | 1,75 |
| CI10.CIU | 4,50 | 7,64 | 3,43 | 3,44 | 2,47 | 3,13 | 1,57 |
| CI12.TRA | 2 | 2,64 | 1,97 | 2,08 | 1,25 | 1,99 | 0,58 |
| CI14.ANI | 3,38 | 3,28 | 2,13 | 1,79 | 1,66 | 1,40 | 1,02 |
| CI15.JUE | 3,68 | 6 | 2,87 | 3,28 | 2,97 | 2,48 | 1,31 |
| CI16.PRO | 4,76 | 6,42 | 2,63 | 3,38 | 2,40 | 2,38 | 1,62 |
| Promedio | 3,39 | 4,83 | 2,50 | 2,48 | 1,91 | 1,95 | 1,11 |

Tabla 37. Promedios de vocablos en las investigaciones cotejadas.

Se registra un cambio de tendencia con respecto a la capacidad productiva de los CI, como ilustra el gráfico: el grupo de Salamanca predomina en todas las áreas, excepto en el CI14, "los animales", donde los finlandeses arrojan un valor levemente más alto y se colocan en segunda posición, superando los resultados de los estudiantes de Andalucía, eslovenos, chinos e italianos.

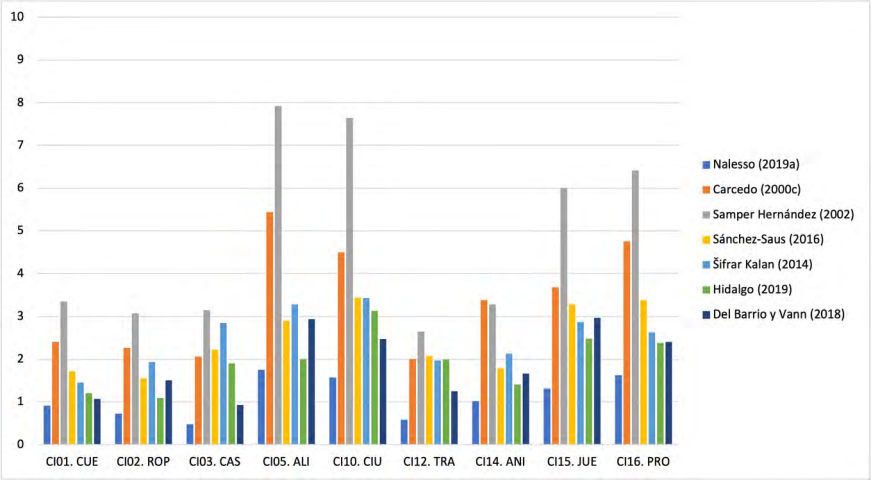


Gráfico 29. Promedios de vocablos por informante en las investigaciones cotejadas.

### Índice de cohesión

Las respuestas de los encuestados de Samper Hernández (2002) y Carcedo (2000c) son más compactas frente a las de otros (0,23 y 0,17). Siguen, por orden decreciente, italianos (0,10) y eslovenos (0,07). Sánchez-Saus (2016) e Hidalgo (2019) presentan índices que se alejan mucho de 1 (0,06 y 0,03), con lo cual la producción se manifiesta abierta y parece que cada sujeto ha actualizado unidades diferentes de sus propios compañeros.

| CI | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| CI01.CUE | 0,27 | 0,34 | 0,13 | 0,09 | 0,20 | 0,05 | 0,15 |
| CI02.ROP | 0,22 | 0,28 | 0,06 | 0,06 | 0,10 | 0,04 | 0,11 |
| CI03.CAS | 0,18 | 0,20 | 0,07 | 0,07 | 0,13 | 0,03 | 0,13 |
| CI05.ALI | 0,14 | 0,19 | 0,06 | 0,06 | 0,07 | 0,03 | 0,08 |
| CI10.CIU | 0,11 | 0,19 | 0,05 | 0,04 | 0,07 | 0,02 | 0,08 |
| CI12.TRA | 0,21 | 0,31 | 0,06 | 0,05 | 0,11 | 0,02 | 0,14 |
| CI14.ANI | 0,18 | 0,27 | 0,07 | 0,06 | 0,10 | 0,04 | 0,11 |
| CI15.JUE | 0,09 | 0,14 | 0,05 | 0,04 | 0,04 | 0,02 | 0,06 |
| CI16.PRO | 0,11 | 0,14 | 0,05 | 0,03 | 0,06 | 0,02 | 0,06 |
| Promedio | 0,17 | 0,23 | 0,07 | 0,06 | 0,10 | 0,03 | 0,10 |

Tabla 38. Índice de cohesión en las investigaciones cotejadas.

El CI01, "partes del cuerpo", es el más cerrado: esto podría deberse a un aprendizaje sistematizado mediante el cual los alumnos estudiaron este tema. Parece que los campos que se presentan en el aula dan lugar a un conocimiento más compacto, tienen un grado de cohesión semántica más alto con respecto a los que no se trabajan, o que se trabajan menos. En el caso opuesto, los centros CI15, "juegos y distracciones", y CI16, "profesiones y oficios", son más difusos, quizás porque cada sujeto activa más asociaciones mentales en virtud de su experiencia personal. Asimismo, es posible que en ciertos temas los índices cambien a causa de la realidad socioambiental en la que viven los informantes (Gómez Molina y Gómez Devís 2004: 83-84).

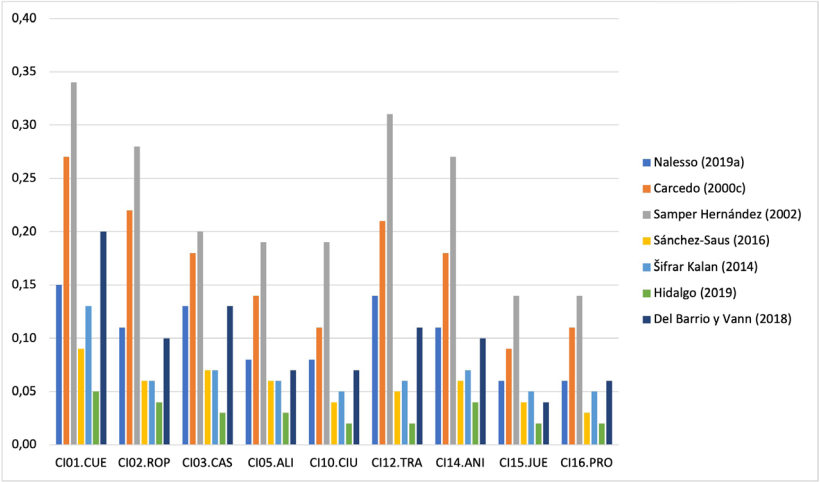


Gráfico 30. Índice de cohesión en las investigaciones cotejadas.

## 3.2 Estudio cualitativo

La segunda parte del capítulo aborda el análisis cualitativo de nuestro corpus: estudiamos el tipo de léxico activado presentando los vocablos más disponibles, que alcanzan un ID igual o superior a 0,1. En este sentido, averiguamos las temáticas y categorías gramaticales más difundidas en cada centro.

Asimismo, computamos la cardinalidad del conjunto (López Chávez 1992), definida también índice de compatibilidad (Ávila Muñoz y Sánchez Sáez 2010, 2011), con el propósito de medir la correspondencia existente en las respuestas aportadas por las variantes de una misma variable y por los grupos B2_a y B2_b.

A continuación, cotejamos el léxico disponible de nuestros informantes con dos corpus de referencia: lo comparamos con las primeras 5.000 palabras[36] más frecuentes del español procedentes del *Corpus de referencia del español actual* de la RAE (CREA) para verificar si están incluidos en este banco de datos y, en segunda instancia, con el *Corpus de aprendices del español* del Instituto Cervantes (CAES) para descubrir si coinciden con las palabras más conocidas y utilizadas por otros aprendientes de ELE.

[36] Con esta cantidad es posible desenvolverse en situaciones comunicativas cotidianas según Alvar Ezquerra (2003: 100).

### 3.2.1 Análisis transversal: resultados generales

El análisis transversal se abre con el estudio de los listados generales del léxico más disponible por centro de interés.[37] Estudiamos, a la vez, cuántos vocablos con ID ≥ 0,1 presenta cada centro y cuáles son las asociaciones más activadas y las clases gramaticales predominantes.

#### C101, "partes del cuerpo"

Los vocablos más disponibles en este campo son diecinueve: *ojos*, *mano*, *cabeza*, *pie*, *pierna*, *brazo*, *boca*, *dedo*, *nariz*, *pelo*, *cara*, *espalda*, *cuello*, *orejas*, *rodilla*, *dientes*, *lengua*, *barriga*, *uñas*. Son sustantivos que se adscriben al estímulo prototípico nocional y se refieren a la parte superior del cuerpo humano.

La categoría predominante es la nominal (71%), pero hay también otras clases y temáticas ligadas a un estímulo secundario: detectamos el 11% de adjetivos relacionados con el aspecto físico (*rubio*, *azul*, *pelirrojo*, *guapo*, *delgado*); el 5% de verbos y sentidos (*oír*, *oído*, *ver*, *vista*, *respirar*, *tacto*); el 12% de sustantivos referentes a la biología (*sangre*, *neurona*, *aparato*). La presencia de unidades multipalabra se limita a una, *cuerpo humano*.

#### C102, "la ropa"

Entre los vocablos con ID ≥ 0,1 aparecen sustantivos referidos a las prendas de vestir y a los complementos, que se encuentran en gran cantidad en todo el listado. En total registramos doce elementos más disponibles: *camiseta*, *zapatos*, *pantalones*, *camisa*, *jersey*, *chaqueta*, *falda*, *vaqueros*, *bufanda*, *calcetines*, *vestido*, *abrigo*.

En los rangos inferiores detectamos un porcentaje escaso de verbos (*ponerse*, *vestirse*) y de otros sustantivos adscritos a diferentes campos semánticos, como los materiales (*seda*, *algodón*). Además, apreciamos la activación de algunas locuciones sustantivas: *gafas de sol*, *lente de contacto*, *ropa interior*, *traje de baño*, *zapatilla de deporte*.

[37] Si no se indica de manera diferente, alistamos en cada centro de interés los vocablos en orden decreciente según el ID obtenido.

### c103, "partes de la casa (sin muebles)"

Los vocablos más disponibles son siete: *cocina*, *cuarto de baño*, *habitación*, *salón*, *jardín*, *cuarto*, *garaje*. De nuevo se trata de sustantivos –la única categoría gramatical presente en este CI– y una sola locución, a la que se añaden otras en rangos más bajos: *cuarto de estar*, *cuarto trasero*, *cuarto matrimonial*, *salón de baile*. Los informantes han activado también palabras relacionadas con las habitaciones y con las partes estructurales de la casa (*pared*, *muro*, *techo*, *puerta*, *portal*, *ventana*, *pasillo*, *escalera*, *corredor*).

### c104, "los muebles de la casa"

Aquí son diez los ítems que superan o igualan el ID preestablecido, entre ellos percatamos una unidad pluriverbal, la única que encontramos en el listado: *cama*, *silla*, *mesa*, *sofá*, *armario*, *sillón*, *ducha*, *lámpara*, *televisión*, *mesilla de noche*.

Todos los vocablos son sustantivos vinculados a temáticas distintas, ya que concedimos una mayor amplitud de las relaciones asociativas: los informantes han escrito unidades que indican electrodomésticos, complementos y decoraciones (*televisión*, *lavadora*, *horno*, *frigorífico*, *lavavajillas*, *cocina*, *refrigerador*, *cuadro*, *alfombra*, *cortina*, *almohada*).

### c105, "alimentos y bebidas"

Este campo es el que cuenta con más vocablos con ID ≥ 0,1 (son veinticinco): *agua*, *cerveza*, *tomate*, *leche*, *manzana*, *naranja*, *vino*, *paella*, *jamón*, *pescado*, *pan*, *arroz*, *carne*, *zumo*, *pasta*, *café*, *bocadillo*, *azúcar*, *huevo*, *queso*, *patata*, *chocolate*, *sangría*, *plátano*, *sal*.

Encontramos asociaciones con comida y bebida que hacen lo propio bajo la forma de sustantivos (93%). Notamos productos típicos de la cocina española e hispanoamericana (*paella*, *jamón*, *sangría*), que aparecen también en posiciones más bajas (*aguacate*, *calimocho*, *frijoles*, *gazpacho*, *palta*, *papa*, *quesadilla*, *tapas*, *tortilla*). Destacamos la presencia de técnicas de cocción (*asado*, *flambeado*, *parrilla*), condimentos y aliños (*sal*, *azúcar*, *aceite*, *vinagre*, *orégano*, *azafrán*).

Extraña el hecho de que en un campo como este no haya aparecido ningún verbo, mientras que se han actualizado varios adjetivos (*dulce*, *salado*, *amargo*, *animal*, *vegetal*) y locuciones (7% del listado): *arroz con leche*, *crema catalana*, *piña colada*, *dulce de leche*, *agua sin gas*, *agua con gas*, *vino tinto*, *vino blanco*, *tinto de verano*, *zumo de naranja*, *zumo de fruta*, *pechuga de pollo*.

Para terminar, aceptamos *Coca-Cola* siendo un ejemplo prototípico de lexicalización de una marca comercial.

### c106, "objetos colocados en la mesa para la comida"

Los vocablos más disponibles son seis sustantivos que designan los objetos más comunes que se utilizan para comer: *plato*, *vaso*, *cuchillo*, *cuchara*, *botella*, *tenedor*.

Si observamos todo el listado, el 91% está formado por nombres de comida y bebida (*agua*, *alimento*, *comida*), utensilios (*mantel*, *olla*, *botella*, *bandeja*, *copa*, *tenedor*, *servilleta*, *jarra*), condimentos (*aceite*, *sal*, *pimienta*, *vinagre*) y otras lexías (*desayuno*, *mesa*, *comedor*). Contamos también con colocaciones (*plato hondo*, *plato llano*, *vino tinto*, *vino blanco*) y verbos (*cambiar*, *poner*, *tomar*, *pulir*).

### c107, "la cocina y sus utensilios"

Los lexemas más disponibles son ocho: *horno*, *plato*, *cuchillo*, *cuchara*, *sartén*, *vaso*, *frigorífico*, *nevera*. En lo que se refiere a la categoría gramatical predominante, los sustantivos cubren el 92% de la lista frente al 8% de verbos (*abrir*, *cocinar*, *enharinar*, *apagar*, *mezclar*). Hemos encontrado la locución *libro de cocina* y las unidades compuestas *lavaplatos* y *lavavajillas*. Aparecen también electrodomésticos y utensilios a lo largo de toda la lista, como en el centro anterior. Según Hidalgo (2019: 262) se trata de un fenómeno de índole psicolingüística definido «aprovechamiento léxico», para el cual los encuestados se valen de las respuestas aportadas antes. De ahí que se puedan encontrar secuencias formadas por unidades léxicas idénticas, incluso en el mismo orden. Este fenómeno suele producirse cuando los informantes no saben qué palabras escribir porque apenas conocen vocablos relacionados con un tema o ya han escrito todos los que saben.

### c108, "la escuela: muebles y materiales"

Las unidades con ID ≥ 0,1 se adscriben principalmente al segundo estímulo semántico, los materiales escolares: *libro*, *bolígrafo*, *lápiz*, *cuaderno*, *mochila*, *ordenador*, *papel*, *borrador*.

El 7% de la lista se compone de verbos (*escribir*, *ir*, *pinchar*, *subrayar*, *volver*) y el 93% de sustantivos. Si desglosamos este último conjunto contamos con la siguiente distribución de temáticas: 68% materiales (*goma de*

*borrar*, *marcador*, *proyector*, *registro*, *sacapuntas*, *papel*, *mapa*), 15% muebles (*armario*, *cátedra*, *mesa*, *perchero*, *pupitre*, *reloj*, *silla*), 7% personas (*alumno*, *clase*, *compañero*, *estudiante*, *profesor*), 2% misceláneo (*aula*, *luz*, *edificio*).

### C109, "iluminación y calefacción"

En este campo son tres los vocablos más disponibles (*lámpara*, *luz*, *fuego*) y presentan un ID notablemente elevado con respecto a lo que se posiciona después, *sol*, que en términos de frecuencia de aparición alcanza 12 ocurrencias mientras que estas logran 80, 59 y 39 respectivamente.

En general, se trata de sustantivos (91%) que responden a sendos estímulos del CI. En la mayoría de los casos, son nombres que pertenecen al ámbito de la calefacción (*calentador*, *termosifón*, *temperatura*, *calorífico*, *calefacción*, *chimenea*, *estufa*, *ventilador*, *leña*). Los dos adjetivos detectados, *caliente* y *frío*, cubren el 4% de la lista y tampoco los verbos son muchos, forman el 5% (*calentar*, *apagar*, *encender*).

Rastreamos vocablos relacionados con otras temáticas (*electricidad*, *pantalla*, *noche*, *estrella*, *circuito*, *enchufe*) que admitimos debido a la dificultad que, a nuestro juicio, conlleva este centro para los informantes. El análisis revela también la actualización de unidades multipalabra: *aire acondicionado*, *energía eléctrica*, *energía renovable*, *hacer frío*, *hacer calor*, *central nuclear*.

### C110, "la ciudad"

Este centro de interés es uno de los más ricos en la producción de vocablos (en total 157), los más disponibles son diecisiete y comprenden nombres de edificios y construcciones de la ciudad (*escuela*, *casa*, *edificio*, *iglesia*, *plaza*, *estación*, *ayuntamiento*, *árbol*), medios de transporte (*coche*, *autobús*, *tren*), servicios (*tienda*, *parque*, *bar*, *restaurante*), vías públicas (*calle*, *carretera*). En la lista aparecen sinónimos, hiperónimos, hipónimos, merónimos y holónimos de estas lexías. Es más, los informantes aportan sustantivos referidos a habitantes, diversión, tiendas, restaurantes, lugares de trabajo y otros vocablos (*centro*, *policía*, *fuente*, *río*, *aeropuerto*, *derecha*, *cartel*, *muralla*, *lampión*, *playa*, *rincón*, *sitio*, *luz*, *arte*, *billete*, *flor*). Apreciamos la presencia de la palabra *carro*, típica de la variedad diatópica hispanoamericana.

En lo referente a las categorías gramaticales –además de los sustantivos– las locuciones abarcan el 6% de la lista, los verbos el 9% y los adjetivos el 3%.

### C11, "el campo"

Los informantes asocian este estímulo con las personas que viven en el campo y con sus oficios, lugares, animales y productos más característicos, los medios de transporte que allí se utilizan, el clima y tiempo atmosférico. Además, han activado algunas relaciones secundarias vinculadas con sensaciones (*tranquilidad*, *silencio*, *paz*, *libertad*), entre las cuales aparece la locución *al aire libre*. Contamos con pocos verbos (*pasear*, *cultivar*, *andar*, *correr*, *jugar*, *trabajar*) y adjetivos (*verde*, *fértil*).

Las nueve unidades más disponibles son sustantivos, perfectamente encajados en el ámbito semántico del campo: *árbol*, *animal*, *flor*, *tierra*, *hierba*, *vaca*, *campesino*, *planta*, *caballo*.

### C12, "medios de transporte"

La lista presenta diez elementos estrictamente ligados a los medios de transporte: *coche*, *autobús*, *tren*, *bicicleta*, *avión*, *taxi*, *metro*, *motocicleta*, *barco*, *tranvía*. En los rangos inferiores, donde se hallan vocablos menos disponibles, se añaden otras formas de desplazamiento (*a pie*, *caballo*, *patinetes*, *a nado*, *correr*, *caminar*, *andar*), títulos de viaje (*billete*), documentos (*carné de conducir*) y palabras relacionadas con el tráfico (*esmog*, *horario*, *parada*, *gasolina*).

El porcentaje de sustantivos arrojados en total es del 81%, dentro del cual hay americanismos: *carro*, *colectivo*, *guagua*, *máquina*. Los verbos cubren el 9%: *conducir*, *tomar*, *coger*, *correr*, *subir*, *esperar*, *viajar*, *caminar*, *ir*, *andar*. Percatamos la sigla AVE y algunas locuciones como *a pie*, *a nado*, *transporte público*, *carné de conducir*, *hacer dedo*, *globo aerostático*.

### C13, "trabajos del campo y del jardín"

En el conjunto total de vocablos arrojados, los que revelan un ID ≥ 0,1 son cuatro, entre los que aparecen los verbos *plantar* y *cortar*, además de los nombres *jardinero* y *campesino*. *Dar agua* es la única combinación que encontramos.

La mayoría de los lexemas pertenece al ámbito prototípico del centro (*cultivar*, *agricultor*, *regar*, *granjero*, *pastor*, *arar*, *recoger*, *cosechar*, *gau-*

*cho*), pero se detectan también otras asociaciones que incluyen animales (*gato*, *perro*, *vaca*, *oveja*, *burro*, *mariposa*), productos (*flor*, *fruta*, *verdura*, *mantequilla*, *aceite*), herramientas (*tijeras*, *arado*), lugares (*campo*, *jardín*).

### C114, "los animales"

Los doce vocablos más disponibles son sustantivos referidos a animales: *perro*, *gato*, *caballo*, *vaca*, *pez*, *pájaro*, *león*, *cerdo*, *gallina*, *tigre*, *conejo*, *toro*. Sorprende el hecho de que no haya surgido ninguna otra asociación, como por ejemplo sus partes del cuerpo, los lugares donde viven y productos de origen animal.

Apreciamos la aparición de los siguientes americanismos: *tigrillo*, *hicotea*, *chancho*, *cochino*, *alpaca* y de las colocaciones, *caballito de mar* y *estrella marina*.

### C115, "juegos y distracciones"

He aquí trece vocablos con ID ≥ 0,1 y corresponden al 10% de la lista (*fútbol*, *baloncesto*, *leer*, *libro*, *bailar*, *cartas*, *pelota*, *televisión*, *ordenador*, *videojuegos*, *cine*, *escuchar música*, *natación*) que, por lo general, presenta una gran variedad temática y tipológica. Apreciamos la activación de juguetes, deportes, lugares, personas y animales de compañía. Registramos *Monopoly* y *Play Station* al tratarse de lexicalización de una marca comercial.

En la mayoría de los casos se trata de sustantivos, pero contabilizamos un 27% de verbos (*correr*, *nadar*, *pasear*, *cantar*, *dibujar*, *salir*, *viajar*, *descansar*, *esquiar*, *chatear*, *escribir*, *patinar*, *fotografiar*), cantidad que destaca con respecto a otros CI. Asimismo, las unidades multipalabra alcanzan un porcentaje del 11%: *juego de mesa*, *parque de diversiones*, *instrumento musical*, *dibujos animados*, *soldadito de plomo*, *redes sociales*, *escuchar música*, *salir con amigos*, *ir de compras*, *quedar con amigos*, *pasárselo bien*, *ir de vacaciones*, *ir de tapas*, *ir de bares*.

### C116, "profesiones y oficios"

Las lexías más disponibles son sustantivos, como era de esperar: *profesor*, *médico*, *cocinero*, *abogado*, *doctor*, *camarero*, *fontanero*, *actor*, *estudiante*, *empleado*. Igualmente, en el listado total se observa un 98% de nombres comunes y solo un 2% de verbos.

Los nombres se reparten en dos macroáreas: la primera (94%) agrupa profesiones y profesionales (*enfermero*, *peluquero*, *traductor*, *canguro*, *escritor*, *jefe*, *periodista*, *albañil*, *sastre*, *cirujano*, *empresario*, *estilista*, *logopeda)*, que recopilamos en la forma masculina singular para dar uniformidad a la muestra, manteniendo los heterónimos (*actor*, *actriz*, *azafata*, *preste*, *cura*, *monja*); la segunda (4%) reúne lugares y tipos de trabajo (*oficina*, *banco*, *trabajo*, *tienda*, *estudio*, *escuela*, *fábrica*, *negocios*, *restaurante*, *hospital*). Notamos la presencia de los americanismos *plomero*, *mucamo*, *botón* y de las locuciones *guía de turismo*, *ama de casa*, *señora de la limpieza*, *cuidador de perro.*

### 3.2.2 Análisis transversal: resultados por variable

Aplicando el mismo protocolo realizamos el análisis de los resultados según las tres variables contempladas, al que añadimos el estudio de la cardinalidad del conjunto para establecer los porcentajes de compatibilidad entre las respuestas.

| | **Cardinalidad del conjunto** | | |
|---|---|---|---|
| **CI** | ***Sexo*** | ***Nivel de ELE*** | ***Conocimiento de otras LE*** |
| CI01.CUE | 65% | 77,27% | 68,18% |
| CI02.ROP | 69,23% | 76,92% | 69,23% |
| CI03.CAS | 77,78% | 55,56% | 77,78% |
| CI04.MUE | 50% | 72,73% | 50% |
| CI05.ALI | 55,17% | 60,61% | 60,71% |
| CI06.MES | 28,57% | 85,71% | 100% |
| CI07.COC | 30% | 30% | 37,50% |
| CI08.ESC | 63,64% | 100% | 81,82% |
| CI09.ILU | 0% | 100% | 100% |
| CI10.CIU | 40% | 59,09% | 63,16% |
| CI11.CAM | 33,34% | 43,75% | 60% |
| CI12.TRA | 100% | 100% | 90% |
| CI13.TRC | 75% | 50% | 50% |

| | | | |
|---|---|---|---|
| CI14.ANI | 42,86% | 64,71% | 66,67% |
| CI15.JUE | 37,50% | 55,56% | 58,82% |
| CI16.PRO | 33,34% | 46,67% | 53,85% |

Tabla 39. Cardinalidad del conjunto según variables.

**Variable sexo**

**CI01, "partes del cuerpo"**

Los lexemas más disponibles compartidos son trece (*mano*, *cabeza*, *ojos*, *pie*, *boca*, *pelo*, *brazo*, *dedo*, *rodilla*, *cara*, *pierna*, *nariz*, *espalda*). Las mujeres arrojan seis unidades que los hombres no escriben o que en su lista no alcanzan el ID requerido (*orejas*, *cuello*, *dientes*, *lengua*, *barriga*, *uñas*). Ellos presentan una sola palabra exclusiva (*cejas*). De este modo la cardinalidad computada es del 65%.

**CI02, "la ropa"**

En este campo se produce una coincidencia de nueve ítems (*camiseta*, *zapatos*, *camisa*, *pantalones*, *calcetines*, *vaqueros*, *bufanda*, *chaqueta*, *jersey*), las primeras posiciones de las listas coinciden (*camiseta*, *zapatos*). Las mujeres presentan trece vocablos con ID ≥ 0,1 (añaden *falda*, *abrigo*, *vestido*, *traje* que no están en la lista de los hombres, formadas por los elementos comunes). A partir de estos valores, la compatibilidad es del 69,23%.

**CI03, "partes de la casa (sin muebles)"**

La cardinalidad aquí es del 77,78%: todos los vocablos con ID mayor o igual a 0,1 aportados por las mujeres (*cocina*, *cuarto de baño*, *habitación*, *salón*, *jardín*, *cuarto*, *garaje*) corresponden a los de los hombres, que presentan dos exclusivos (*escalera*, *puerta*).

**CI04, "los muebles de la casa"**

Los vocablos compartidos son cinco (*cama*, *armario*, *sofá*, *silla*, *mesa*); las mujeres activan otras cinco unidades con ID ≥ 0,1 sobrepasando a sus

compañeros (*sillón*, *ducha*, *lámpara*, *televisión*, *mesilla de noche*). La cardinalidad computada es del 50%.

### C105, "alimentos y bebidas"

Detectamos un desnivel notable entre el número de respuestas de los dos conjuntos que alcanzan el ID requerido: veintisiete para las mujeres y dieciocho para los hombres. Las primeras actualizan once vocablos que ellos no tienen (*zumo*, *café*, *azúcar*, *huevo*, *bocadillo*, *queso*, *chocolate*, *patata*, *sal*, *tapas*, *aceite*). Ellos, por su parte, arrojan dos divergentes (*limón* y *Coca-Cola*). Los lexemas compartidos son: *agua*, *cerveza*, *vino*, *naranja*, *paella*, *pan*, *carne*, *pescado*, *tomate*, *leche*, *jamón*, *arroz*, *manzana*, *sangría*, *pasta*, *plátano*. Estos datos suponen un índice de cardinalidad del 55,17%.

### C106, "objetos colocados en la mesa para la comida"

La compatibilidad aquí es inferior con respecto a lo visto hasta ahora, 28,57%, pese a que se haya registrado cierta correspondencia en las primeras posiciones de las listas. Las mujeres sobrepasan a sus compañeros en quince elementos (*cucharilla*, *servilleta*, *sal*, *aceite*, *mantel*, *copa*, *taza*, *jarra*, *ensaladera*, *agua*, *sartén*, *plato llano*, *bebida*, *plato hondo*, *cucharada*), en consecuencia, las lexías compartidas son seis y coinciden con la entera lista masculina (*plato*, *cuchillo*, *vaso*, *cuchara*, *tenedor*, *botella*).

### C107, "la cocina y sus utensilios"

He aquí tres coincidencias entre las variantes (*horno*, *nevera*, *plato*): el grupo femenino tiene cinco palabras que no están en la lista del masculino (*cuchillo*, *cuchara*, *vaso*, *frigorífico*, *sartén*), donde se hallan dos unidades que las mujeres no presentan entre las más disponibles (*refrigerador*, *olla*). Por esta razón, hay una cardinalidad del conjunto del 30%.

### C108, "la escuela: muebles y materiales"

Se pone de relieve una coincidencia de siete vocablos, a saber, todas las lexías presentes en la lista de los hombres (*libro*, *lápiz*, *cuaderno*, *bolígrafo*, *silla*, *mochila*, *pizarra*). Las mujeres, por su parte, tienen cuatro vocablos divergentes (*mesa*, *ordenador*, *papel*, *borrador*), lo que lleva a una compatibilidad del 63,64%.

**C109, "iluminación y calefacción"**

No computamos correspondencias entre los vocablos: los hombres aportan en su listado dos ítems que superan o igualan el ID de 0,1 (*lámpara*, *luz*) mientras que las compañeras no presentan ninguno que tiene la característica necesaria, por eso la cardinalidad es 0%.

**C110, "la ciudad"**

En este campo se produce un total de ocho coincidencias: *casa*, *plaza*, *calle*, *iglesia*, *parque*, *edificio*, *autobús*, *estación*. Las mujeres escriben un total de once vocablos exclusivos con ID ≥ 0,1 (*coche*, *tienda*, *escuela*, *bar*, *carretera*, *restaurante*, *árbol*, *tren*, *museo*, *ayuntamiento*, *universidad*) mientras que los varones ofrecen solo un divergente (*centro*). Basándonos en estos datos, la compatibilidad es del 40%.

**C111, "el campo"**

El índice llega al 33,34% desde el momento en el que los vocablos compartidos son cuatro (*árbol*, *vaca*, *animal*, *campesino*); las mujeres activan siete exclusivos (*flor*, *tierra*, *hierba*, *planta*, *caballo*, *naturaleza*, *casa*) y los hombres uno (*gallina*).

**C112, "medios de transporte"**

El índice es de gran interés en este estímulo, ya que por primera vez el número de vocablos más disponibles coincide en los sociolectos que presentan las mismas unidades: *coche*, *autobús*, *tren*, *bicicleta*, *avión*, *metro*, *motocicleta*, *taxi*, *tranvía*, *barco*. Se alcanza el 100% de compatibilidad y, es más, los lexemas de los rangos 1-5 se corresponden perfectamente.

**C113, "trabajos del campo y del jardín"**

Los vocablos de los hombres (*campesino*, *jardinero*, *cortar*) aparecen en la lista de las mujeres, que proporcionan una sola unidad adicional (*plantar*), lo que lleva a computar una cardinalidad del 75%.

**C114, "los animales"**

Las unidades compartidas coinciden con la lista de vocablos más disponibles de los participantes masculinos, formada por seis elementos: *perro*, *gato*, *caballo*, *cerdo*, *vaca*, *pez*. Por su parte, el listado de las mujeres

tiene ocho unidades: *león*, *pájaro*, *gallina*, *conejo*, *tigre*, *toro*, *mariposa*, *elefante*. Los resultados de los cálculos revelan una cardinalidad del 42,86%.

### CI15, "juegos y distracciones"

La compatibilidad llega al 37,50% puesto que las piezas léxicas compartidas son seis (*fútbol*, *baloncesto*, *videojuegos*, *pelota*, *ordenador*, *televisión*), los varones tienen como más disponible un solo vocablo (*voleibol*) que no está en la lista de las mujeres, las cuales, a su vez, ofrecen nueve lexemas exclusivos (*leer*, *cartas*, *libro*, *bailar*, *cine*, *escuchar música*, *correr*, *natación*, *pasear*).

### CI16, "profesiones y oficios"

Las listas de mujeres y hombres comparten cuatro vocablos (*profesor*, *abogado*, *doctor*, *cocinero*). Ellas sobrepasan a los compañeros en siete unidades (*médico*, *camarero*, *fontanero*, *actor*, *secretario*, *estudiante*, *empleado*), ya que ellos solo ofrecen uno exclusivo (*bombero*), implicando una compatibilidad del 33,34%.

## Variable nivel de ELE

### CI01, "partes del cuerpo"

El índice de cardinalidad es bastante elevado, pues logra un 77,27%. Los vocablos compartidos son diecisiete y coinciden con las aportaciones de los informantes de nivel B1: *ojos*, *mano*, *cabeza*, *pie*, *pierna*, *nariz*, *pelo*, *boca*, *brazo*, *cara*, *dedo*, *espalda*, *cuello*, *rodilla*, *orejas*, *lengua*, *barriga*. Los de nivel B2 presentan cinco unidades adicionales con ID ≥ 0,1 (*uñas*, *dientes*, *hombro*, *labios*, *cejas*).

### CI02, "la ropa"

Otra vez, los elementos comunes se corresponden perfectamente con la lista de vocablos de los estudiantes de nivel umbral (*camiseta*, *zapatos*, *pantalones*, *vaqueros*, *jersey*, *chaqueta*, *camisa*, *falda*, *vestido*, *abrigo*) ya que los avanzados arrojan una cantidad mayor de tres (*bufanda*, *calcetines*, *blusa*). La cardinalidad sigue siendo elevada: 76,92%.

**CI03, "partes de la casa (sin muebles)"**

El índice disminuye al 55,56% ya que la suma disyuntiva entre los dos grupos es mayor con respecto a lo visto hasta ahora. Los informantes de nivel B1 tienen una lexía exclusiva (*garaje*) y los B2 seis (*cuarto*, *techo*, *garaje*, *puerta*, *escalera*, *ventana*). En total los vocablos compartidos son cinco: *cocina*, *cuarto de baño*, *habitación*, *salón*, *jardín*.

**CI04, "los muebles de la casa"**

Los vocablos totales de los B2 son once entre los cuales tres (*mesilla de noche*, *televisión*, *cajón*) son exclusivos de este grupo, por tanto, los B1 tienen una lista de ocho unidades con ID ≥ 0,1 (*cama*, *silla*, *mesa*, *sofá*, *armario*, *sillón*, *ducha*, *lámpara*). Esto lleva a una cardinalidad del 72,73%.

**CI05, "alimentos y bebidas"**

Se produce en este CI una compatibilidad del 60,61% debido a la alta coincidencia registrada en las posiciones iniciales, pese a las divergencias de las finales donde averiguamos que los B1 proporcionan tres voces que no están en el otro listado (*sangría*, *tapas*, *aceite*) y los B2 hacen lo propio con diez (*huevo*, *zanahoria*, *chocolate*, *patata*, *cebolla*, *sal*, *galleta*, *pollo*, *limón*, *ensalada*). Los vocablos compartidos son: *agua*, *leche*, *tomate*, *cerveza*, *naranja*, *paella*, *pan*, *manzana*, *pasta*, *vino*, *pescado*, *arroz*, *jamón*, *zumo*, *café*, *carne*, *azúcar*, *plátano*, *bocadillo*, *queso*.

**CI06, "objetos colocados en la mesa para la comida"**

La cardinalidad se revela muy elevada, 85,71%, porque la lista de vocablos con ID mayor o igual a 0,1 comprende seis unidades por cada grupo (*plato*, *vaso*, *cuchillo*, *cuchara*, *botella*, *tenedor*) y solo hay una divergente en el B2 (*cucharilla*). Todas las lexías coincidentes se colocan exactamente en los mismos rangos en los dos grupos.

**CI07, "la cocina y sus utensilios"**

Hay una coincidencia de tres vocablos (*horno*, *plato*, *vaso*) y cada variante presenta en su lista algunos ítems exclusivos con ID ≥ 0,1: los B1 proporcionan tres (*frigorífico*, *lavavajillas*, *nevera*) y los B2 cuatro (*cuchillo*, *cuchara*, *sartén*, *olla*). Con todo esto la cardinalidad es del 30%.

### CI08, "la escuela: muebles y materiales"

El índice llega a su máximo en este tema: no hay ninguna lexía divergente. Asimismo, se detecta cierta correspondencia en el orden de las aportaciones según rango, en particular las posiciones 4, 5, 10, 11 coinciden: *mesa*, *silla*, *libro*, *lápiz*, *cuaderno*, *pizarra*, *bolígrafo*, *ordenador*, *mochila*, *papel*, *borrador*.

### CI09, "iluminación y calefacción"

De nuevo, la cardinalidad del conjunto es del 100% ya que cada lista tiene tres voces que, además, se presentan según la misma ordenación: *lámpara*, *luz*, *fuego*.

### CI10, "la ciudad"

Los B1 tienen cinco unidades que no aparecen en la lista de los avanzados (*ayuntamiento*, *árbol*, *museo*, *cine*, *universidad*), que a su vez presentan cuatro exclusivas (*tren*, *bicicleta*, *restaurante*, *palacio*). Los vocablos correspondientes son trece (*calle*, *coche*, *tienda*, *casa*, *parque*, *autobús*, *edificio*, *escuela*, *plaza*, *iglesia*, *carretera*, *bar*, *estación*). La compatibilidad, notablemente reducida con respecto a los CI anteriores, es del 59,09%.

### CI11, "el campo"

Los alumnos B2 producen una lista de vocablos con ID ≥ 0,1 de trece elementos (los exclusivos son *tierra*, *planta*, *fruta*, *gallina*, *verdura*, *tractor*), mientras que los B1 alcanzan un total de diez (los exclusivos son *naturaleza*, *perro*, *casa*). Las unidades compartidas son siete (*árbol*, *animal*, *flor*, *vaca*, *caballo*, *hierba*, *campesino*). El índice de cardinalidad es del 43,75%.

### CI12, "medios de transporte"

La compatibilidad alcanzada aquí es del 100% porque la única diferencia entre los registros es la distinta colocación de los vocablos en las listas ya que ambos grupos arrojan el mismo número y las mismas lexías con ID ≥ 0,1: *coche*, *autobús*, *tren*, *bicicleta*, *avión*, *barco*, *motocicleta*, *metro*, *taxi*, *tranvía*.

### C113, "trabajos del campo y del jardín"

Aquí cuatro lexemas coinciden (*jardinero*, *campesino*, *plantar*, *cortar*): los B2 tienen en su lista dos vocablos que los B1 no presentan, por lo tanto, la compatibilidad es del 50%.

### C114, "los animales"

El índice llega a un 64,71%. Aparecen once vocablos coincidentes (*perro*, *gato*, *caballo*, *vaca*, *pez*, *pájaro*, *león*, *cerdo*, *conejo*, *tigre*, *gallina*) y seis que no están compartidos y se reparten entre las variantes: los estudiantes B1 ofrecen dos piezas exclusivas (*toro*, *burro*) y los B2 cuatro (*elefante*, *pollo*, *cebra*, *ave*).

### C115, "juegos y distracciones"

El grupo B1 presenta en total dieciséis vocablos más disponibles, frente a los doce de los compañeros. En efecto, tiene seis voces divergentes (*cine*, *jugar*, *pasear*, *natación*, *tocar*, *correr*) mientras que los B2 solo logran dos (*videojuegos*, *nadar*) y los ítems compartidos son diez (*fútbol*, *baloncesto*, *leer*, *cartas*, *libro*, *ordenador*, *televisión*, *bailar*, *pelota*, *escuchar música*), lo que lleva a calcular una cardinalidad del conjunto del 55,56%.

### C116, "profesiones y oficios"

El índice es del 46,67% debido a que los vocablos comunes son siete (*profesor*, *camarero*, *cocinero*, *médico*, *doctor*, *abogado*, *actor*) y los aprendientes de nivel umbral revelan tres lexías adicionales en su lista (*fontanero*, *estudiante*, *entrenador*), los de nivel avanzado hacen lo mismo con cinco (*secretario*, *empleado*, *bombero*, *policía*, *enfermero*).

## Variable conocimiento de otras LE

### C101, "partes del cuerpo"

La cardinalidad del conjunto es 68,18% ya que hay quince lexemas compartidos (*ojos*, *mano*, *cabeza*, *pie*, *pelo*, *nariz*, *boca*, *pierna*, *dientes*, *rodilla*, *espalda*, *brazo*, *barriga*, *uñas*, *lengua*) y cada variante aporta algunos elementos divergentes: el grupo =2 LE tiene tres (*cara*, *dedo*, *hombro*) y el otro cuatro (*boca*, *cara*, *cuello*, *orejas*).

**C102, "la ropa"**

La correspondencia es de nueve vocablos, que coinciden con la lista de los estudiantes que conocen dos LE (*camiseta*, *zapatos*, *chaqueta*, *pantalones*, *falda*, *jersey*, *camisa*, *vaqueros*, *calcetines*). El otro grupo ha escrito más unidades con ID ≥ 0,1, pues, tiene cuatro voces exclusivas (*bufanda*, *abrigo*, *vestido*, *traje*), con lo cual la compatibilidad es del 69,23%.

**C103, "partes de la casa (sin muebles)"**

Computamos un índice bastante elevado (77,78%) puesto que la discrepancia entre los dos conjuntos es solo de dos palabras (*escalera*, *techo*) que están en el listado del grupo =2 LE. En total los vocablos en común son siete (*cocina*, *cuarto de baño*, *habitación*, *salón*, *jardín*, *cuarto*, *garaje*).

**C104, "los muebles de la casa"**

Los vocablos que aparecen en ambas listas son seis (*cama*, *silla*, *mesa*, *sofá*, *armario*, *sillón*) y la cardinalidad es del 50% ya que detectamos el mismo número de lexías divergentes en cada una de las opciones. Los encuestados >2 LE proporcionan entre sus respuestas la unidad pluriverbal *mesilla de noche*, que no está entre las de los compañeros =2 LE, que, a su vez, aportan *televisión*, *frigorífico*, *espejo*, ausentes en la otra lista.

**C105, "alimentos y bebidas"**

La cardinalidad es del 60,71%. Los vocablos compartidos son diecisiete (*agua*, *cerveza*, *vino*, *naranja*, *zumo*, *café*, *manzana*, *tomate*, *jamón*, *paella*, *pan*, *leche*, *pescado*, *carne*, *bocadillo*, *huevo*, *chocolate*) y cada conjunto ofrece algunos peculiares: en la lista de los alumnos que saben menos LE aparecen tres (*sangría*, *zanahoria*, *galleta*) mientras que los que conocen más ofrecen ocho (*arroz*, *pasta*, *queso*, *azúcar*, *patata*, *plátano*, *sal*, *tapas*).

**C106, "objetos colocados en la mesa para la comida"**

No hay ningún vocablo divergente, todos están en las listas de los dos grupos: *plato*, *vaso*, *cuchillo*, *botella*, *cuchara*, *tenedor*. La única diferencia se detecta en los rangos 4 y 5 que intercambian las lexías *botella* y *cuchara*. Por eso, la cardinalidad logra el 100%.

**c107, "la cocina y sus utensilios"**

En este campo contabilizamos la compatibilidad más baja en absoluto, se trata del 37,50%. Los vocablos comunes son tres (*horno*, *cuchillo*, *sartén*), cada grupo tiene por lo menos uno exclusivo: los informantes =2 LE presentan *nevera*; los >2 LE hacen lo propio con *plato*, *cuchara*, *vaso*, *frigorífico*.

**c108, "la escuela: muebles y materiales"**

Las unidades compartidas son nueve (*libro*, *bolígrafo*, *lápiz*, *cuaderno*, *pizarra*, *mesa*, *silla*, *papel*, *ordenador*) y hacen parte de la lista de los estudiantes que conocen dos LE. Del otro lado, la lista de los compañeros es más larga ya que contiene once elementos, de los cuales dos son divergentes (*mochila*, *borrador*). El índice aquí es del 81,82%.

**c109, "iluminación y calefacción"**

En este centro de interés la cardinalidad es del 100%, no detectamos ninguna diferencia, ni del número de vocablos con ID ≥ 0,1 ni de su colocación en las listas (*lámpara*, *luz*, *fuego*).

**c110, "la ciudad"**

Los vocablos compartidos son doce (*calle*, *tienda*, *edificio*, *casa*, *coche*, *iglesia*, *plaza*, *parque*, *bar*, *autobús*, *carretera*, *escuela*) y cada variante tiene algunos exclusivos: dos en el grupo =2 LE (*árbol*, *museo*) y cinco en el >2 LE (*estación*, *tren*, *ayuntamiento*, *restaurante*, *universidad*). Según eso, la cardinalidad del conjunto es 63,16%.

**c111, "el campo"**

El índice es del 60% debido a los seis vocablos en común (*árbol*, *animal*, *campesino*, *caballo*, *vaca*, *flor*) y a los peculiares de cada grupo: uno de los informantes que conocen dos LE (*perro*) y tres de los que saben más idiomas (*hierba*, *tierra*, *planta*).

**c112, "medios de transporte"**

Hay un solo lexema divergente (*barco*) que se encuentra en el listado de los >2 LE, los elementos comunes son nueve (*coche*, *avión*, *autobús*,

*tren*, *bicicleta*, *metro*, *tranvía*, *taxi*, *motocicleta*). Esto hace que la compatibilidad llegue a un 90%.

**C113, "trabajos del campo y del jardín"**

La cardinalidad del conjunto en este caso es del 50%: la lista de vocablos más disponibles del conjunto =2 LE engloba las lexías compartidas (*jardinero*, *campesino*) y los >2 LE ofrecen, por su parte, dos vocablos adicionales (*plantar*, *cortar*).

**C114, "los animales"**

En este centro de interés las unidades comunes que alcanzan el ID ≥ 0,1 son diez (*perro*, *gato*, *caballo*, *cerdo*, *pez*, *pájaro*, *vaca*, *león*, *tigre*, *gallina*) y la cardinalidad es del 66,67% puesto que cada muestreo tiene vocablos exclusivos. El primero actualiza *pollo* mientras que el segundo proporciona *conejo*, *toro*, *elefante*, *burro*.

**C115, "juegos y distracciones"**

Hay diez vocablos comunes: *fútbol*, *baloncesto*, *leer*, *libro*, *televisión*, *bailar*, *ordenador*, *pelota, videojuegos, cine*. Al analizar por separado las listas, observamos que el grupo =2 LE ha activado tres exclusivos (*Play Station*, *nadar*, *discoteca)* y el otro cinco (*cartas*, *escuchar música*, *natación*, *correr*), por tanto, el porcentaje de cardinalidad es 58,82%.

**C116, "profesiones y oficios"**

El índice de compatibilidad es del 53,85%, los vocablos en común son siete (*profesor, médico, abogado, fontanero, cocinero, doctor, camarero*) y ambas variantes presentan por lo menos uno divergente. Los aprendientes =2 LE tienen una unidad más (*jardinero*) y los >2 LE presentan cinco adicionales (*actor*, *empleado*, *estudiante*, *secretario*, *cantante*).

### 3.2.3 Análisis longitudinal

En esta sección damos cuenta de la compatibilidad de las respuestas según el momento de realización de la prueba, valorando el grado de coincidencia entre el comienzo y el final del año académico para compro-

bar si tras la asistencia a un curso de ELE los vocablos con ID ≥ 0,1 activados son los mismos o cambian en cantidad y calidad.

| CI | Cardinalidad del conjunto |
|---|---|
| CI01.CUE | 95,45% |
| CI02.ROP | 61,90% |
| CI03.CAS | 64,29% |
| CI04.MUE | 53,33% |
| CI05.ALI | 52,63% |
| CI06.MES | 70% |
| CI07.COC | 60% |
| CI08.ESC | 61,11% |
| CI09.ILU | 30% |
| CI10.CIU | 50% |
| CI11.CAM | 44,44% |
| CI12.TRA | 83,33% |
| CI13.TRC | 57,14% |
| CI14.ANI | 56,52% |
| CI15.JUE | 45,83% |
| CI16.PRO | 100% |

Tabla 40. Cardinalidad del conjunto según la fecha de la prueba.

### CI01, "las partes del cuerpo"

En la primera prueba los encuestados han activado una palabra más, *cejas*, y la correspondencia es de veintiuno lexemas: *mano*, *ojos*, *pierna*, *cabeza*, *pie*, *boca*, *dedo*, *brazo*, *nariz*, *pelo*, *cara*, *orejas*, *espalda*, *cuello*, *rodilla*, *uñas*, *lengua*, *dientes*, *hombro*, *labios*, *barriga*. La cardinalidad llega al 95,45%.

**C102, "la ropa"**

En este campo el corpus B2_b tiene ocho lexemas exclusivos (*sudadera*, *sombrero*, *bolso*, *cinturón*, *traje*, *minifalda*, *zapatillas*, *tacones*), ya que los demás vocablos que componen la lista de ítems con ID ≥ 0,1 están compartidos con el B2_a (*camiseta*, *pantalones*, *zapatos*, *camisa*, *falda*, *bufanda*, *chaqueta*, *jersey*, *calcetines*, *vaqueros*, *abrigo*, *vestido*, *blusa*). La compatibilidad del conjunto es del 61,90%.

**C103, "partes de la casa (sin muebles)"**

Cada conjunto presenta aquí algunas palabras divergentes: en la primera prueba constatamos dos (*puerta*, *ventana*) y en la segunda tres (*pasillo*, *comedor*, *piso*), mientras que las comunes son nueve (*cocina*, *cuarto de baño*, *cuarto*, *salón*, *jardín*, *habitación*, *techo*, *garaje*, *escalera*). Todo esto permite contabilizar un índice del 64,29%.

**C104, "los muebles de la casa"**

La cardinalidad del 53,33% se computa a partir de las ocho voces comunes (*silla*, *mesa*, *cama*, *armario*, *sofá*, *sillón*, *mesilla de noche*, *lámpara*) y de la suma disyuntiva que ve tres vocablos peculiares en B2_a (*ducha, televisión, cajón*) y cuatro en B2_b (*horno*, *estantería*, *lavadora*, *nevera*).

**C105, "alimentos y bebidas"**

He aquí veinte elementos en común (*agua*, *leche*, *tomate*, *cerveza*, *naranja*, *paella*, *pan*, *manzana*, *pasta*, *vino*, *pescado*, *arroz*, *jamón*, *zumo*, *zanahoria*, *plátano*, *pollo*, *limón*, *bocadillo*, *queso*). El corpus B2_a tiene diez divergentes (*huevo*, *café*, *carne*, *chocolate*, *patata*, *azúcar*, *cebolla*, *sal*, *galleta*, *ensalada*) y el B2_b presenta ocho exclusivos (*sangría*, *fresa*, *tequila*, *gazpacho*, *patata*, *hamburguesa*, *tortilla*). Estos datos llevan a computar una compatibilidad del 52,63%.

**C106, "objetos colocados en la mesa para la comida"**

Hay una alta correlación de los vocablos ubicados en las primeras posiciones de las listas, la cardinalidad es del 70%. Los ítems compartidos son siete y coinciden con el listado de la primera prueba (*plato*, *vaso*, *cu-*

*chillo*, *cuchara*, *botella*, *tenedor*, *cucharilla*). En la segunda, los encuestados presentan tres unidades adicionales con ID ≥ 0,1 (*mantel*, *servilleta*, *sal*).

**C107, "la cocina y sus utensilios"**

Es posible que en este CI se hayan reciclado las respuestas del anterior según el fenómeno del aprovechamiento léxico: las lexías compartidas son seis (*horno*, *cuchillo*, *plato*, *cuchara*, *sartén*, *olla*) y repiten las que acabamos de comentar. Asimismo, se ofrecen una palabra exclusiva (*vaso*) en la primera prueba y tres en la segunda (*nevera*, *tenedor*, *horno microondas*), lo que supone que la cardinalidad sea del 60%.

**C108, "la escuela: muebles y materiales"**

El índice de compatibilidad calculado es del 61,11%: son once los vocablos comunes en las dos suministraciones de la prueba (*bolígrafo*, *libro*, *silla*, *lápiz*, *cuaderno*, *mesa*, *pizarra*, *mochila*, *ordenador*, *papel*, *borrador)*, que aparecen todas en la lista de la primera. Contamos con siete exclusivos en la segunda (*estuche*, *profesor*, *estudiar*, *estudiante*, *escribir*, *leer*, *sacapuntas*).

**C109, "iluminación y calefacción"**

Este campo semántico revela el menor porcentaje de compatibilidad con respecto a los demás, pues no supera el 30%. Las unidades presentes en ambos listados son tres (*lámpara*, *luz*, *fuego*) y coinciden con los vocablos con ID ≥ 0,1 aportados por los informantes durante la primera prueba. Las exclusivas del corpus B2_b son siete (*bombilla*, *calor*, *sol*, *radiador*, *iluminar*, *calefacción*, *encender*). Esto corrobora el incremento del bagaje léxico a final del año académico, ya que los vocablos más disponibles pasan de tres a diez.

**C110, "la ciudad"**

Disponemos de trece ítems compartidos (*calle*, *coche*, *tienda*, *casa*, *parque*, *autobús*, *escuela*, *plaza*, *iglesia*, *carretera*, *bar*, *estación*, *tren*, *bicicleta*, *restaurante*, *palacio*) que se suman a los peculiares de cada muestreo. La cardinalidad del conjunto es del 50%.

### C11, "el campo"

Con una intersección de ocho vocablos (*árbol*, *animal*, *tierra*, *hierba*, *flor*, *campesino*, *vaca*, *caballo*) y una suma disyuntiva de diez, que se divide respectivamente en cinco elementos exclusivos en cada corpus (B2_a: *planta*, *fruta*, *gallina*, *verdura*, *tractor*; B2_b: *campo*, *cortijo*, *tractor*, *cultivar*, *río*) calculamos un índice de compatibilidad del 44,44%.

### C12, "medios de transporte"

En un total de veintitrés vocablos que tienen un ID mayor o igual a 0,1, diez están presentes en ambos listados y corresponden con la lista procedente de la prueba suministrada a comienzo del año académico (*coche*, *autobús*, *tren*, *bicicleta*, *avión*, *taxi*, *metro*, *motocicleta*, *barco*, *tranvía*). Las unidades exclusivas extraídas de la otra prueba son dos (*a pie*, *carro*). La compatibilidad es del 83,33%.

### C13, "trabajos del campo y del jardín"

En este centro de interés la cardinalidad del conjunto es del 57,14% puesto que los dos sub-corpus comparten cuatro vocablos (*jardinero*, *plantar*, *campesino*, *cortar*) y al realizar la segunda encuesta los informantes han actualizado tres más con ID ≥ 0,1 (*cultivar*, *regar*, *flor*).

### C14, "los animales"

Los vocablos comunes son trece (*perro*, *gato*, *caballo*, *vaca*, *pájaro*, *cerdo*, *león*, *pez*, *gallina*, *elefante*, *tigre*, *pollo*, *conejo*) y los diferentes diez: dos son exclusivos del B2_a (*cebra*, *ave*) y ocho del B2_b (*ratón*, *jirafa*, *burro*, *tortuga*, *hámster*, *mono*, *mariposa*, *serpiente*). Conforme a esto, la compatibilidad del conjunto es del 56,52%.

### C15, "juegos y distracciones"

En la segunda aplicación de la prueba los vocablos que alcanzan el ID requerido son once (*jugar*, *móvil*, *correr*, *cine*, *tenis*, *deporte*, *pasear*, *balón*, *redes sociales*, *voleibol*, *natación*), que se añaden a los doce activados en la primera (*fútbol*, *baloncesto*, *leer*, *cartas*, *libro*, *videojuegos*, *ordenador*, *televisión*, *bailar*, *pelota*, *nadar*, *escuchar música*). La correlación de estos datos lleva a una cardinalidad del conjunto del 52,17%.

### C116, "profesiones y oficios"

Computamos un índice de compatibilidad del 52,38% debido a que los lexemas comunes son once (*profesor*, *médico*, *abogado*, *cocinero*, *doctor*, *actor*, *secretario*, *bombero*, *policía*, *camarero*, *enfermero*), hay uno exclusivo en el B2_a (*empleado*) y nueve en el B2_b (*peluquero*, *estudiante*, *jardinero*, *juez*, *escritor*, *panadero*, *cantante*, *trabajar*, *bailarín*). De nuevo, esto parece confirmar que los encuestados al terminar el año académico activan un caudal mayor de vocablos más disponibles, pasando de doce a veinte.

## 3.2.4 Análisis comparativo

En esta última sección cotejamos los 1.490 vocablos recopilados en el corpus con las listas extraídas del CREA y del CAES con el objetivo de examinar cuántas y cuáles son las unidades compartidas con el léxico disponible de nuestros informantes. Además de comparar las listas enteras de los dos repertorios, tomamos como medidas de corte las primeras 500 y 1.000 entradas para averiguar la cantidad de lexemas compartida en intervalos menores.[38]

La primera comparación se desarrolla a partir de la lista de frecuencia proporcionada por el CREA, que encierra las 5.000 palabras más frecuentes de la lengua española.

| Vocablos CREA | Vocablos compartidos | |
|---|---|---|
| | Total | Porcentaje |
| 500 | 32 | 6,40% |
| 1.000 | 76 | 7,60% |
| 5.000 | 278 | 5,56% |

Tabla 41. Vocablos compartidos con el CREA.

Observamos que, pese a la diferente conformación interna de nuestro corpus con respecto al CREA, el vocabulario disponible de los informantes está incluido en el léxico más frecuente del español, aunque de manera reducida. El 20% de estas unidades compartidas coincide con los vocablos

[38] Claro está que no se trata de un análisis exhaustivo ya que nuestro corpus fue editado y lematizado, mientras que el CREA y el CAES contienen palabras funcionales, nombres propios y formas conjugadas, flexivas, derivadas, precisamente porque proceden de material auténtico.

más disponibles de nuestro corpus general, esto es, los que tienen un ID ≥ 0,1. En orden alfabético se trata de: *actor*, *animal*, *árbol*, *arroz*, *avión*, *bailar*, *bar*, *barco*, *botella*, *brazo*, *caballo*, *café*, *calle*, *cara*, *carne*, *carretera*, *cartas*, *cerveza*, *cine*, *cocina*, *cortar*, *cuarto*, *cuello*, *dientes*, *doctor*, *edificio*, *estación*, *estudiante*, *flor*, *fuego*, *fútbol*, *gato*, *habitación*, *horno*, *iglesia*, *jardín*, *leche*, *lengua*, *león*, *libro*, *luz*, *naranja*, *nariz*, *ojos*, *orejas*, *papel*, *pasta*, *pelo*, *pelota*, *perro*, *pie*, *pierna*, *planta*, *plato*, *plaza*, *profesor*, *restaurante*, *sal*, *salón*, *silla*, *sillón*, *taxi*, *televisión*, *tienda*, *tierra*, *toro*, *tren*, *vaso*, *vestido*, *vino*.

El análisis gramatical de este conjunto corrobora la tendencia surgida a lo largo de los análisis transversal y longitudinal, revelando una masiva presencia de sustantivos y una escasa aparición de verbos y adjetivos.

En segunda instancia, cotejamos la lista de frecuencia procedente del CAES de la cual tenemos en cuenta exclusivamente los datos del nivel intermedio[39] y quitamos las direcciones de correo electrónico registradas, ya que no hacen parte del léxico español. Por eso, consideramos una lista compuesta por un total de 5.364 ítems.

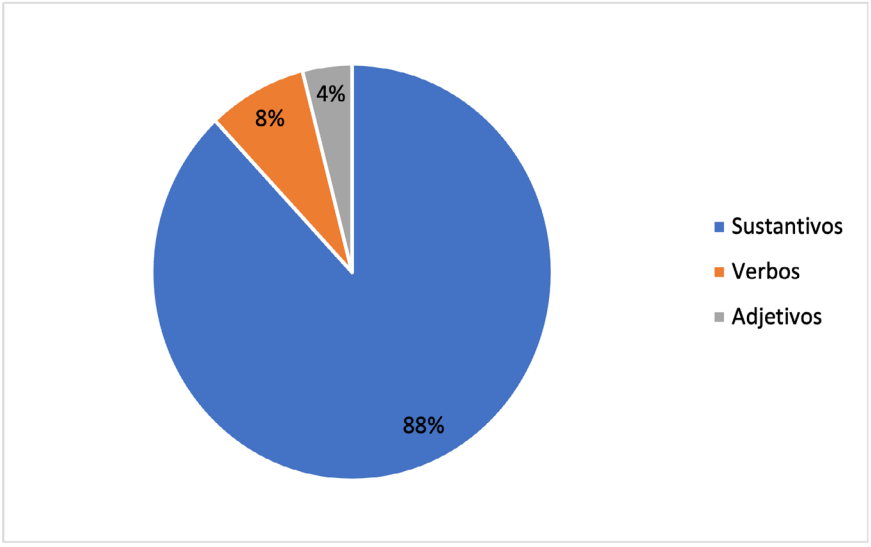


Gráfico 31. Categorías gramaticales de los vocablos compartidos con el CREA.

[39] Como este corpus recoge todos los niveles del MCER, sumamos los datos relativos a los niveles B1 y B2 para realizar el cotejo conforme al dominio del español que tienen nuestros informantes.

| Lista frecuencia CAES | Vocablos compartidos | |
|---|---|---|
| | Total | Porcentaje |
| 500 | 74 | 14,80% |
| 1.000 | 153 | 15,30% |
| 5.364 | 559 | 10,42% |

Tabla 42. Vocablos compartidos con el CAES.

De una primera ojeada sobresale una mayor coincidencia de nuestro repertorio con estos datos con respecto a los del CREA. Los porcentajes de compatibilidad casi doblan los calculados antes. Llegamos a un índice de 15,30% como máximo al considerar la lista de las 1.000 palabras, intervalo en el que se detectan más correspondencias. Las cifras bajan ligeramente en los demás, pero siguen siendo mayores si las contrastamos con el CREA.

En lo que se refiere a la subdivisión según categoría gramatical, otra vez, se nota una elevada aparición de sustantivos (88,73%) seguidos por una limitada cantidad de verbos (8,23%), adjetivos (2,86%) y un solo adverbio (0,18%).

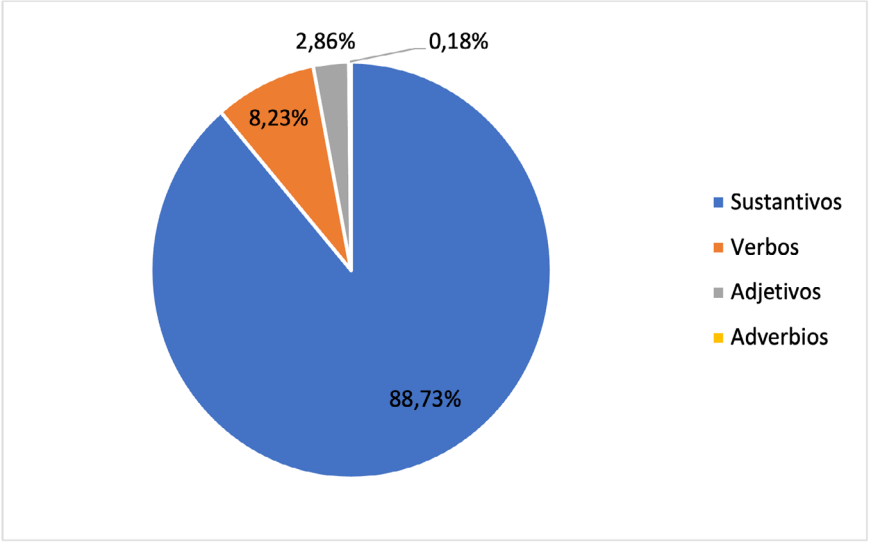


Gráfico 32. Categorías gramaticales de los vocablos compartidos con el CAES.

Sin duda, el caudal de vocablos compartidos depende del tipo de prueba a la que se someten los informantes. Los tests de disponibilidad léxica intentan abarcar todos los intereses humanos y ser lo más variados posi-

ble, mientras que las tareas utilizadas para la recopilación del CAES están más circunscritas (por ejemplo, para el nivel B1: escribir una carta a un amigo, escribir una carta con una queja por pérdida de equipaje a una compañía aérea, escribir una historia; para el nivel B2: redactar una solicitud de admisión en un programa universitario motivando la petición y preguntando sobre las características del programa, escribir un ensayo sobre la importancia de las nuevas tecnologías en la sociedad actual).

# Capítulo 4.
# Análisis de la riqueza léxica

## 4.1 Estudio cuantitativo

En esta primera parte del capítulo profundizamos el estudio de la competencia léxica de nuestros informantes midiendo la riqueza de vocabulario producida en los textos escritos durante las pruebas suministradas.

En primera instancia, el análisis transversal discute los datos generales y por variable de variación léxica y densidad. Seguidamente, extendemos el análisis con el estudio de los datos estadísticos descriptivos que informan sobre la simetría o la dispersión del comportamiento intragrupal e intergrupal. Igualmente, el análisis longitudinal compara los resultados de la primera prueba con los de la segunda para comprobar si ha habido un incremento de los índices y, en consecuencia, del conocimiento léxico. El análisis comparativo cierra esta sección con el cotejo entre nuestros resultados con los de otras investigaciones para averiguar si aprendientes de ELE de igual dominio lingüístico y de distinta procedencia (LM, entorno cultural y educativo) presentan la misma riqueza léxica.

### 4.1.1 Análisis transversal: resultados generales

Tras los procesos de edición y digitalización del material recogido en la primera suministración de la prueba, los datos generales extraídos de los cómputos expresan la variedad y la densidad léxica a partir de un total

de 10.000 *tokens* y de 1.072 *types* repartidos en el conjunto textual (100 textos). En detalle, los índices se distribuyen como sigue en cada relato.

| Texto | TTR | Variación léxica | Índice de hápax | Densidad léxica | IAT |
|---|---|---|---|---|---|
| 1 | 0,64 | 64% | 1,23 | 48% | 2,08 |
| 2 | 0,53 | 53% | 1,51 | 55% | 1,82 |
| 3 | 0,66 | 66% | 1,43 | 56% | 1,79 |
| 4 | 0,62 | 62% | 1,24 | 50% | 2 |
| 5 | 0,61 | 61% | 1,27 | 59% | 1,69 |
| 6 | 0,57 | 57% | 1,50 | 55% | 1,82 |
| 7 | 0,60 | 60% | 1,43 | 59% | 1,69 |
| 8 | 0,60 | 60% | 1,43 | 55% | 1,82 |
| 9 | 0,60 | 60% | 1,43 | 55% | 1,82 |
| 10 | 0,64 | 64% | 1,36 | 57% | 1,75 |
| 11 | 0,62 | 62% | 1,32 | 59% | 1,69 |
| 12 | 0,65 | 65% | 1,48 | 60% | 1,67 |
| 13 | 0,63 | 63% | 1,40 | 52% | 1,92 |
| 14 | 0,59 | 59% | 1,51 | 55% | 1,82 |
| 15 | 0,68 | 68% | 1,39 | 63% | 1,59 |
| 16 | 0,53 | 53% | 1,77 | 52% | 1,92 |
| 17 | 0,65 | 65% | 1,35 | 53% | 1,89 |
| 18 | 0,61 | 61% | 1,45 | 43% | 2,33 |
| 19 | 0,53 | 53% | 1,47 | 51% | 1,96 |
| 20 | 0,66 | 66% | 1,14 | 50% | 2 |
| 21 | 0,59 | 59% | 1,51 | 53% | 1,89 |
| 22 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 23 | 0,59 | 59% | 1,51 | 52% | 1,92 |

| | | | | | |
|---|---|---|---|---|---|
| 24 | 0,66 | 66% | 1,27 | 55% | 1,82 |
| 25 | 0,71 | 71% | 1,18 | 56% | 1,79 |
| 26 | 0,67 | 67% | 1,31 | 55% | 1,82 |
| 27 | 0,59 | 59% | 1,40 | 55% | 1,82 |
| 28 | 0,56 | 56% | 1,65 | 60% | 1,67 |
| 29 | 0,58 | 58% | 1,41 | 55% | 1,82 |
| 30 | 0,64 | 64% | 1,56 | 60% | 1,67 |
| 31 | 0,62 | 62% | 1,07 | 53% | 1,89 |
| 32 | 0,75 | 75% | 1,29 | 64% | 1,56 |
| 33 | 0,64 | 64% | 1,25 | 53% | 1,89 |
| 34 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 35 | 0,64 | 64% | 1,25 | 56% | 1,79 |
| 36 | 0,69 | 69% | 1,41 | 66% | 1,52 |
| 37 | 0,59 | 59% | 1,55 | 51% | 1,96 |
| 38 | 0,69 | 69% | 1,33 | 54% | 1,85 |
| 39 | 0,63 | 63% | 1,34 | 50% | 2 |
| 40 | 0,54 | 54% | 1,69 | 49% | 2,04 |
| 41 | 0,63 | 63% | 1,54 | 61% | 1,64 |
| 42 | 0,70 | 70% | 1,27 | 58% | 1,72 |
| 43 | 0,61 | 61% | 1,42 | 60% | 1,67 |
| 44 | 0,62 | 62% | 1,38 | 59% | 1,69 |
| 45 | 0,57 | 57% | 1,43 | 54% | 1,85 |
| 46 | 0,66 | 66% | 1,35 | 60% | 1,67 |
| 47 | 0,66 | 66% | 1,29 | 60% | 1,67 |
| 48 | 0,64 | 64% | 1,36 | 57% | 1,75 |
| 49 | 0,65 | 65% | 1,25 | 55% | 1,82 |
| 50 | 0,61 | 61% | 1,45 | 54% | 1,85 |

| | | | | | |
|---|---|---|---|---|---|
| 51 | 0,67 | 67% | 1,37 | 61% | 1,64 |
| 52 | 0,54 | 54% | 1,64 | 50% | 2 |
| 53 | 0,62 | 62% | 1,38 | 50% | 2 |
| 54 | 0,60 | 60% | 1,50 | 54% | 1,85 |
| 55 | 0,59 | 59% | 1,31 | 53% | 1,89 |
| 56 | 0,62 | 62% | 1,41 | 53% | 1,89 |
| 57 | 0,55 | 55% | 1,45 | 48% | 2,08 |
| 58 | 0,65 | 65% | 1,38 | 63% | 1,59 |
| 59 | 0,71 | 71% | 1,25 | 57% | 1,75 |
| 60 | 0,66 | 66% | 1,38 | 56% | 1,79 |
| 61 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 62 | 0,67 | 67% | 1,31 | 60% | 1,67 |
| 63 | 0,62 | 62% | 1,32 | 53% | 1,89 |
| 64 | 0,64 | 64% | 1,33 | 55% | 1,82 |
| 65 | 0,61 | 61% | 1,45 | 58% | 1,72 |
| 66 | 0,63 | 63% | 1,50 | 55% | 1,82 |
| 67 | 0,54 | 54% | 1,59 | 50% | 2 |
| 68 | 0,66 | 66% | 1,29 | 57% | 1,75 |
| 69 | 0,68 | 68% | 1,28 | 59% | 1,69 |
| 70 | 0,53 | 53% | 1,56 | 49% | 2,04 |
| 71 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 72 | 0,64 | 64% | 1,28 | 52% | 1,92 |
| 73 | 0,67 | 67% | 1,34 | 52% | 1,92 |
| 74 | 0,57 | 57% | 1,46 | 50% | 2 |
| 75 | 0,73 | 73% | 1,30 | 56% | 1,79 |
| 76 | 0,66 | 66% | 1,47 | 55% | 1,82 |
| 77 | 0,70 | 70% | 1,37 | 61% | 1,64 |

| | | | | | |
|---|---|---|---|---|---|
| 78 | 0,58 | 58% | 1,41 | 46% | 2,17 |
| 79 | 0,56 | 56% | 1,65 | 55% | 1,82 |
| 80 | 0,59 | 59% | 1,40 | 58% | 1,72 |
| 81 | 0,72 | 72% | 1,26 | 59% | 1,69 |
| 82 | 0,67 | 67% | 1,34 | 59% | 1,69 |
| 83 | 0,64 | 64% | 1,33 | 58% | 1,72 |
| 84 | 0,67 | 67% | 1,29 | 52% | 1,92 |
| 85 | 0,61 | 61% | 1,45 | 54% | 1,85 |
| 86 | 0,65 | 65% | 1,41 | 57% | 1,75 |
| 87 | 0,63 | 63% | 1,31 | 49% | 2,04 |
| 88 | 0,66 | 66% | 1,27 | 53% | 1,89 |
| 89 | 0,65 | 65% | 1,30 | 48% | 2,08 |
| 90 | 0,61 | 61% | 1,42 | 53% | 1,89 |
| 91 | 0,63 | 63% | 1,29 | 52% | 1,92 |
| 92 | 0,58 | 58% | 1,53 | 54% | 1,85 |
| 93 | 0,60 | 60% | 1,50 | 57% | 1,75 |
| 94 | 0,69 | 69% | 1,35 | 57% | 1,75 |
| 95 | 0,67 | 67% | 1,34 | 58% | 1,72 |
| 96 | 0,66 | 66% | 1,29 | 55% | 1,82 |
| 97 | 0,67 | 67% | 1,29 | 54% | 1,85 |
| 98 | 0,59 | 59% | 1,48 | 56% | 1,79 |
| 99 | 0,67 | 67% | 1,37 | 57% | 1,75 |
| 100 | 0,61 | 61% | 1,39 | 50% | 2 |
| Promedio | 0,63 | 63% | 1,39 | 55% | 1,84 |

Tabla 43. Índices de RL.

### Diversidad léxica

La primera etapa consiste en la aplicación de las fórmulas que diferencian los *tokens* de los *types* y aportan información sobre la variedad del corpus y de cada texto: *Type/Token Ratio* (TTR), variación léxica, índice de hápax.

Recordamos que el TTR puede variar, idealmente, de 0 a 1: el valor 1 es improbable de alcanzar en un escrito que no sea una oración ya que implica que no se repita ninguna palabra; análogamente, es imposible bajar hasta 0 porque siempre tiene que aparecer un *type* como mínimo de manera que un texto exista. Por ende, cuanto más el resultado se acerca a 1, tanto más el índice revela riqueza léxica. En nuestro caso, el valor mínimo se calcula en cuatro textos (2, 16, 19, 70) que no sobrepasan el 0,53; el máximo se detecta en el texto 32 que logra un 0,75; la media es de 0,63 y se alcanza en seis textos (13, 39, 41, 66, 87, 91). De ahí que sea posible dividir el corpus en tres franjas: el 46% se coloca bajo la media mientras que el 48% la supera, la oscilación va de 0,64 a 0,75.

La *Type/Token Ratio* es buena: sobre un total de 100 ítems que componen un texto, más de cincuenta unidades son diferentes.

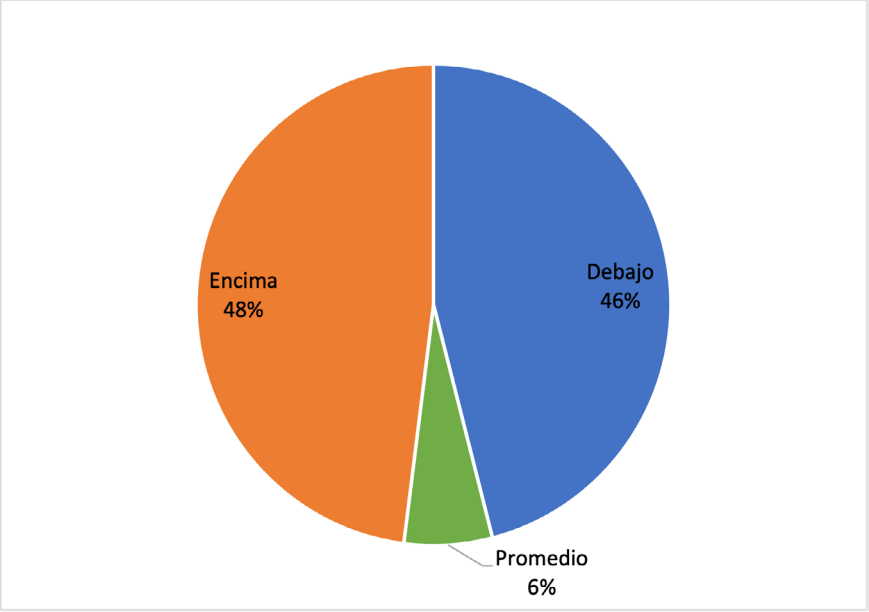


Gráfico 33. Distribución de los textos según el promedio de TTR.

Al traducir en porcentaje estos datos obtenemos el índice de la variación léxica. Los valores mínimo y máximo y el promedio coinciden con los vistos arriba: respectivamente 53%, 75% y 63%, por lo que el gráfico siguiente representa la misma distribución en tres ramos.

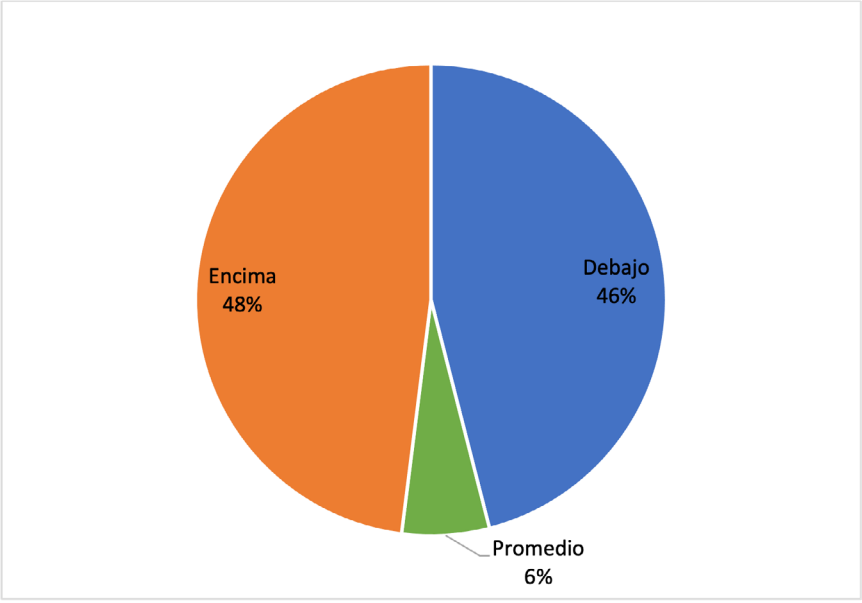


Gráfico 34. Distribución de los textos según el promedio de la variación léxica.

En suma, el corpus se divide en tres secciones: una presenta una variedad de vocabulario media que va del 53% al 59% (veinticuatro textos); otra encierra la gran mayoría de los textos que alcanzan una diversidad comprendida entre el 60% y el 69% (sesenta y nueve textos); la última se compone de los que consiguen hasta un 75% de variación (siete textos).

El índice de hápax concurre a establecer la diversidad léxica. En este caso, las palabras hápax constituyen un dato relevante que supone un buen nivel de riqueza léxica puesto que en promedio se contabilizan 46 hápax sobre una media de 63 vocablos (en el total de 100 palabras por texto). Destaca una gran oscilación de valores, hasta el doble, que supone una capacidad potencial de variación intragrupal muy diferente: el número mínimo de hápax es 30 (texto 25) y el máximo es 60 (texto 16). Destacamos, de nuevo, que cuanto más pequeño es el valor obtenido en un texto, más rico es su vocabulario ya que la riqueza léxica es inversamente proporcional al índice: el texto 31 logra el resultado menor (1,07),

donde se cuentan 58 hápax en un total de 62 *types*; mientras que el texto 16 alcanza el mayor (1,77), en el que los hápax son 30 y los vocablos 53. El 4% de los textos se queda en la media de 1,39 (textos 15, 34, 61, 100), el 44% la supera y el 52% queda por debajo, lo que implica una mayoría de textos más ricos.

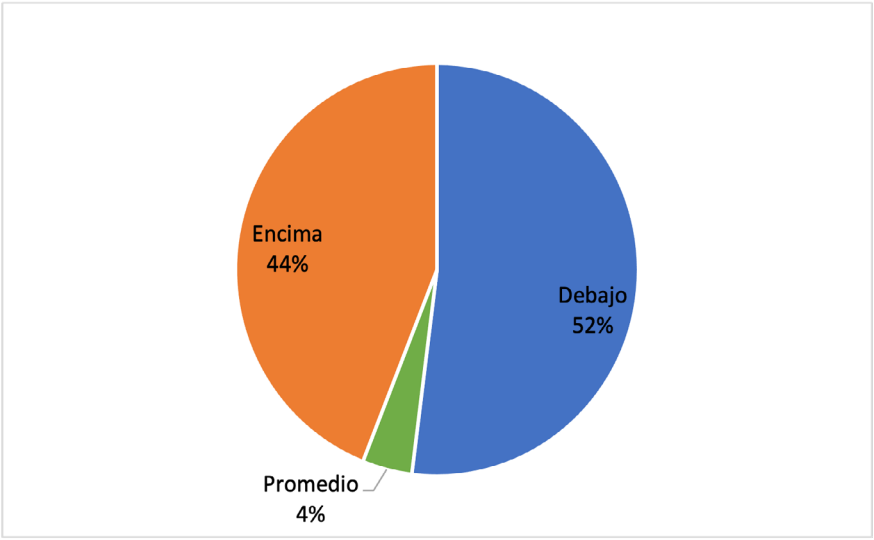


Gráfico 35. Distribución de los textos según el promedio del índice de hápax.

Los datos revelan valores del índice medio-bajos y un grado alto de riqueza ya que la mayoría de los textos se aproximan más a 1 que a 2.[40] Aún más, gran parte de los textos (81%) está bajo el 1,50 que es equidistante de los puntos de referencia.

### Densidad léxica

La segunda etapa se centra en el uso de las palabras nocionales a partir de la relación entre *tokens* y *tokens* léxicos expresada por los índices de densidad léxica e intervalo de aparición de palabras nocionales (IAT).

La densidad revela que más de la mitad de las palabras de un texto es nocional en la mayoría de las aportaciones. El corpus se divide en tres

[40] Se toma como medida de corte porque destaca cuanto el valor se aleja del ideal, si bien no se haya establecido un límite máximo universal.

grupos conforme a los intervalos porcentuales calculados: diez textos manifiestan datos más pobres (de 43% a 49%); setenta y seis textos tienen una densidad media (de 50% a 59%); catorce textos consiguen un rendimiento mayor (de 60% a 66%).

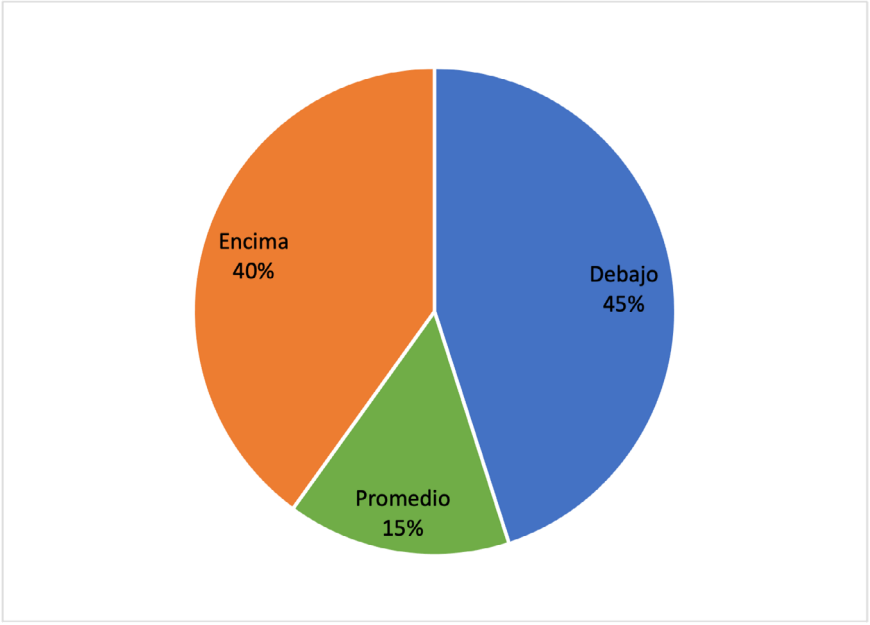


Gráfico 36. Distribución de los textos según el promedio de la densidad léxica.

Este resultado es notable si consideramos que solo el 10% de los relatos no llega al 50% de densidad, el 8% lo iguala y el 82% lo sobrepasa. El promedio (55%) se alcanza en los textos 2, 6, 8, 9, 14, 24, 26, 27, 29, 49, 64, 66, 76, 79, 96; el índice mínimo (43%) en el 18 y el máximo (66%) en el 36.

El IAT revela un empleo reducido de las palabras nocionales a favor de las funcionales: la media es de 1,84. Esto indica la presencia de un ítem léxico casi cada dos gramaticales y no representa un alto grado de riqueza ya que es posible alcanzar un resultado mejor cuando el intervalo se acerca a 1: «un intervalo de voces de contenido léxico de 2,1 detecta menor riqueza léxica que si se consigue un intervalo entre lexías de 1,2» (Reyes Díaz 2010: 148). El texto 36 obtiene el índice más bajo (1,52) y, por tanto, el uso de palabras semánticas es mayor con respecto a los otros. Por el contrario, el que más se distancia del valor ideal es el relato 18, donde el IAT es 2,33.

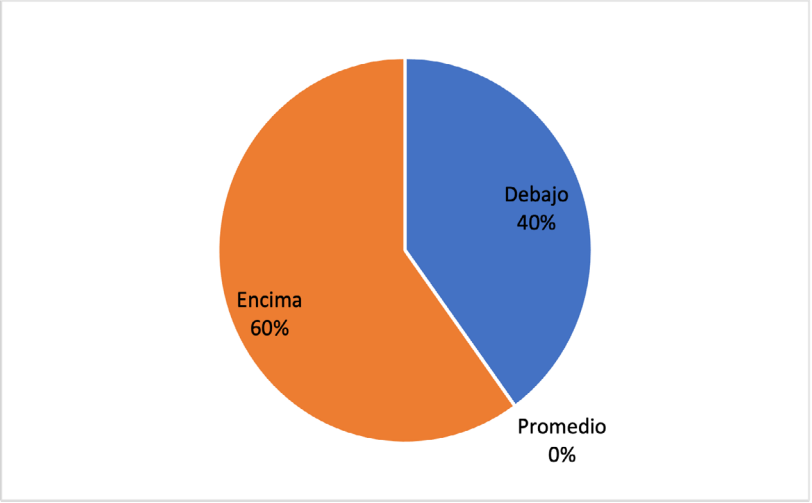


Gráfico 37. Distribución de los textos según el promedio de IAT.

La representación gráfica destaca que el 40% de los encuestados ha escrito textos que se quedan debajo de la media y proporcionan datos ligeramente mejores, pero todavía escasos puesto que las cifras se avecinan más a 2 que a 1. El 60% de los textos que está por encima consigue un resultado peor, superando el 2.

### 4.1.2 Análisis transversal: resultados por variable

Estudiamos ahora la riqueza léxica en función de las tres variables sociolingüísticas contempladas según las mismas medidas utilizadas antes, añadiendo el análisis de los descriptores estadísticos.

#### Variable sexo

Recordamos que el número de encuestados de los dos conjuntos no es homogéneo (87 mujeres y 13 hombres) por lo que llevamos a cabo la comparación según los promedios para que el desnivel no desvirtúe el análisis. La media general de palabras por texto es de 100 en cada agrupación, las mujeres escriben un total de 63 vocablos y los hombres 62.

A continuación, presentamos los resultados completos a partir de los cuales establecemos el grado de riqueza.

| *Mujer* | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,64 | 64% | 1,23 | 48% | 2,08 |
| 2 | 0,53 | 53% | 1,51 | 55% | 1,82 |
| 3 | 0,57 | 57% | 1,50 | 55% | 1,82 |
| 4 | 0,60 | 60% | 1,43 | 59% | 1,69 |
| 5 | 0,60 | 60% | 1,43 | 55% | 1,82 |
| 6 | 0,60 | 60% | 1,43 | 55% | 1,82 |
| 7 | 0,62 | 62% | 1,32 | 59% | 1,69 |
| 8 | 0,65 | 65% | 1,48 | 60% | 1,67 |
| 9 | 0,63 | 63% | 1,40 | 52% | 1,92 |
| 10 | 0,59 | 59% | 1,51 | 55% | 1,82 |
| 11 | 0,68 | 68% | 1,39 | 63% | 1,59 |
| 12 | 0,53 | 53% | 1,77 | 52% | 1,92 |
| 13 | 0,65 | 65% | 1,35 | 53% | 1,89 |
| 14 | 0,61 | 61% | 1,45 | 43% | 2,33 |
| 15 | 0,53 | 53% | 1,47 | 51% | 1,96 |
| 16 | 0,66 | 66% | 1,14 | 50% | 2 |
| 17 | 0,59 | 59% | 1,51 | 53% | 1,89 |
| 18 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 19 | 0,59 | 59% | 1,51 | 52% | 1,92 |
| 20 | 0,66 | 66% | 1,27 | 55% | 1,82 |
| 21 | 0,71 | 71% | 1,18 | 56% | 1,79 |
| 22 | 0,67 | 67% | 1,31 | 55% | 1,82 |
| 23 | 0,59 | 59% | 1,40 | 55% | 1,82 |
| 24 | 0,56 | 56% | 1,65 | 60% | 1,67 |
| 25 | 0,58 | 58% | 1,41 | 55% | 1,82 |
| 26 | 0,64 | 64% | 1,56 | 60% | 1,67 |
| 27 | 0,62 | 62% | 1,07 | 53% | 1,89 |
| 28 | 0,75 | 75% | 1,29 | 64% | 1,56 |

| | | | | | |
|---|---|---|---|---|---|
| 29 | 0,64 | 64% | 1,25 | 53% | 1,89 |
| 30 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 31 | 0,64 | 64% | 1,25 | 56% | 1,79 |
| 32 | 0,69 | 69% | 1,41 | 66% | 1,52 |
| 33 | 0,59 | 59% | 1,55 | 51% | 1,96 |
| 34 | 0,69 | 69% | 1,33 | 54% | 1,85 |
| 35 | 0,63 | 63% | 1,34 | 50% | 2 |
| 36 | 0,54 | 54% | 1,69 | 49% | 2,04 |
| 37 | 0,63 | 63% | 1,54 | 61% | 1,64 |
| 38 | 0,70 | 70% | 1,27 | 58% | 1,72 |
| 39 | 0,61 | 61% | 1,42 | 60% | 1,67 |
| 40 | 0,62 | 62% | 1,38 | 59% | 1,69 |
| 41 | 0,57 | 57% | 1,43 | 54% | 1,85 |
| 42 | 0,66 | 66% | 1,35 | 60% | 1,67 |
| 43 | 0,66 | 66% | 1,29 | 60% | 1,67 |
| 44 | 0,64 | 64% | 1,36 | 57% | 1,75 |
| 45 | 0,65 | 65% | 1,25 | 55% | 1,82 |
| 46 | 0,61 | 61% | 1,45 | 54% | 1,85 |
| 47 | 0,67 | 67% | 1,37 | 61% | 1,64 |
| 48 | 0,60 | 60% | 1,50 | 54% | 1,85 |
| 49 | 0,59 | 59% | 1,31 | 53% | 1,89 |
| 50 | 0,55 | 55% | 1,45 | 48% | 2,08 |
| 51 | 0,65 | 65% | 1,38 | 63% | 1,59 |
| 52 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 53 | 0,67 | 67% | 1,31 | 60% | 1,67 |
| 54 | 0,62 | 62% | 1,32 | 53% | 1,89 |
| 55 | 0,63 | 63% | 1,50 | 55% | 1,82 |
| 56 | 0,66 | 66% | 1,29 | 57% | 1,75 |
| 57 | 0,68 | 68% | 1,28 | 59% | 1,69 |
| 58 | 0,53 | 53% | 1,56 | 49% | 2,04 |

| | | | | | |
|---|---|---|---|---|---|
| 59 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 60 | 0,64 | 64% | 1,28 | 52% | 1,92 |
| 61 | 0,67 | 67% | 1,34 | 52% | 1,92 |
| 62 | 0,57 | 57% | 1,46 | 50% | 2 |
| 63 | 0,73 | 73% | 1,30 | 56% | 1,79 |
| 64 | 0,66 | 66% | 1,47 | 55% | 1,82 |
| 65 | 0,70 | 70% | 1,37 | 61% | 1,64 |
| 66 | 0,58 | 58% | 1,41 | 46% | 2,17 |
| 67 | 0,56 | 56% | 1,65 | 55% | 1,82 |
| 68 | 0,59 | 59% | 1,40 | 58% | 1,72 |
| 69 | 0,72 | 72% | 1,26 | 59% | 1,69 |
| 70 | 0,67 | 67% | 1,34 | 59% | 1,69 |
| 71 | 0,64 | 64% | 1,33 | 58% | 1,72 |
| 72 | 0,67 | 67% | 1,29 | 52% | 1,92 |
| 73 | 0,61 | 61% | 1,45 | 54% | 1,85 |
| 74 | 0,65 | 65% | 1,41 | 57% | 1,75 |
| 75 | 0,63 | 63% | 1,31 | 49% | 2,04 |
| 76 | 0,66 | 66% | 1,27 | 53% | 1,89 |
| 77 | 0,65 | 65% | 1,30 | 48% | 2,08 |
| 78 | 0,61 | 61% | 1,42 | 53% | 1,89 |
| 79 | 0,63 | 63% | 1,29 | 52% | 1,92 |
| 80 | 0,58 | 58% | 1,53 | 54% | 1,85 |
| 81 | 0,60 | 60% | 1,50 | 57% | 1,75 |
| 82 | 0,69 | 69% | 1,35 | 57% | 1,75 |
| 83 | 0,67 | 67% | 1,34 | 58% | 1,72 |
| 84 | 0,66 | 66% | 1,29 | 55% | 1,82 |
| 85 | 0,67 | 67% | 1,29 | 54% | 1,85 |
| 86 | 0,59 | 59% | 1,48 | 56% | 1,79 |
| 87 | 0,67 | 67% | 1,37 | 57% | 1,75 |
| Promedio | 0,63 | 63% | 1,39 | 55% | 1,83 |

Tabla 44. Índices de RL según la variante mujer.

| *Hombre* | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,66 | 66% | 1,43 | 56% | 1,79 |
| 2 | 0,62 | 62% | 1,24 | 50% | 2 |
| 3 | 0,61 | 61% | 1,27 | 59% | 1,69 |
| 4 | 0,64 | 64% | 1,36 | 57% | 1,75 |
| 5 | 0,54 | 54% | 1,64 | 50% | 2 |
| 6 | 0,62 | 62% | 1,38 | 50% | 2 |
| 7 | 0,62 | 62% | 1,41 | 53% | 1,89 |
| 8 | 0,71 | 71% | 1,25 | 57% | 1,75 |
| 9 | 0,66 | 66% | 1,38 | 56% | 1,79 |
| 10 | 0,64 | 64% | 1,33 | 55% | 1,82 |
| 11 | 0,61 | 61% | 1,45 | 58% | 1,72 |
| 12 | 0,54 | 54% | 1,59 | 50% | 2 |
| 13 | 0,61 | 61% | 1,39 | 50% | 2 |
| Promedio | 0,62 | 62% | 1,39 | 54% | 1,86 |

Tabla 45. Índices de RL según la variante hombre.

### Diversidad léxica

Los descriptivos estadísticos revelan una relación similar entre *tokens* y *types* en los textos, ya que no se detecta una gran discrepancia entre las secciones. El TTR es satisfactorio ya que los datos se acercan más a 1 que a 0. Se observa, en particular, una ligera superioridad de las informantes en el valor máximo.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | Mujer | Hombre |
| Media | 0,63 | 0,62 |
| Mediana | 0,63 | 0,62 |

| | | |
|---|---|---|
| Mínimo | 0,53 | 0,54 |
| Máximo | 0,75 | 0,71 |

Tabla 46. Estadísticos descriptivos de TTR para la variable sexo.

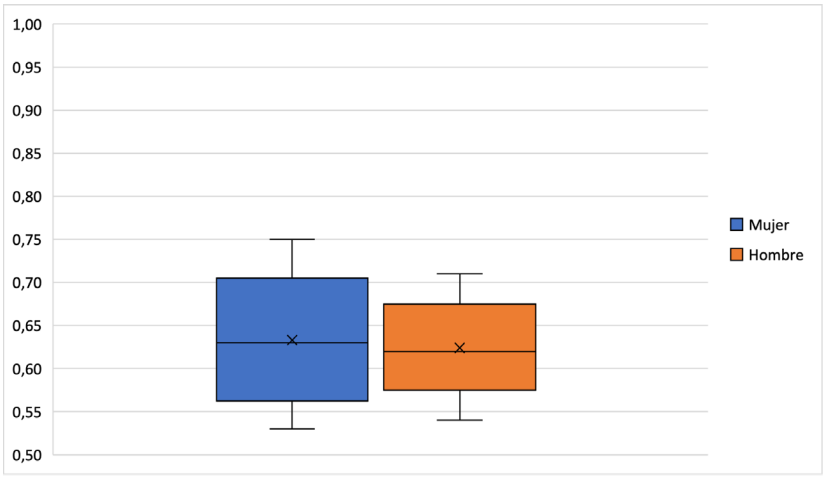


Gráfico 38. Diagrama de cajas de TTR para la variable sexo.

Las respuestas de las mujeres ponen de relieve una desviación intragrupal mayor debido a la mayor extensión de los bigotes de la caja azul, que marca una oscilación superior de los datos con respecto a la caja naranja.

En lo que atañe a la variación léxica, si traducimos en porcentaje los datos descritos, concuerdan con lo que acabamos de discutir: las mujeres tocan el vértice mayor (75%), el menor (53%) y el promedio sigue a su favor (63%), ya que cambia de un punto con respecto de la media alcanzada por los informantes (62%).

| **Descriptivos** | **Variante** | |
|---|---|---|
| | Mujer | Hombre |
| Media | 63% | 62% |
| Mediana | 63% | 62% |
| Mínimo | 53% | 54% |
| Máximo | 75% | 71% |

Tabla 47. Estadísticos descriptivos de la variación léxica para la variable sexo.

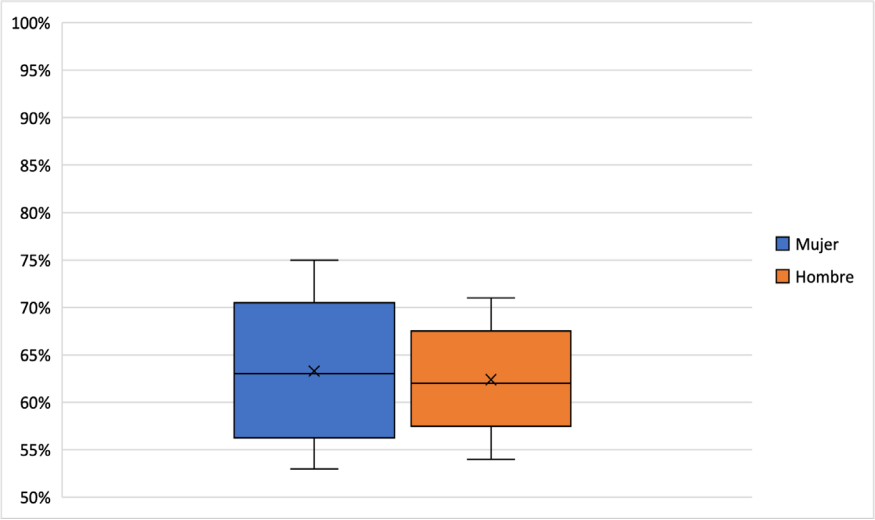


Gráfico 39. Diagrama de cajas de la variación léxica para la variable sexo.

Al examinar pormenorizadamente los valores de cada relato, se hace patente que todos alcanzan más del 50% de variedad textual (el índice mínimo es 53%), con lo cual el grado de repetición de las unidades léxicas es bastante bajo. Entonces, los porcentajes de *types* son mayores en relación con los *tokens*.

Asimismo, el índice de hápax manifiesta una notable oscilación de los resultados en general, pero si tomamos como punto de partida medias y medianas no hay diferencia entre las dos variantes (1,39 y 1,38). Las estudiantes arrojan valores más altos (1,77), en particular en el texto 116 donde hay 30 hápax y 53 vocablos (56%), y a la vez valores más bajos (1,07), sobre todo en el texto 131 se hallan 58 hápax sobre un total de 62 *types* (93%).

| **Descriptivos** | **Variante** | |
|---|---|---|
| | Mujer | Hombre |
| Media | 1,39 | 1,39 |
| Mediana | 1,38 | 1,38 |
| Mínimo | 1,07 | 1,24 |
| Máximo | 1,77 | 1,64 |

Tabla 48. Estadísticos descriptivos del índice de hápax para la variable sexo.

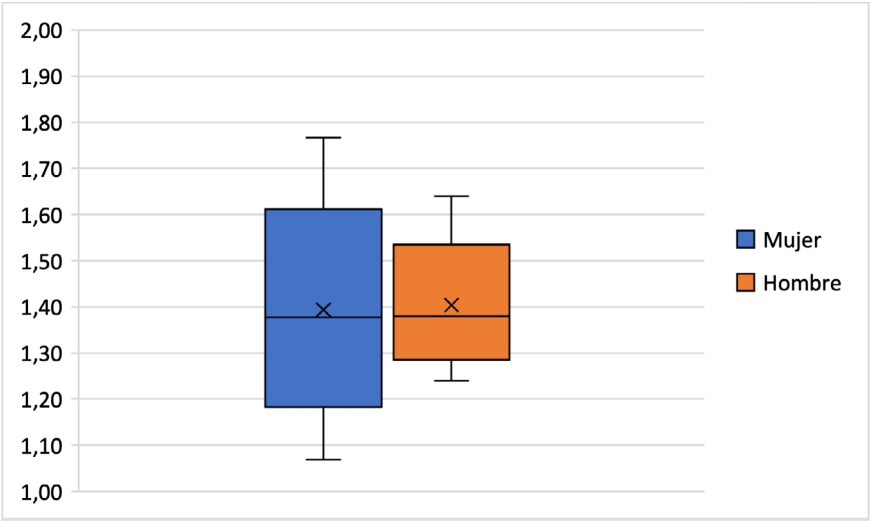


Gráfico 40. Diagrama de cajas del índice de hápax para la variable sexo.

Las mujeres aportan un caudal mayor de hápax como indican las partes inferiores de los diagramas, donde se colocan los resultados mejores (la extensión del bigote, que baja hasta el valor mínimo, y del Q1, que sube hasta la mediana). Los hombres presentan un bigote más corto que implica una menor cantidad de hápax. Por otra parte, la presencia de estas unidades en su producción resulta más compacta y la desviación se reduce (de 1,69 a 2). De todas formas, el 85% de ambas variantes tiene un índice que no supera el 1,50 y, de acuerdo con esto, no se detecta ninguna supremacía: los resultados son satisfactorios en sendos casos.

### Densidad léxica

La aplicación del índice de la densidad léxica es significativa, por lo general, ya que la mayoría de las producciones alcanza o supera el 50%. La riqueza de los dos subgrupos se reparte equitativamente pese a una ligera diferencia entre los valores medios: media y mediana de las mujeres se solapan (55%) mientras que la media matemática (54%) se posiciona por debajo de la mediana de un punto (55%) para los hombres.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | Mujer | Hombre |
| Media | 55% | 54% |

| | | |
|---|---|---|
| Mediana | 55% | 55% |
| Mínimo | 43% | 50% |
| Máximo | 66% | 59% |

Tabla 49. Estadísticos descriptivos de la densidad léxica para la variable sexo.

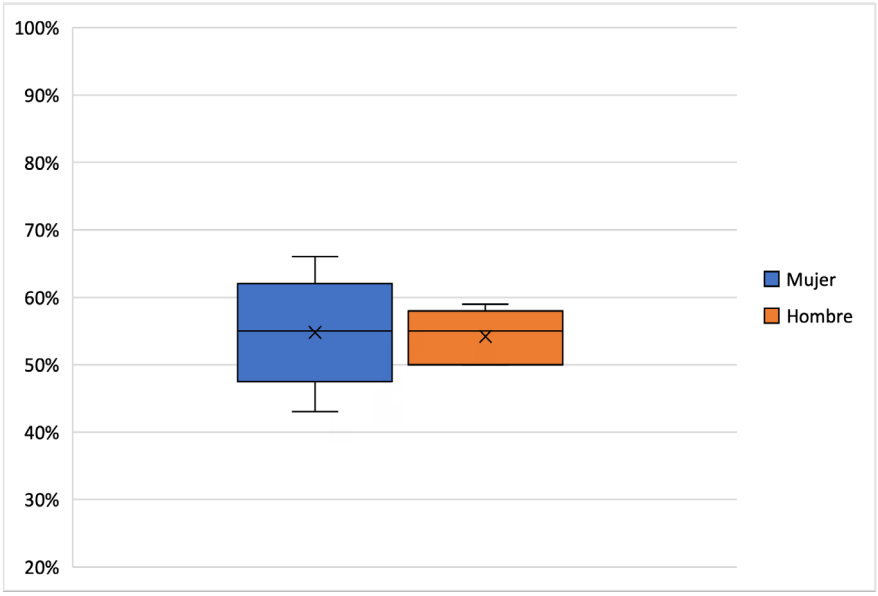


Gráfico 41. Diagrama de cajas de la densidad léxica para la variable sexo.

Las participantes siguen presentando el valor superior (66%) y el inferior (43%), con lo cual el índice es más difuso y la desviación es mayor ya que el intervalo de datos entre cuartiles es más amplio. De ahí que la caja azul sea más alargada con respecto a la naranja que presenta valores más compactos, de 50% a 59%.

Al desglosar los resultados de cada relato, sobresale que el 100% de los hombres iguala o sobrepasa el 50% frente al 89% de las producciones de la componente femenina: la mayoría de los varones aporta una cantidad más elevada de palabras semánticas. El diagrama que representa los resultados de las mujeres se extiende hasta el extremo inferior a partir del Q1 –que coincide con el límite del diagrama masculino– donde se hallan índices más pequeños.

Por su parte, los datos del IAT revelan que, otra vez, las mujeres ofrecen simultáneamente el mejor (1,52) y el peor (2,33) resultado. Las medianas coinciden entre los dos muestreos (1,82) y las medias cambian, de poco, favoreciendo las encuestadas ya que el promedio (1,83) es levemente más pequeño con respecto al de los compañeros (1,86).

| **Descriptivos** | **Variante** | |
|---|---|---|
| | Mujer | Hombre |
| Media | 1,83 | 1,86 |
| Mediana | 1,82 | 1,82 |
| Mínimo | 1,52 | 1,69 |
| Máximo | 2,33 | 2 |

Tabla 50. Estadísticos descriptivos de IAT para la variable sexo.

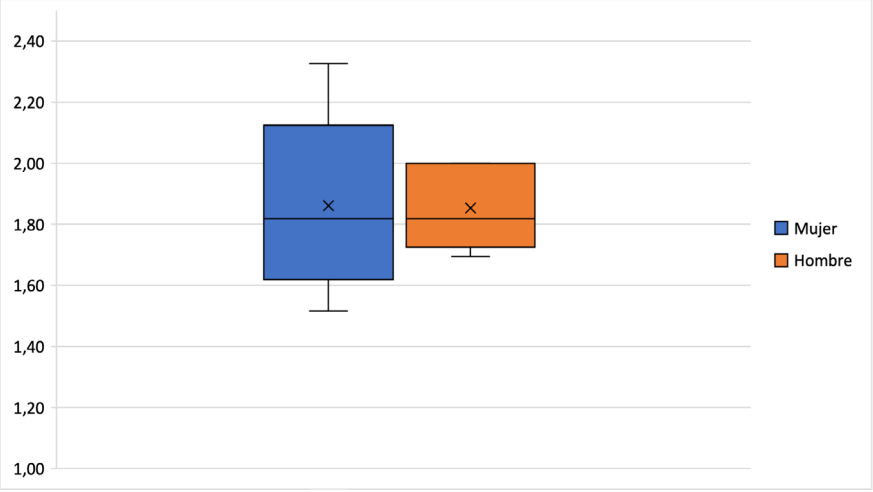


Gráfico 42. Diagrama de cajas de IAT para la variable sexo.

En definitiva, los valores promediales manifiestan cierta simetría entre las dos variantes. La caja naranja se solapa con la parte central de la azul poniendo de relieve la desviación menor de los varones (de 1,69 a 2) y mayor de las alumnas que logran cifras más variadas (de 1,52 a 2,33). En todo caso, el grado de riqueza léxica no es alto porque los resultados de la

fórmula se avecinan más a 2 que a 1, el valor ideal.[41] El gráfico ilustra la situación descrita: los intervalos que van del Q2 (la línea de la mediana) al Q3 y suben hasta el Q4, son más amplios comparados con la distancia comprendida entre la mediana y el Q1, en particular en la caja azul.

### Variable nivel de ELE

La repartición de los informantes presenta dos secciones homogéneas de 50 sujetos que revelan un nivel de riqueza léxica casi idéntico: el promedio por informante de *tokens* es de 100 en ambas variantes y el promedio de *types* es de 62 en el grupo de nivel B1 y de 63 en el B2.

| ***Nivel B1*** | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,64 | 64% | 1,23 | 48% | 2,08 |
| 2 | 0,53 | 53% | 1,51 | 55% | 1,82 |
| 3 | 0,66 | 66% | 1,43 | 56% | 1,79 |
| 4 | 0,62 | 62% | 1,24 | 50% | 2 |
| 5 | 0,61 | 61% | 1,27 | 59% | 1,69 |
| 6 | 0,57 | 57% | 1,50 | 55% | 1,82 |
| 7 | 0,60 | 60% | 1,43 | 59% | 1,69 |
| 8 | 0,60 | 60% | 1,43 | 55% | 1,82 |
| 9 | 0,60 | 60% | 1,43 | 55% | 1,82 |
| 10 | 0,64 | 64% | 1,36 | 57% | 1,75 |
| 11 | 0,62 | 62% | 1,32 | 59% | 1,69 |
| 12 | 0,65 | 65% | 1,48 | 60% | 1,67 |
| 13 | 0,63 | 63% | 1,40 | 52% | 1,92 |
| 14 | 0,59 | 59% | 1,51 | 55% | 1,82 |
| 15 | 0,68 | 68% | 1,39 | 63% | 1,59 |
| 16 | 0,53 | 53% | 1,77 | 52% | 1,92 |
| 17 | 0,65 | 65% | 1,35 | 53% | 1,89 |

[41] Tomamos esta medida de corte, tal cual planteamos en el análisis del índice de hápax.

| | | | | | |
|---|---|---|---|---|---|
| 18 | 0,61 | 61% | 1,45 | 43% | 2,33 |
| 19 | 0,53 | 53% | 1,47 | 51% | 1,96 |
| 20 | 0,66 | 66% | 1,14 | 50% | 2 |
| 21 | 0,59 | 59% | 1,51 | 53% | 1,89 |
| 22 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 23 | 0,59 | 59% | 1,51 | 52% | 1,92 |
| 24 | 0,66 | 66% | 1,27 | 55% | 1,82 |
| 25 | 0,71 | 71% | 1,18 | 56% | 1,79 |
| 26 | 0,67 | 67% | 1,31 | 55% | 1,82 |
| 27 | 0,59 | 59% | 1,40 | 55% | 1,82 |
| 28 | 0,56 | 56% | 1,65 | 60% | 1,67 |
| 29 | 0,58 | 58% | 1,41 | 55% | 1,82 |
| 30 | 0,64 | 64% | 1,56 | 60% | 1,67 |
| 31 | 0,62 | 62% | 1,07 | 53% | 1,89 |
| 32 | 0,75 | 75% | 1,29 | 64% | 1,56 |
| 33 | 0,64 | 64% | 1,25 | 53% | 1,89 |
| 34 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 35 | 0,64 | 64% | 1,25 | 56% | 1,79 |
| 36 | 0,69 | 69% | 1,41 | 66% | 1,52 |
| 37 | 0,59 | 59% | 1,55 | 51% | 1,96 |
| 38 | 0,69 | 69% | 1,33 | 54% | 1,85 |
| 39 | 0,63 | 63% | 1,34 | 50% | 2 |
| 40 | 0,54 | 54% | 1,69 | 49% | 2,04 |
| 41 | 0,63 | 63% | 1,54 | 61% | 1,64 |
| 42 | 0,70 | 70% | 1,27 | 58% | 1,72 |
| 43 | 0,61 | 61% | 1,42 | 60% | 1,67 |
| 44 | 0,62 | 62% | 1,38 | 59% | 1,69 |
| 45 | 0,57 | 57% | 1,43 | 54% | 1,85 |
| 46 | 0,66 | 66% | 1,35 | 60% | 1,67 |
| 47 | 0,66 | 66% | 1,29 | 60% | 1,67 |
| 48 | 0,64 | 64% | 1,36 | 57% | 1,75 |
| 49 | 0,65 | 65% | 1,25 | 55% | 1,82 |

| | | | | | |
|---|---|---|---|---|---|
| 50 | 0,61 | 61% | 1,45 | 54% | 1,85 |
| Promedio | 0,62 | 62% | 1,39 | 55% | 1,82 |

Tabla 51. Índices de RL según la variante nivel B1.

| ***Nivel B2*** | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,67 | 67% | 1,37 | 61% | 1,64 |
| 2 | 0,54 | 54% | 1,64 | 50% | 2 |
| 3 | 0,62 | 62% | 1,38 | 50% | 2 |
| 4 | 0,60 | 60% | 1,50 | 54% | 1,85 |
| 5 | 0,59 | 59% | 1,31 | 53% | 1,89 |
| 6 | 0,62 | 62% | 1,41 | 53% | 1,89 |
| 7 | 0,55 | 55% | 1,45 | 48% | 2,08 |
| 8 | 0,65 | 65% | 1,38 | 63% | 1,59 |
| 9 | 0,71 | 71% | 1,25 | 57% | 1,75 |
| 10 | 0,66 | 66% | 1,38 | 56% | 1,79 |
| 11 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 12 | 0,67 | 67% | 1,31 | 60% | 1,67 |
| 13 | 0,62 | 62% | 1,32 | 53% | 1,89 |
| 14 | 0,64 | 64% | 1,33 | 55% | 1,82 |
| 15 | 0,61 | 61% | 1,45 | 58% | 1,72 |
| 16 | 0,63 | 63% | 1,50 | 55% | 1,82 |
| 17 | 0,54 | 54% | 1,59 | 50% | 2 |
| 18 | 0,66 | 66% | 1,29 | 57% | 1,75 |
| 19 | 0,68 | 68% | 1,28 | 59% | 1,69 |
| 20 | 0,53 | 53% | 1,56 | 49% | 2,04 |
| 21 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 22 | 0,64 | 64% | 1,28 | 52% | 1,92 |
| 23 | 0,67 | 67% | 1,34 | 52% | 1,92 |
| 24 | 0,57 | 57% | 1,46 | 50% | 2 |

| | | | | | |
|---|---|---|---|---|---|
| 25 | 0,73 | 73% | 1,30 | 56% | 1,79 |
| 26 | 0,66 | 66% | 1,47 | 55% | 1,82 |
| 27 | 0,70 | 70% | 1,37 | 61% | 1,64 |
| 28 | 0,58 | 58% | 1,41 | 46% | 2,17 |
| 29 | 0,56 | 56% | 1,65 | 55% | 1,82 |
| 30 | 0,59 | 59% | 1,40 | 58% | 1,72 |
| 31 | 0,72 | 72% | 1,26 | 59% | 1,69 |
| 32 | 0,67 | 67% | 1,34 | 59% | 1,69 |
| 33 | 0,64 | 64% | 1,33 | 58% | 1,72 |
| 34 | 0,67 | 67% | 1,29 | 52% | 1,92 |
| 35 | 0,61 | 61% | 1,45 | 54% | 1,85 |
| 36 | 0,65 | 65% | 1,41 | 57% | 1,75 |
| 37 | 0,63 | 63% | 1,31 | 49% | 2,04 |
| 38 | 0,66 | 66% | 1,27 | 53% | 1,89 |
| 39 | 0,65 | 65% | 1,30 | 48% | 2,08 |
| 40 | 0,61 | 61% | 1,42 | 53% | 1,89 |
| 41 | 0,63 | 63% | 1,29 | 52% | 1,92 |
| 42 | 0,58 | 58% | 1,53 | 54% | 1,85 |
| 43 | 0,60 | 60% | 1,50 | 57% | 1,75 |
| 44 | 0,69 | 69% | 1,35 | 57% | 1,75 |
| 45 | 0,67 | 67% | 1,34 | 58% | 1,72 |
| 46 | 0,66 | 66% | 1,29 | 55% | 1,82 |
| 47 | 0,67 | 67% | 1,29 | 54% | 1,85 |
| 48 | 0,59 | 59% | 1,48 | 56% | 1,79 |
| 49 | 0,67 | 67% | 1,37 | 57% | 1,75 |
| 50 | 0,61 | 61% | 1,39 | 50% | 2 |
| Promedio | 0,63 | 63% | 1,39 | 54% | 1,85 |

Tabla 52. Índices de RL según la variante nivel B2.

### Diversidad léxica

La *Type/Token Ratio* destaca una distribución casi equivalente de los datos. El valor mínimo es el mismo para los dos niveles (0,53). Los estudiantes más avanzados sobrepasan, de poco, a los compañeros en los valores estadísticos de medio (0,01), mientras que el grupo intermedio supera en el extremo máximo, donde alcanza un 0,75 (contra el 0,73 de los otros).

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B1 | B2 |
| Media | 0,62 | 0,63 |
| Mediana | 0,63 | 0,64 |
| Mínimo | 0,53 | 0,53 |
| Máximo | 0,75 | 0,73 |

Tabla 53. Estadísticos descriptivos de TTR para la variable nivel de ELE.

Los muestreos presentan una oscilación intragrupal casi igual, parece que a este punto del proceso de aprendizaje el grado de lengua adquirido no cambia la capacidad productiva en términos de variedad. En este sentido, las cajas presentan aproximadamente los mismos intervalos entre cuartiles.

La tabla 54 corrobora la ausencia de un efecto de este factor en la variación léxica. Las dos agrupaciones comparten el resultado más bajo (53%), los encuestados de nivel menor tocan el extremo superior (75%) y los otros alcanzan una media (63%) y una mediana (64%) mayores, pero, al fin y al cabo, la oscilación es mínima.

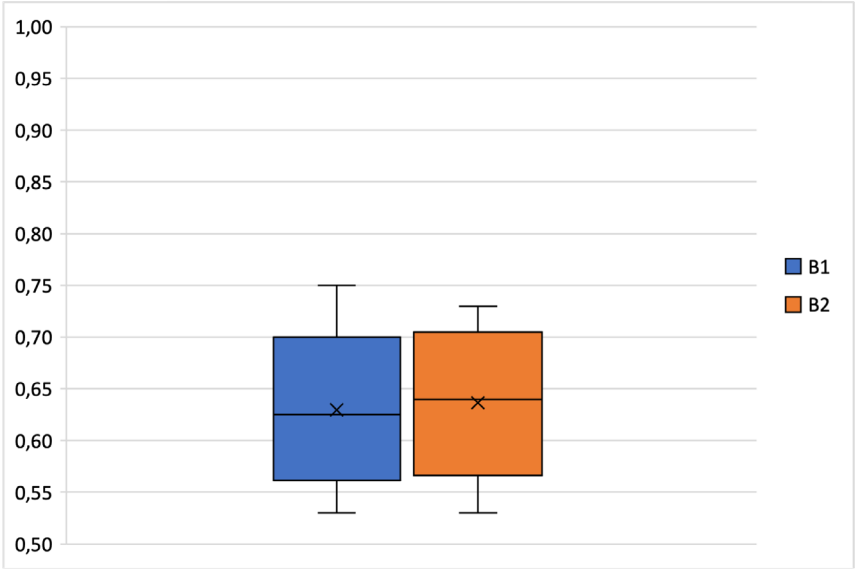


Gráfico 43. Diagrama de cajas de TTR para la variable nivel de ELE.

| Descriptivos | Variante | |
|---|---|---|
| | B1 | B2 |
| Media | 62% | 63% |
| Mediana | 63% | 64% |
| Mínimo | 53% | 53% |
| Máximo | 75% | 73% |

Tabla 54. Estadísticos descriptivos de la variación léxica para la variable nivel de ELE.

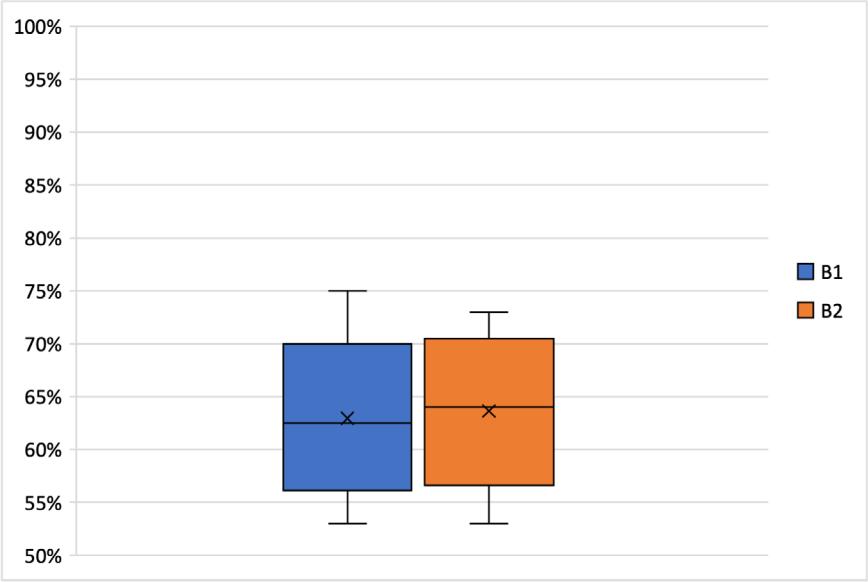


Gráfico 44. Diagrama de cajas de la variación léxica para la variable nivel de ELE.

En suma, la variedad supera el 50% del índice sin distinción de nivel y, sin embargo, el grado de competencia lingüística no ejerce un influjo tan fuerte como para alterar los resultados.

El índice de hápax pone de relieve una gran oscilación de los resultados, si excluimos la media que es igual en los dos conjuntos. En efecto, los alumnos de nivel B1 aportan, a la vez, el mejor valor (1,07) y el peor (1,77). Obtienen el índice más pequeño en el texto 131 que contiene la mayor cantidad de hápax y el valor más grande en el 116 que denota una escasa presencia. Los B2 arrojan datos más cerrados que van de 1,25 a 1,65.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B1 | B2 |
| Media | 1,39 | 1,39 |
| Mediana | 1,39 | 1,37 |
| Mínimo | 1,07 | 1,25 |
| Máximo | 1,77 | 1,65 |

Tabla 55. Estadísticos descriptivos del índice de hápax para la variable nivel de ELE.

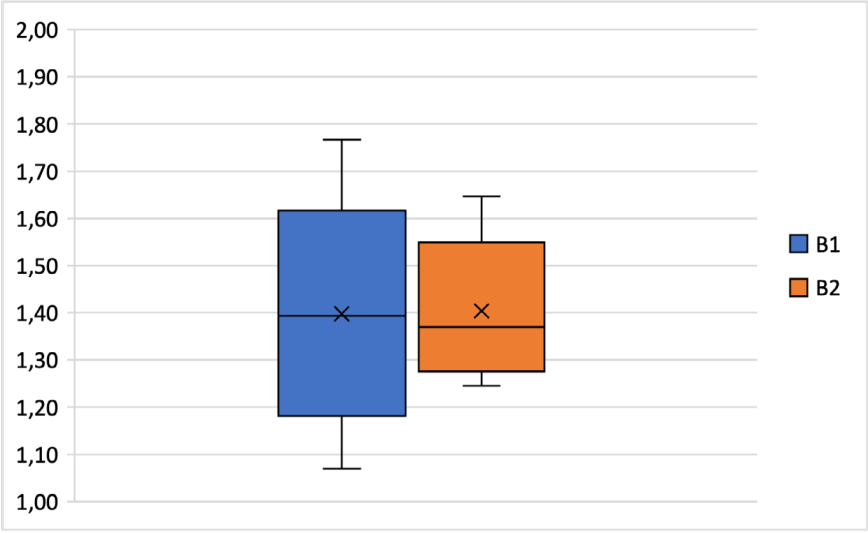


Gráfico 45. Diagrama de cajas del índice de hápax para la variable nivel de ELE.

Los resultados de los informantes de nivel B1 oscilan más que los de nivel B2, donde la variación intragrupal es menor, por eso la caja azul es más alargada. La naranja, de hecho, es más pequeña como la extensión de los bigotes que apenas superan el Q3 de la azul y no logran el límite del Q1. De todas formas, el 80% de los relatos de los B1 y el 90% de los B2 tienen un índice de hápax aceptable.

### Densidad léxica

La densidad revela que los resultados de los encuestados de nivel intermedio varían más en el uso de las palabras semánticas comparados con los otros, que tienen una menor desviación. Se registra mayor variación, entonces, en los B1 que llegan a los extremos superior (66%) e inferior (43%) del corpus, con una oscilación de 23 puntos frente a los 17 de los B2.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B1 | B2 |
| Media | 55% | 54% |
| Mediana | 55% | 55% |
| Mínimo | 43% | 46% |
| Máximo | 66% | 63% |

Tabla 56. Estadísticos descriptivos de la densidad léxica para la variable nivel de ELE.

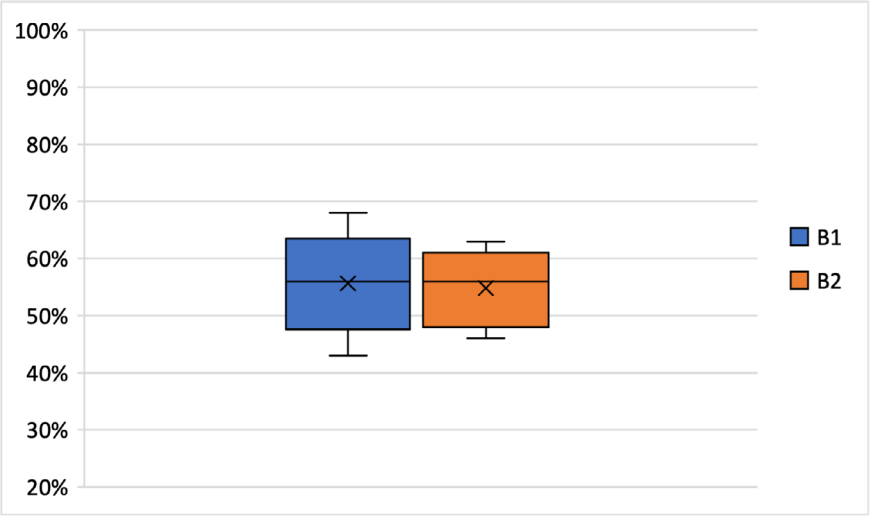


Gráfico 46. Diagrama de cajas de la densidad léxica para la variable nivel de ELE.

La variación intergrupal no es tan significativa para que el nivel de español afecte a este índice, como se desprende de las proporciones de los intervalos entre cuartiles. Lo que llama la atención es la longitud diferente de los bigotes que marcan los extremos de cada agrupación. Examinando el índice de cada texto, el 92% de los B1 iguala o sobrepasa el 50% de densidad léxica frente al 88% de los B2. La mayoría de los alumnos de nivel inferior ha escrito una cantidad mayor de palabras temáticas. El diagrama marca esta diferencia: la caja azul sobrepasa la naranja alargándose hasta el extremo superior, donde se colocan los resultados mejores.

Para terminar, el IAT pone de manifiesto una situación parecida a lo visto hasta ahora: el índice es alto y supone la presencia de una palabra nocional casi cada dos palabras funcionales. La media matemática y la mediana favorecen, ligeramente, el nivel B1, cuyos textos tocan los polos mínimo y máximo manifestando una mayor desviación intragrupal (desde 1,52 hasta 2,33, contra los estudiantes avanzados que van de 1,59 a 2,17).

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B1 | B2 |
| Media | 1,82 | 1,85 |

| | | |
|---|---|---|
| Mediana | 1,82 | 1,84 |
| Mínimo | 1,52 | 1,59 |
| Máximo | 2,33 | 2,17 |

Tabla 57. Estadísticos descriptivos de IAT para la variable nivel de ELE.

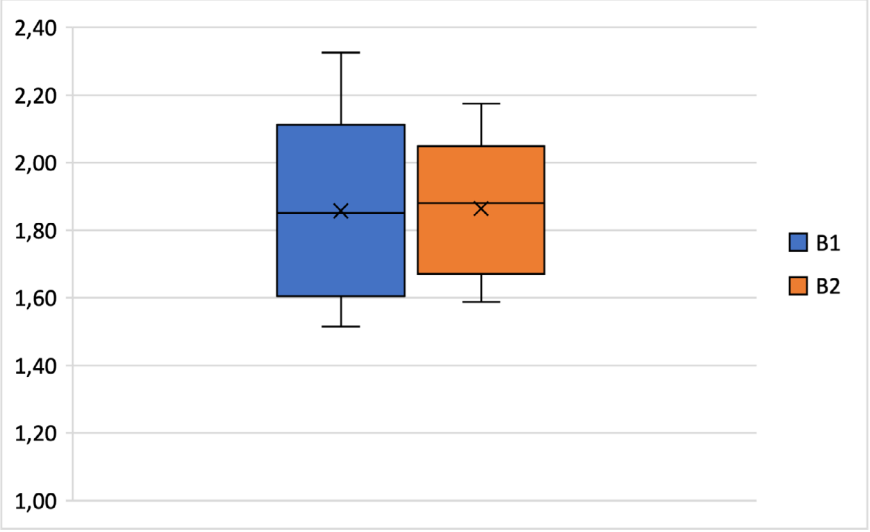


Gráfico 47. Diagrama de cajas de IAT para la variable nivel de ELE.

La disparidad de la desviación intergrupal es evidente y radica en la configuración de los diagramas, tanto en el tamaño de las cajas como en la amplitud intercuartil y la extensión de los bigotes. Los relatos de los alumnos de nivel B1 se caracterizan por una oscilación mayor mientras que los de los participantes de nivel B2 son más compactos, esto confirma la tendencia surgida anteriormente.

### Variable conocimiento de otras LE

En el ámbito de la disponibilidad léxica son numerosos los estudios que incluyen la variable conocimiento de otras lenguas extranjeras, mientras que este análisis es pionero en el campo de aplicación de la riqueza léxica, ya que no tenemos constancia de investigaciones que examinan este condicionante. Partiendo de la *Hipótesis de la Interdependencia Lingüística* averiguamos si un sujeto que ya tiene habilidades lingüísticas

relacionadas con el uso de un idioma extranjero se apoya en dichas habilidades para aprender un nuevo idioma.

Disponemos de 19 informantes que conocen dos LE y 81 que conocen más, con lo cual trabajamos con los valores promediales partiendo de una media de 100 *tokens* por cada texto y de 61 *types* en los textos del grupo =2 LE y 63 en el >2 LE.

| ***Conoce 2 LE*** | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,54 | 54% | 1,24 | 50% | 2 |
| 2 | 0,62 | 62% | 1,36 | 50% | 2 |
| 3 | 0,60 | 60% | 1,40 | 54% | 1,85 |
| 4 | 0,55 | 55% | 1,65 | 48% | 2,08 |
| 5 | 0,56 | 56% | 1,41 | 55% | 1,82 |
| 6 | 0,67 | 67% | 1,39 | 59% | 1,69 |
| 7 | 0,67 | 67% | 1,41 | 58% | 1,72 |
| 8 | 0,66 | 66% | 1,69 | 55% | 1,82 |
| 9 | 0,61 | 61% | 1,54 | 50% | 2 |
| 10 | 0,62 | 62% | 1,29 | 50% | 2 |
| 11 | 0,64 | 64% | 1,64 | 57% | 1,75 |
| 12 | 0,59 | 59% | 1,38 | 55% | 1,82 |
| 13 | 0,56 | 56% | 1,50 | 60% | 1,67 |
| 14 | 0,58 | 58% | 1,45 | 55% | 1,82 |
| 15 | 0,61 | 61% | 1,65 | 48% | 2,08 |
| 16 | 0,69 | 69% | 1,34 | 66% | 1,52 |
| 17 | 0,54 | 54% | 1,34 | 49% | 2,04 |
| 18 | 0,63 | 63% | 1,29 | 61% | 1,64 |
| 19 | 0,66 | 66% | 1,39 | 60% | 1,67 |
| Promedio | 0,61 | 61% | 1,44 | 55% | 1,84 |

Tabla 58. Índices de RL según la variante LE =2.

| *Conoce más de 2 LE* | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,67 | 67% | 1,23 | 61% | 1,64 |
| 2 | 0,59 | 59% | 1,51 | 53% | 1,89 |
| 3 | 0,62 | 62% | 1,43 | 53% | 1,89 |
| 4 | 0,65 | 65% | 1,27 | 63% | 1,59 |
| 5 | 0,71 | 71% | 1,50 | 57% | 1,75 |
| 6 | 0,66 | 66% | 1,43 | 56% | 1,79 |
| 7 | 0,61 | 61% | 1,43 | 48% | 2,08 |
| 8 | 0,67 | 67% | 1,43 | 60% | 1,67 |
| 9 | 0,62 | 62% | 1,32 | 53% | 1,89 |
| 10 | 0,64 | 64% | 1,48 | 55% | 1,82 |
| 11 | 0,61 | 61% | 1,40 | 58% | 1,72 |
| 12 | 0,63 | 63% | 1,51 | 55% | 1,82 |
| 13 | 0,54 | 54% | 1,39 | 50% | 2 |
| 14 | 0,66 | 66% | 1,77 | 57% | 1,75 |
| 15 | 0,68 | 68% | 1,35 | 59% | 1,69 |
| 16 | 0,53 | 53% | 1,45 | 49% | 2,04 |
| 17 | 0,65 | 65% | 1,47 | 52% | 1,92 |
| 18 | 0,64 | 64% | 1,14 | 52% | 1,92 |
| 19 | 0,67 | 67% | 1,51 | 52% | 1,92 |
| 20 | 0,57 | 57% | 1,27 | 50% | 2 |
| 21 | 0,73 | 73% | 1,51 | 56% | 1,79 |
| 22 | 0,66 | 66% | 1,27 | 55% | 1,82 |
| 23 | 0,70 | 70% | 1,18 | 61% | 1,64 |
| 24 | 0,58 | 58% | 1,31 | 46% | 2,17 |
| 25 | 0,59 | 59% | 1,56 | 58% | 1,72 |
| 26 | 0,72 | 72% | 1,07 | 59% | 1,69 |
| 27 | 0,64 | 64% | 1,29 | 58% | 1,72 |
| 28 | 0,67 | 67% | 1,25 | 52% | 1,92 |
| 29 | 0,61 | 61% | 1,25 | 54% | 1,85 |

| | | | | | |
|---|---|---|---|---|---|
| 30 | 0,65 | 65% | 1,55 | 57% | 1,75 |
| 31 | 0,63 | 63% | 1,33 | 49% | 2,04 |
| 32 | 0,66 | 66% | 1,34 | 53% | 1,89 |
| 33 | 0,65 | 65% | 1,27 | 48% | 2,08 |
| 34 | 0,61 | 61% | 1,42 | 53% | 1,89 |
| 35 | 0,63 | 63% | 1,38 | 52% | 1,92 |
| 36 | 0,58 | 58% | 1,43 | 54% | 1,85 |
| 37 | 0,60 | 60% | 1,35 | 57% | 1,75 |
| 38 | 0,69 | 69% | 1,36 | 57% | 1,75 |
| 39 | 0,67 | 67% | 1,25 | 54% | 1,85 |
| 40 | 0,59 | 59% | 1,45 | 56% | 1,79 |
| 41 | 0,67 | 67% | 1,37 | 57% | 1,75 |
| 42 | 0,64 | 64% | 1,31 | 48% | 2,08 |
| 43 | 0,53 | 53% | 1,41 | 55% | 1,82 |
| 44 | 0,66 | 66% | 1,38 | 56% | 1,79 |
| 45 | 0,61 | 61% | 1,25 | 59% | 1,69 |
| 46 | 0,57 | 57% | 1,38 | 55% | 1,82 |
| 47 | 0,60 | 60% | 1,39 | 59% | 1,69 |
| 48 | 0,60 | 60% | 1,31 | 55% | 1,82 |
| 49 | 0,60 | 60% | 1,32 | 55% | 1,82 |
| 50 | 0,62 | 62% | 1,33 | 59% | 1,69 |
| 51 | 0,65 | 65% | 1,45 | 60% | 1,67 |
| 52 | 0,63 | 63% | 1,50 | 52% | 1,92 |
| 53 | 0,59 | 59% | 1,59 | 55% | 1,82 |
| 54 | 0,68 | 68% | 1,29 | 63% | 1,59 |
| 55 | 0,53 | 53% | 1,28 | 52% | 1,92 |
| 56 | 0,65 | 65% | 1,56 | 53% | 1,89 |
| 57 | 0,61 | 61% | 1,27 | 43% | 2,33 |
| 58 | 0,53 | 53% | 1,28 | 51% | 1,96 |
| 59 | 0,66 | 66% | 1,34 | 50% | 2 |

| | | | | | |
|---|---|---|---|---|---|
| 60 | 0,59 | 59% | 1,46 | 53% | 1,89 |
| 61 | 0,65 | 65% | 1,30 | 52% | 1,92 |
| 62 | 0,59 | 59% | 1,47 | 52% | 1,92 |
| 63 | 0,66 | 66% | 1,37 | 55% | 1,82 |
| 64 | 0,71 | 71% | 1,41 | 56% | 1,79 |
| 65 | 0,67 | 67% | 1,40 | 55% | 1,82 |
| 66 | 0,64 | 64% | 1,26 | 60% | 1,67 |
| 67 | 0,62 | 62% | 1,33 | 53% | 1,89 |
| 68 | 0,75 | 75% | 1,29 | 64% | 1,56 |
| 69 | 0,64 | 64% | 1,45 | 53% | 1,89 |
| 70 | 0,64 | 64% | 1,41 | 56% | 1,79 |
| 71 | 0,59 | 59% | 1,31 | 51% | 1,96 |
| 72 | 0,69 | 69% | 1,27 | 54% | 1,85 |
| 73 | 0,63 | 63% | 1,30 | 50% | 2 |
| 74 | 0,70 | 70% | 1,42 | 58% | 1,72 |
| 75 | 0,61 | 61% | 1,29 | 60% | 1,67 |
| 76 | 0,62 | 62% | 1,53 | 59% | 1,69 |
| 77 | 0,57 | 57% | 1,50 | 54% | 1,85 |
| 78 | 0,66 | 66% | 1,35 | 60% | 1,67 |
| 79 | 0,64 | 64% | 1,29 | 57% | 1,75 |
| 80 | 0,65 | 65% | 1,48 | 55% | 1,82 |
| 81 | 0,61 | 61% | 1,37 | 54% | 1,85 |
| Promedio | 0,63 | 63% | 1,38 | 55% | 1,83 |

Tabla 59. Índices de RL según la variante LE >2.

### Diversidad léxica

En principio, la *Type/Token Ratio* se inclina a favor de los estudiantes que conocen más de dos LE ya que los promedios mejores se localizan en sus producciones, así como el valor máximo. Sin embargo, esto no pesa demasiado en los resultados del grupo que alcanza un nivel mayor de riqueza léxica.

| Descriptivos | Variante | |
|---|---|---|
| | LE =2 | LE >2 |
| Media | 0,61 | 0,63 |
| Mediana | 0,61 | 0,64 |
| Mínimo | 0,54 | 0,53 |
| Máximo | 0,69 | 0,75 |

Tabla 60. Estadísticos descriptivos de TTR para la variable conocimiento de otras LE.

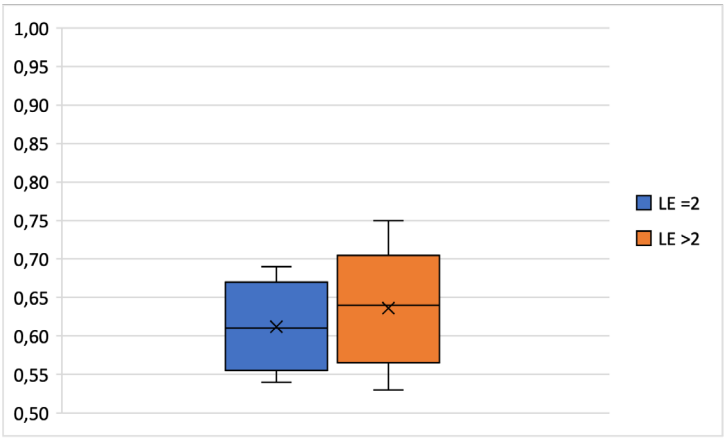


Gráfico 48. Diagrama de cajas de TTR para la variable conocimiento de otras LE.

El bigote del diagrama naranja, correspondiente a los datos recabados de los textos de los alumnos >2 LE, se extiende hasta el punto extremo superior (0,75). Análogamente, se observa una cantidad notable de datos que sobrepasan los resultados obtenidos por los compañeros, que no superan el 0,69. De acuerdo con esta distribución, vemos una disimetría en el comportamiento de las dos variantes que marca una superioridad del 9% de los informantes del grupo LE >2.

La variación léxica corrobora la tesis de Cummins confirmado un aumento de la riqueza en los textos de los informantes que saben más de dos LE. Los demás estudiantes tienen un índice reducido, adecuado a su nivel de ELE desde el momento en el que todos lograron más del 50% de variación, hecho que, sin duda, permite afirmar que la variedad léxica satisface las expectativas.

| Descriptivos | Variante | |
|---|---|---|
| | LE =2 | LE >2 |
| Media | 61% | 63% |
| Mediana | 61% | 64% |
| Mínimo | 54% | 53% |
| Máximo | 69% | 75% |

Tabla 61. Estadísticos descriptivos de la variación léxica para la variable conocimiento de otras LE.

De todas formas, no sorprende el desnivel expresado por la asimetría de los dos diagramas que patentiza el rendimiento mejor de un grupo frente al otro. La misma tendencia se manifiesta en el análisis del índice de hápax, los informantes que conocen más de dos LE alcanzan el mejor resultado (1,07) que supone un incremento del 16%, pese a que también toquen el índice peor (1,77). Asimismo, su media y mediana se revelan más bajas que las del otro grupo (respectivamente de 4% y 2%).

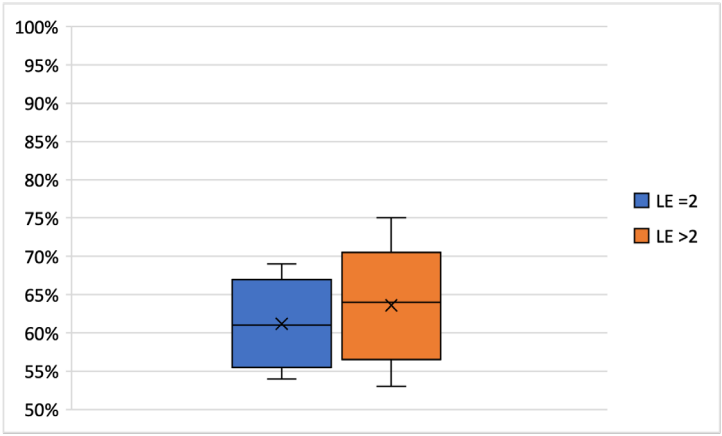


Gráfico 49. Diagrama de cajas de la variación léxica para la variable conocimiento de otras LE.

| Descriptivos | Variante | |
|---|---|---|
| | LE =2 | LE >2 |
| Media | 1,44 | 1,38 |

| | | |
|---|---|---|
| Mediana | 1,40 | 1,37 |
| Mínimo | 1,24 | 1,07 |
| Máximo | 1,69 | 1,77 |

Tabla 62. Estadísticos descriptivos del índice de hápax para la variable conocimiento de otras LE.

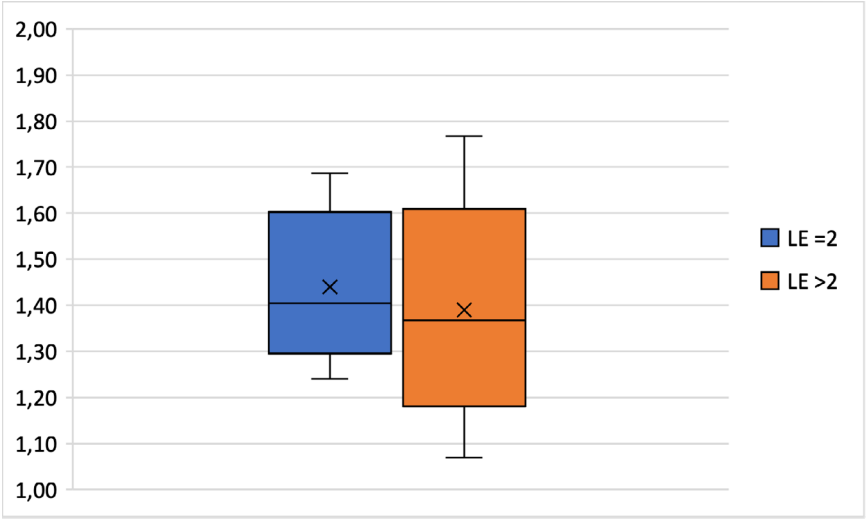


Gráfico 50. Diagrama de cajas del índice de hápax para la variable conocimiento de otras LE.

La mayor parte de los textos que están posicionados por debajo del Q2, que contiene los datos mejores, son los escritos de los alumnos del grupo LE >2: el 88% de sus relatos tiene un índice igual o menor de 1,50 frente al 74% de la variante =2 LE. Se patentiza una producción mayor de hápax ya que, comparada con la naranja, la amplitud intercuartil Q2-Q3 azul es reducida. Todo esto revela la influencia de la variable en la capacidad productiva de los participantes.

### Densidad léxica

El estudio de la densidad léxica supone un cambio de tendencia, los encuestados que conocen dos LE arrojan resultados más altos. Los datos estadísticos son mayores, pese a que la desviación propia del otro mues-

treo sea superior (de 43% a 64% *versus* de 48% a 66%). Por su parte, media y mediana se corresponden llegando a un porcentaje apreciable del 55%.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | LE =2 | LE >2 |
| Media | 55% | 55% |
| Mediana | 55% | 55% |
| Mínimo | 48% | 43% |
| Máximo | 66% | 64% |

Tabla 63. Estadísticos descriptivos de la densidad léxica para la variable conocimiento de otras LE.

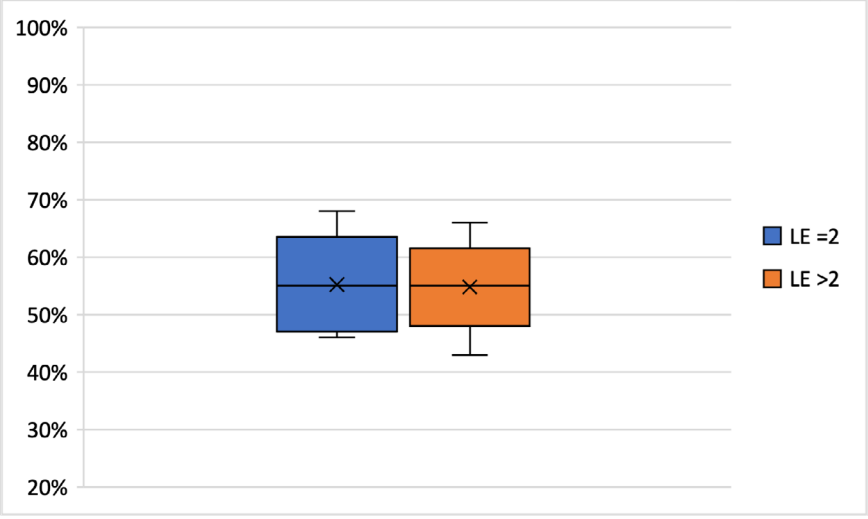


Gráfico 51. Diagrama de cajas de la densidad léxica para la variable conocimiento de otras LE.

La variante LE =2 presenta un rendimiento mejor, la amplitud intercuartil Q2-Q4 de la caja azul se alarga más, el bigote superior sube hasta el punto máximo de 66% y el inferior no sobrepasa el 48%. Al contrario, el bigote inferior de la otra variante baja hasta el 43%, obteniendo el peor índice en absoluto. Examinando cada relato, el 84% de los escritos del primer muestreo igualan o sobrepasan el 50% de densidad frente al 91% del segundo, que aporta una cantidad más elevada de palabras de contenido nocional.

En general, sabemos que los resultados del IAT son altos e implican un escaso grado de riqueza léxica ya que el entero corpus proporciona un índice mínimo de 1,52. De todos modos, de nuevo, predomina un rendimiento mejor de los informantes del grupo =2 LE que presentan cifras ligeramente más pequeñas. Las medianas coinciden en los dos muestreos (1,82) y las medias cambian solo de 0,01.

| Descriptivos | Variante | |
|---|---|---|
| | LE =2 | LE >2 |
| Media | 1,84 | 1,83 |
| Mediana | 1,82 | 1,82 |
| Mínimo | 1,52 | 1,56 |
| Máximo | 2,08 | 2,33 |

Tabla 64. Estadísticos descriptivos de IAT para la variable conocimiento de otras LE.

El diagrama naranja muestra que los datos del grupo >2 LE son elevados y conllevan una menor riqueza léxica. El intervalo Q3-Q4 sube a 2,33 alcanzando el valor máximo del corpus, esto es, el peor. La amplitud intercuartil y la desviación intragrupal azules son menores y oscilan entre valores más pequeños apuntando una mayor riqueza en el conjunto LE =2.

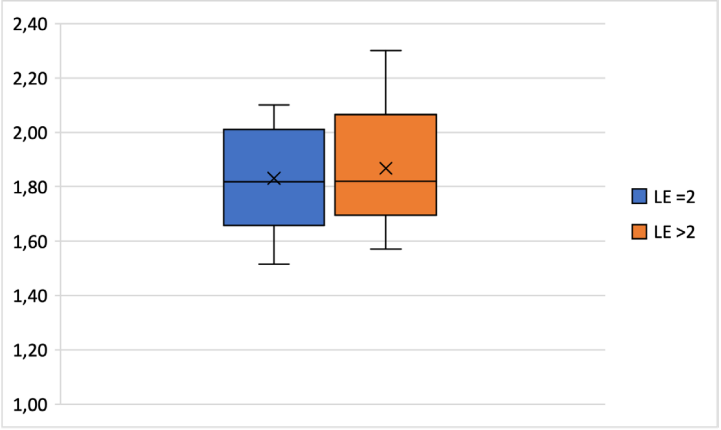


Gráfico 52. Diagrama de cajas de IAT para la variable conocimiento de otras LE.

### 4.1.3 Análisis longitudinal

En este apartado comparamos el material recogido durante las dos suministraciones de la prueba para comprobar si existe una relación aso-

ciativa entre el nivel lingüístico y la cantidad de lexías conocidas y utilizadas. Aplicamos el mismo protocolo empírico midiendo la variedad y la densidad léxica y estudiando los descriptores estadísticos. Partimos de un promedio por texto de 100 *tokens*, 61 *types* en la primera prueba (B2_a) y 63 en la segunda (B2_b).

| ***B2_a*** | | | | | |
|---|---|---|---|---|---|
| **Texto** | **TTR** | **Variación léxica** | **Índice de hápax** | **Densidad léxica** | **IAT** |
| 1 | 0,67 | 67% | 1,37 | 61% | 1,64 |
| 2 | 0,54 | 54% | 1,64 | 50% | 2 |
| 3 | 0,62 | 62% | 1,38 | 50% | 2 |
| 4 | 0,60 | 60% | 1,50 | 54% | 1,85 |
| 5 | 0,59 | 59% | 1,31 | 53% | 1,89 |
| 6 | 0,62 | 62% | 1,41 | 53% | 1,89 |
| 7 | 0,55 | 55% | 1,45 | 48% | 2,08 |
| 8 | 0,65 | 65% | 1,38 | 63% | 1,59 |
| 9 | 0,71 | 71% | 1,25 | 57% | 1,75 |
| 10 | 0,66 | 66% | 1,38 | 56% | 1,79 |
| 11 | 0,61 | 61% | 1,39 | 48% | 2,08 |
| 12 | 0,67 | 67% | 1,31 | 60% | 1,67 |
| 13 | 0,62 | 62% | 1,32 | 53% | 1,89 |
| 14 | 0,64 | 64% | 1,33 | 55% | 1,82 |
| 15 | 0,61 | 61% | 1,45 | 58% | 1,72 |
| 16 | 0,63 | 63% | 1,50 | 55% | 1,82 |
| 17 | 0,54 | 54% | 1,59 | 50% | 2 |
| 18 | 0,66 | 66% | 1,29 | 57% | 1,75 |
| 19 | 0,68 | 68% | 1,28 | 59% | 1,69 |
| 20 | 0,53 | 53% | 1,56 | 49% | 2,04 |
| 21 | 0,65 | 65% | 1,27 | 52% | 1,92 |
| 22 | 0,64 | 64% | 1,28 | 52% | 1,92 |
| 23 | 0,67 | 67% | 1,34 | 52% | 1,92 |
| 24 | 0,57 | 57% | 1,46 | 50% | 2 |

| | | | | | |
|---|---|---|---|---|---|
| 25 | 0,73 | 73% | 1,30 | 56% | 1,79 |
| 26 | 0,66 | 66% | 1,47 | 55% | 1,82 |
| 27 | 0,70 | 70% | 1,37 | 61% | 1,64 |
| 28 | 0,58 | 58% | 1,41 | 46% | 2,17 |
| 29 | 0,56 | 56% | 1,65 | 55% | 1,82 |
| 30 | 0,59 | 59% | 1,40 | 58% | 1,72 |
| 31 | 0,72 | 72% | 1,26 | 59% | 1,69 |
| 32 | 0,67 | 67% | 1,34 | 59% | 1,69 |
| 33 | 0,64 | 64% | 1,33 | 58% | 1,72 |
| 34 | 0,67 | 67% | 1,29 | 52% | 1,92 |
| 35 | 0,61 | 61% | 1,45 | 54% | 1,85 |
| 36 | 0,65 | 65% | 1,41 | 57% | 1,75 |
| 37 | 0,63 | 63% | 1,31 | 49% | 2,04 |
| 38 | 0,66 | 66% | 1,27 | 53% | 1,89 |
| 39 | 0,65 | 65% | 1,30 | 48% | 2,08 |
| 40 | 0,61 | 61% | 1,42 | 53% | 1,89 |
| 41 | 0,63 | 63% | 1,29 | 52% | 1,92 |
| 42 | 0,58 | 58% | 1,53 | 54% | 1,85 |
| 43 | 0,60 | 60% | 1,50 | 57% | 1,75 |
| 44 | 0,69 | 69% | 1,35 | 57% | 1,75 |
| 45 | 0,67 | 67% | 1,34 | 58% | 1,72 |
| 46 | 0,66 | 66% | 1,29 | 55% | 1,82 |
| 47 | 0,67 | 67% | 1,29 | 54% | 1,85 |
| 48 | 0,59 | 59% | 1,48 | 56% | 1,79 |
| 49 | 0,67 | 67% | 1,37 | 57% | 1,75 |
| 50 | 0,61 | 61% | 1,39 | 50% | 2 |
| Promedio | 0,63 | 63% | 1,39 | 54% | 1,85 |

Tabla 65. Índices de RL en la primera suministración de la prueba.

| ***B2_b*** |
|---|

| Texto | TTR | Variación léxica | Índice de hápax | Densidad léxica | IAT |
|---|---|---|---|---|---|
| 1 | 0,57 | 57% | 1,63 | 58% | 1,72 |
| 2 | 0,46 | 46% | 1,59 | 42% | 2,38 |
| 3 | 0,67 | 67% | 1,22 | 52% | 1,92 |
| 4 | 0,69 | 69% | 1,25 | 56% | 1,79 |
| 5 | 0,59 | 59% | 1,37 | 54% | 1,85 |
| 6 | 0,60 | 60% | 1,67 | 60% | 1,67 |
| 7 | 0,59 | 59% | 1,48 | 55% | 1,82 |
| 8 | 0,62 | 62% | 1,35 | 51% | 1,96 |
| 9 | 0,55 | 55% | 1,57 | 54% | 1,85 |
| 10 | 0,63 | 63% | 1,29 | 53% | 1,89 |
| 11 | 0,62 | 62% | 1,44 | 51% | 1,96 |
| 12 | 0,62 | 62% | 1,35 | 52% | 1,92 |
| 13 | 0,61 | 61% | 1,39 | 54% | 1,85 |
| 14 | 0,66 | 66% | 1,40 | 58% | 1,72 |
| 15 | 0,58 | 58% | 1,41 | 57% | 1,75 |
| 16 | 0,61 | 61% | 1,36 | 62% | 1,61 |
| 17 | 0,70 | 70% | 1,27 | 58% | 1,72 |
| 18 | 0,51 | 51% | 1,16 | 53% | 1,89 |
| 19 | 0,62 | 62% | 1,35 | 53% | 1,89 |
| 20 | 0,60 | 60% | 1,36 | 56% | 1,79 |
| 21 | 0,60 | 60% | 1,36 | 51% | 1,96 |
| 22 | 0,67 | 67% | 1,20 | 56% | 1,79 |
| 23 | 0,62 | 62% | 1,35 | 55% | 1,82 |
| 24 | 0,60 | 60% | 1,54 | 53% | 1,89 |
| 25 | 0,60 | 60% | 1,50 | 55% | 1,82 |
| 26 | 0,58 | 58% | 1,45 | 55% | 1,82 |
| 27 | 0,68 | 68% | 1,21 | 55% | 1,82 |
| 28 | 0,62 | 62% | 1,51 | 55% | 1,82 |
| 29 | 0,67 | 67% | 1,26 | 55% | 1,82 |

| | | | | | |
|---|---|---|---|---|---|
| 30 | 0,66 | 66% | 1,22 | 56% | 1,79 |
| 31 | 0,62 | 62% | 1,41 | 50% | 2 |
| 32 | 0,57 | 57% | 1,46 | 47% | 2,13 |
| 33 | 0,60 | 60% | 1,40 | 56% | 1,79 |
| 34 | 0,60 | 60% | 1,43 | 58% | 1,72 |
| 35 | 0,62 | 62% | 1,41 | 58% | 1,72 |
| 36 | 0,56 | 56% | 1,51 | 50% | 2 |
| 37 | 0,68 | 68% | 1,21 | 54% | 1,85 |
| 38 | 0,61 | 61% | 1,49 | 60% | 1,67 |
| 39 | 0,60 | 60% | 1,43 | 53% | 1,89 |
| 40 | 0,67 | 67% | 1,34 | 56% | 1,79 |
| 41 | 0,60 | 60% | 1,30 | 58% | 1,72 |
| 42 | 0,62 | 62% | 1,44 | 55% | 1,82 |
| 43 | 0,58 | 58% | 1,29 | 52% | 1,92 |
| 44 | 0,55 | 55% | 1,57 | 54% | 1,85 |
| 45 | 0,63 | 63% | 1,31 | 57% | 1,75 |
| 46 | 0,60 | 60% | 1,46 | 57% | 1,75 |
| 47 | 0,61 | 61% | 1,36 | 61% | 1,64 |
| 48 | 0,60 | 60% | 1,43 | 57% | 1,75 |
| 49 | 0,60 | 60% | 1,25 | 57% | 1,75 |
| 50 | 0,63 | 63% | 1,34 | 57% | 1,75 |
| Promedio | 0,61 | 61% | 1,39 | 55% | 1,83 |

Tabla 66. Índices de RL en la segunda suministración de la prueba.

### Diversidad léxica

El rendimiento de los alumnos es mejor en la primera prueba: esto extraña ya que, como pensábamos y es esperable, debería darse el resultado opuesto. A comienzo del año académico los valores de TTR son más altos, incluso la media (0,63) y la mediana (0,64), que superan los de la segunda (0,61).

| Descriptivos | Variante | |
|---|---|---|
| | B2_a | B2_b |
| Media | 0,63 | 0,61 |
| Mediana | 0,64 | 0,61 |
| Mínimo | 0,53 | 0,46 |
| Máximo | 0,73 | 0,70 |

Tabla 67. Estadísticos descriptivos de TTR según la fecha de la prueba.

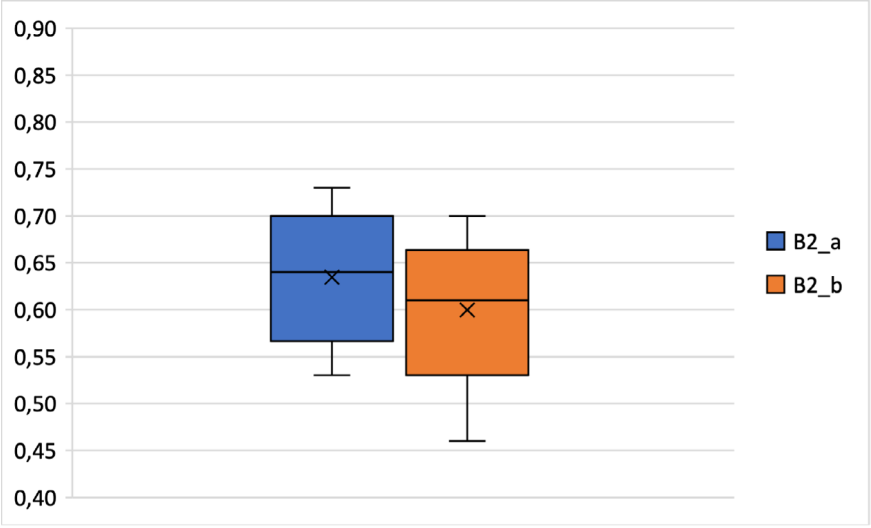


Gráfico 53. Diagrama de cajas de TTR según la fecha de la prueba.

Sobresale la oscilación intragrupal mayor de los datos del corpus B2_b, cuyos valores pasan de 0,70 a 0,46. El corpus B2_a presenta una oscilación de 0,53 a 0,70 que supone una menor desviación interna y aporta valores satisfactorios, mejores con respecto a los conseguidos en la segunda prueba, como enseñan la colocación y el tamaño de los diagramas.

Obviamente, la variación léxica, cuyos datos expresan los valores de la *Type/Token Ratio* en tantos por ciento, presenta el mismo dato: no cabe duda de que los resultados mejores se consiguieron en la primera suministración de la prueba. El valor mínimo baja del 7% confirmando un

empeoramiento de la riqueza léxica, que choca con las expectativas sobre el proceso de adquisición lingüística que supondrían un desarrollo de la competencia según avance el aprendizaje. Los diagramas corroboran el bajón sufrido por los informantes a final del año académico.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B2_a | B2_b |
| Media | 63% | 61% |
| Mediana | 64% | 61% |
| Mínimo | 53% | 46% |
| Máximo | 73% | 70% |

Tabla 68. Estadísticos descriptivos de la variación léxica según la fecha de la prueba.

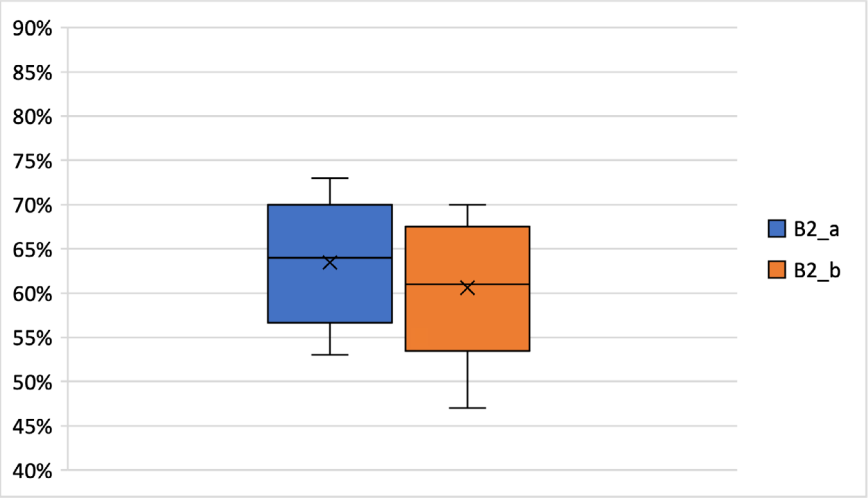


Gráfico 54. Diagrama de cajas de la variación léxica según la fecha de la prueba.

En lo referente a los hápax, los resultados se revelan parecidos: la media matemática es de 1,39 en ambos corpus y la mediana cambia solo de 0,01. En general, su cantidad es bastante buena ya que en B2_a el 90% de los escritos tiene un índice igual o superior a 1,50; en B2_b el porcentaje aumenta hasta el 94%. Los datos son medio-bajos y suponen un grado aceptable de riqueza, pero mejorable porque deberían avecinase más a 1.

| Descriptivos | Variante | |
|---|---|---|
| | B2_a | B2_b |
| Media | 1,39 | 1,39 |
| Mediana | 1,37 | 1,38 |
| Mínimo | 1,25 | 1,16 |
| Máximo | 1,65 | 1,67 |

Tabla 69. Estadísticos descriptivos del índice de hápax según la fecha de la prueba.

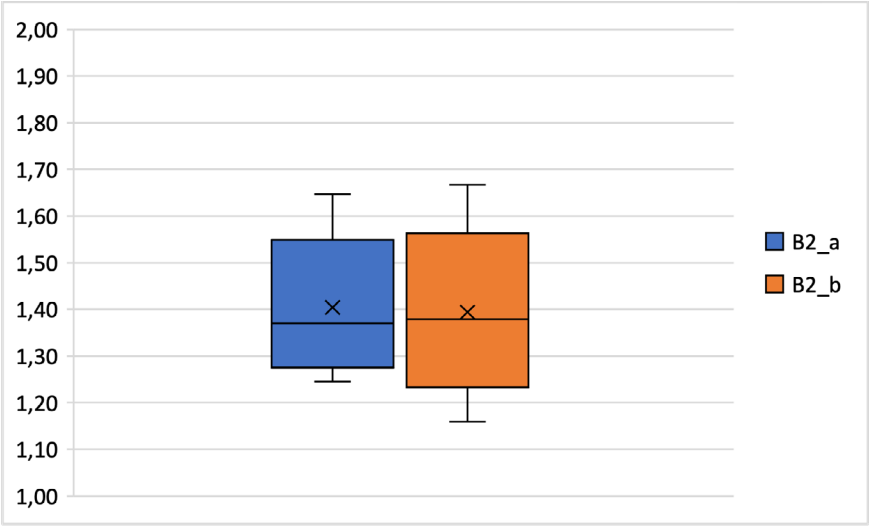


Gráfico 55. Diagrama de cajas del índice de hápax según la fecha de la prueba.

La desviación intragrupal del diagrama naranja manifiesta que en la segunda prueba hay una oscilación mayor, los informantes tocan los dos extremos, sea en positivo sea en negativo, aunque el valor máximo aumenta solo del 0,02. El valor mínimo favorece este mismo muestreo, como se deprende de la extensión del bigote inferior, que baja hasta 1,16 incrementando el rendimiento del 0,09.

### Densidad léxica

Los valores medio de la densidad léxica son satisfactorios y están repartidos uniformemente en los corpus, la mayoría de los textos tiene

una cantidad de palabras semánticas que supera el 50%. No obstante, los extremos ponen de manifiesto, de nuevo, un mejor rendimiento en la primera prueba: el valor máximo se diferencia de un punto y el mínimo de cuatro, hecho que extraña si tenemos en cuenta que los alumnos asistieron a un curso entero.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B2_a | B2_b |
| Media | 54% | 55% |
| Mediana | 55% | 55% |
| Mínimo | 46% | 42% |
| Máximo | 63% | 62% |

Tabla 70. Estadísticos descriptivos de la densidad léxica según la fecha de la prueba.

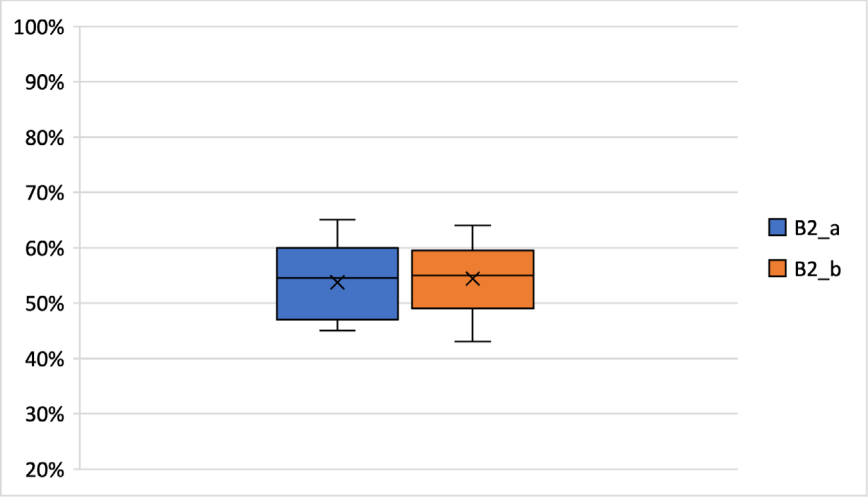


Gráfico 56. Diagrama de cajas de la densidad léxica según la fecha de la prueba.

La desviación inter e intragrupal es parecida, la caja azul solo presenta más oscilación en los intervalos Q1-Q2. En efecto, el 88% de los textos del corpus B2_a y el 96% del B2_b alcanzan o superan el 50% de densidad léxica. Si desglosamos los índices obtenidos por cada relato percatamos cifras semejantes, el uso de las unidades semánticas es bueno.

Terminamos el análisis longitudinal con los resultados del IAT que confirman la tendencia detectada antes. Si bien media y mediana difieran solo de 0,02 el rendimiento es superior en la primera encuesta, cuyos datos comprenden el valor mínimo (1,59) y el máximo (2,17). En la segunda suministración las cifras son más altas, por lo que los resultados empeoran. En particular, hay un aumento de +0,20 que conlleva un índice más alto y denota una reducción de la riqueza léxica a finales del curso.

| **Descriptivos** | **Variante** | |
|---|---|---|
| | B2_a | B2_b |
| Media | 1,85 | 1,83 |
| Mediana | 1,84 | 1,82 |
| Mínimo | 1,59 | 1,61 |
| Máximo | 2,17 | 2,37 |

Tabla 71. Estadísticos descriptivos de IAT según la fecha de la prueba.

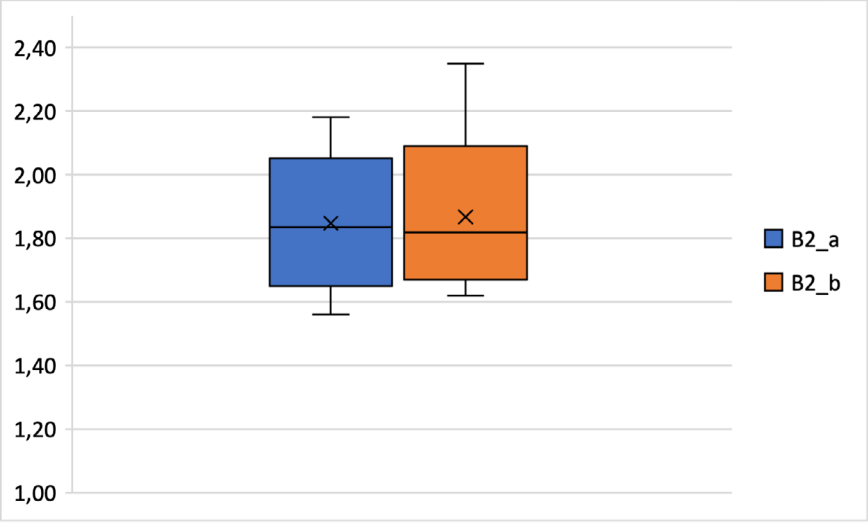


Gráfico 57. Diagrama de cajas de IAT según la fecha de la prueba.

Sobresale el desnivel entre los puntos extremos tocados por los bigotes ya que si examinamos el tamaño de las cajas se observa cierta simetría de los resultados. A más y mejor, al analizar individualmente el índice de cada texto constatamos que los valores del corpus B2_b son un poco más

bajos. Se confirma que la riqueza léxica no es alta desde el momento en el que las cifras se avecinan más a 2 que a 1 y ningún texto presenta un índice inferior a 1,50 (el mínimo es 1,59).

### 4.1.4 Análisis comparativo

Finalmente, comparamos nuestros resultados con los de otras investigaciones sobre la riqueza léxica en aprendientes no nativos de español con el objetivo de destacar correspondencias y diferencias en los índices de TTR e IAT. No tenemos la pretensión de ser exhaustivos, pues, son varios los factores que influyen en el aprendizaje y dependen mucho del entorno en el que se desarrolla y, asimismo, tratamos textos que abarcan varias temáticas. De todos modos, planteamos este cotejo ya que consideramos los informantes de cada estudio modelos estándar del estudiante de nivel intermedio de ELE:

- Cuba Vega y Cuba (2004) miden la riqueza léxica de aprendientes brasileños producidas en textos con extensión fija de 100 *tokens*.
- Berton (2014) trabaja con alumnos suecos que estudian español como L3. Recoge 180 relatos de diferente tamaño, de los cuales tenemos en cuenta los que más se acercan a las características de nuestro corpus, es decir los que llegan a un máximo de 117 *tokens*.
- Wang (2016) examina la producción de discentes sinohablantes en escritos de 100 *tokens*.
- Basso (2017) estudia un corpus de textos escritos por informantes italófonos de diferentes longitudes del que utilizamos los que no exceden los 100 *tokens*.

| **Investigación** | **País** | **Extensión textos** | **Nivel educativo** |
|---|---|---|---|
| Cuba Vega y Cuba (2004) | Brasil | 100 | Universidad |
| Berton (2014) | Suecia | 117 | Bachillerato |
| Wang (2016) | China | 100 | Universidad |

| | | | |
|---|---|---|---|
| Basso (2017) | Italia | 100 | Universidad |
| Nalesso (2019a) | Italia | 100 | Universidad |

Tabla 72. Datos generales de las investigaciones cotejadas.

Si excluimos a Berton (2014), los trabajos se centran en la riqueza léxica de aprendientes universitarios de nivel intermedio y analizan redacciones constituidas por el mismo número de *tokens*, por lo que la comparación resulta equilibrada y coherente.

### Diversidad léxica

La tabla contiene los promedios de la *Type/Token Ratio* que representan el punto de partida para nuestra comparación. La media matemática contabilizada es de 0,63, la cual revela que la diversidad léxica es buena para todos, ya que supera el 50% del texto: la presencia de *types* prevalece sobre la totalidad de los *tokens*.

| **Investigación** | **TTR** |
|---|---|
| Cuba Vega y Cuba (2004) | 0,64 |
| Berton (2014) | 0,70 |
| Wang (2016) | 0,60 |
| Basso (2017) | 0,60 |
| Nalesso (2019a) | 0,63 |
| Promedio | 0,63 |

Tabla 73. Promedios de TTR en las investigaciones cotejadas.

Berton (2014) presenta el mejor índice con una media de 0,70. Siguen los informantes de Cuba Vega y Cuba (2004) con 0,64 y los nuestros con 0,63. Las últimas posiciones están ocupadas por los encuestados de Wang (2016) y los de Basso (2017) que presentan un promedio de 0,60.

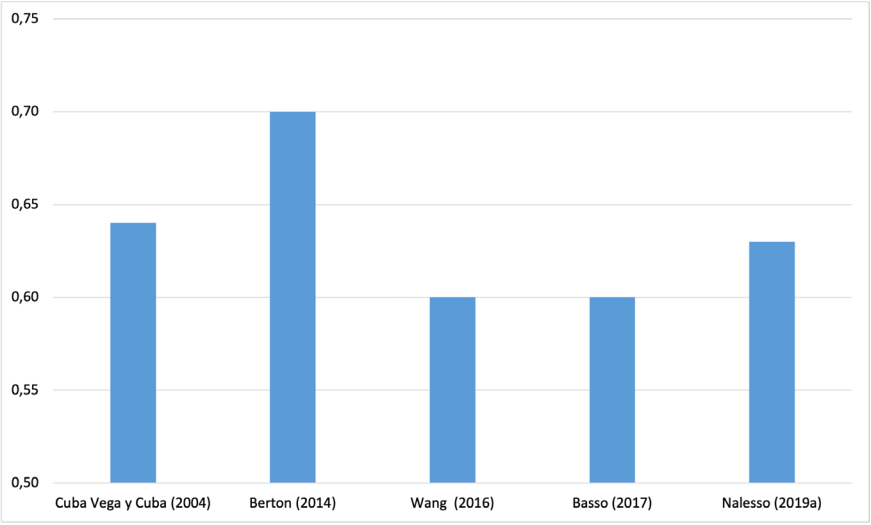


Gráfico 58. Promedios de TTR en las investigaciones cotejadas.

Sorprenden los datos de Berton (2014), que trabaja con alumnos de bachillerato que estudian español como L3 y que superan a los otros –estudiantes universitarios de español L2– de los que se esperaría un rendimiento mejor. Asimismo, extrañan los escasos datos expuestos en Wang (2016), cuyos encuestados arrojan uno de los promedios más pobres si consideramos que en el análisis de la disponibilidad léxica los sinohablantes obtienen resultados sobresalientes (Hidalgo 2019).

### Densidad léxica

El contraste entre el IAT, donde el número de estudios comparables se reduce a cuatro,[42] pone de manifiesto que la densidad de vocabulario tampoco consigue resultados muy diferentes. La media general es de 1,95, cifra que se aleja significativamente del valor ideal de esta medida (recuérdese, 1) e implica un bajo grado de riqueza léxica.

[42] En el trabajo dedicado a los informantes suecos no se estudia ni el IAT ni otro índice que computa la densidad léxica porque según Berton (2014: 28) «[…] no parece[n] útil[es] para este estudio, ya que en las producciones de aprendices principiantes e intermedios como los informantes del presente estudio el uso de palabras funcionales, como por ejemplo los marcadores del discurso, es muy reducido».

| Investigación | IAT |
|---|---|
| Cuba Vega y Cuba (2004) | 1,81 |
| Berton (2014) | – |
| Wang (2016) | 1,80 |
| Basso (2017) | 2,35 |
| Nalesso (2019a) | 1,84 |
| Promedio | 1,95 |

Tabla 74. Promedios de IAT en las investigaciones cotejadas.

Los sinohablantes de Wang (2016) aportan la media mejor (1,80). A continuación, Cuba Vega y Cuba (2004) presentan un IAT igualmente alto de 1,81 seguido por el nuestro que es de 1,84. Por último, Basso (2017) registra el peor índice de 2,35 lo cual sería +0,55 al compararlo con el resultado mejor.

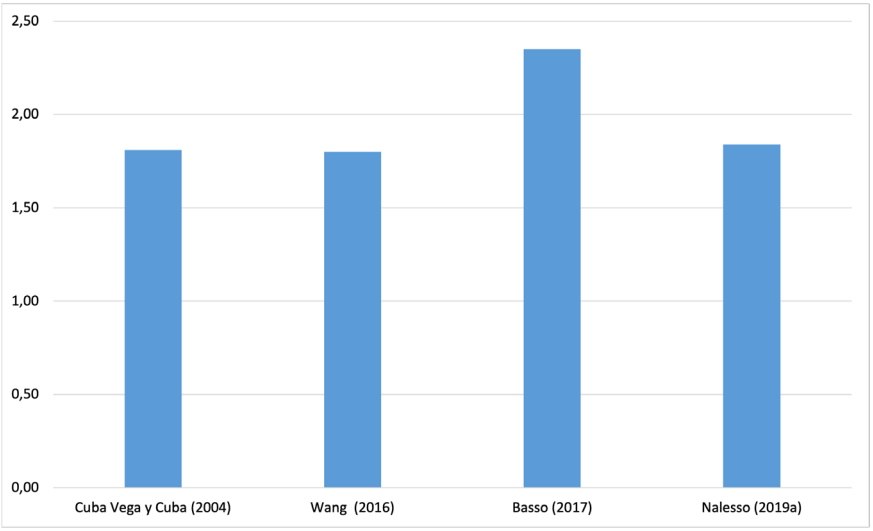


Gráfico 59. Promedios de IAT en las investigaciones cotejadas.

Al excluir Basso (2017), los demás índices varían poco: esto sugiere que los informantes tienen desarrollada la misma capacidad de utilizar las palabras nocionales en la lengua objeto. Sin embargo, se confirma la teoría de Johansson (2008) según la cual una buena variedad léxica no im-

plica una buena densidad. Por cierto, pese a que los dos índices cotejados hayan aportado información diferente, parece que todos los estudiantes han entregado composiciones aceptables en función de su nivel de ELE y saben desempeñar una tarea de producción escrita no extensa y simple.

## 4.2 Aproximación a un estudio cualitativo

Los índices que acabamos de analizar además de proporcionar datos numerales aportan información adicional sobre la capacidad expresiva y la competencia léxica. Sin embargo, un estudio cualitativo implicaría nuevas consideraciones sobre los textos de los informantes desde una óptica que deje de un lado las cifras. Las metodologías cualitativas pueden ser de gran aporte en la investigación sobre riqueza léxica porque permiten profundizar el análisis del material obtenido. Por eso, proponemos un estudio cualitativo preliminar de los textos a través de un análisis morfológico de las unidades léxicas.[43] Para ello, clasificamos las palabras del corpus según categorías gramaticales para indagar, antes, las listas formadas por los *types* aportados y, a continuación, solo por *types* léxicos extraídos del listado de la frecuencia general (tomamos como medida de corte los veinticinco ítems más frecuentes). Además, se anlizan las palabras clave extraídas mediante la herramienta *Keywords* de *Sketch Engine.*

Recordamos que el paso previo al análisis ha sido la lematización de los *tokens* de manera que las formas flexionadas pudiesen ser contadas como un único lema, por lo que en las tablas se encontrarán unidades no marcadas.

### 4.2.1 Listas de frecuencia

En el corpus analizado, resultante de los 100 textos entregados a principio del año académico, el número total de *tokens* es 10.000 y el total de *types*, incluidas las formas flexionadas de un mismo vocablo, es de 1.072. Estos se dividen con una notable disparidad entre palabras funcionales del discurso y de contenido nocional, reflejando los valores de la densidad léxica ya discutidos. La repartición de las unidades léxicas se ilustra en el gráfico 60, cuyos sectores destacan claramente dicho desnivel.

[43] Tratándose de un estudio piloto en este ámbito, ya que no tenemos constancia de otros trabajos de esta naturaleza, esperamos que tanto la metodología propuesta como las modestas conclusiones que sacamos puedan constituir un punto de partida para ulteriores investigaciones.

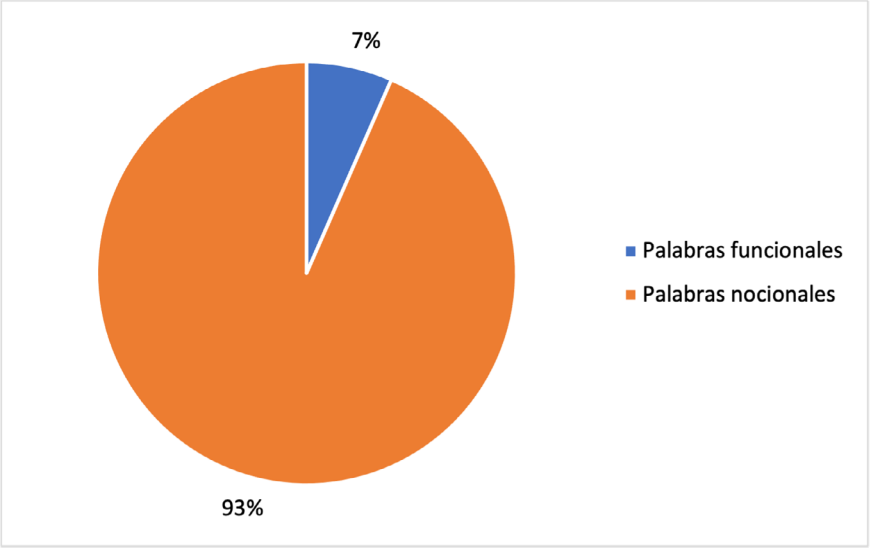


Gráfico 60. Distribución de las palabras en el corpus general de RL.

Los ítems gramaticales se colocan en las primeras posiciones de la lista de las palabras más frecuentes. Artículos, conjunciones, preposiciones y pronombres son las unidades más utilizadas (11 sobre un total de 25), como era esperable, si excluimos el verbo *ser* que consideramos como nocional pese a ser auxiliar y copulativo. El verbo *ir* se ubica en la posición 10; desde la posición 14 encontramos palabras temáticas entre las que sobresale la presencia de otros verbos (*hacer*, *gustar*, *estar*, *tener*, *visitar*), por encima de la aparición de otras categorías.

| Rango | Vocablo | Frecuencia |
|---|---|---|
| 1 | el | 870 |
| 2 | y | 436 |
| 3 | de | 415 |
| 4 | ser | 305 |
| 5 | en | 289 |
| 6 | que | 287 |
| 7 | un | 273 |
| 8 | a | 231 |
| 9 | mi | 185 |

| | | |
|---|---|---|
| 10 | ir | 153 |
| 11 | me | 152 |
| 12 | con | 136 |
| 13 | por | 126 |
| 14 | ciudad | 115 |
| 15 | mucho | 109 |
| 16 | viaje | 107 |
| 17 | hacer | 106 |
| 18 | todo | 95 |
| 19 | más | 95 |
| 20 | día | 94 |
| 21 | gustar | 93 |
| 22 | muy | 92 |
| 23 | estar | 83 |
| 24 | tener | 81 |
| 25 | visitar | 80 |

Tabla 75. Palabras más frecuentes del corpus general de RL.

Al excluir del análisis las unidades funcionales, la lista de las palabras semánticas más frecuentes presenta la siguiente distribución: sustantivos, verbos, adjetivos y adverbios.

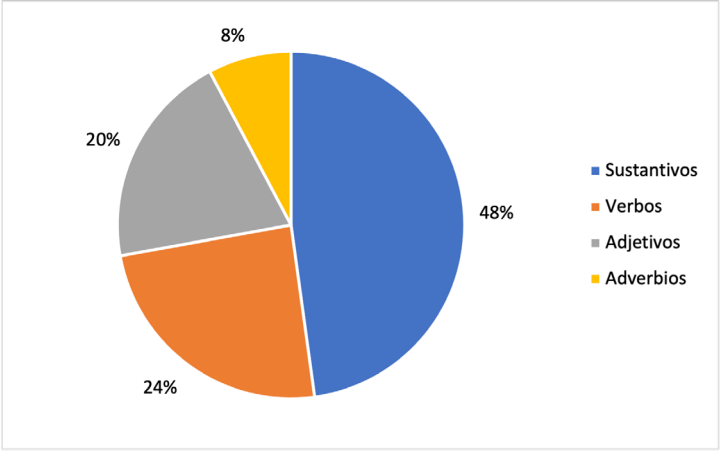


Gráfico 61. Distribución de las palabras nocionales en el corpus general de RL.

La tabla 76 contiene los vocablos que constituyen estos porcentajes, esto es, los más frecuentes sin distinción de categoría.

| Rango | Vocablo | Frecuencia |
|---|---|---|
| 1 | ser | 305 |
| 2 | ir | 153 |
| 3 | ciudad | 115 |
| 4 | mucho | 109 |
| 5 | viaje | 107 |
| 6 | hacer | 106 |
| 7 | todo | 95 |
| 8 | día | 94 |
| 9 | gustar | 93 |
| 10 | muy | 92 |
| 11 | estar | 83 |
| 12 | tener | 81 |
| 13 | visitar | 80 |
| 14 | año | 78 |
| 15 | amigo | 62 |
| 16 | allí | 58 |
| 17 | también | 53 |
| 18 | ver | 47 |
| 19 | pasado | 42 |
| 20 | como | 41 |
| 21 | semana | 39 |
| 22 | llegar | 37 |
| 23 | haber | 37 |
| 24 | lugar | 35 |
| 25 | familia | 32 |

Tabla 76. Palabras nocionales más frecuentes del corpus general de RL.

La mayor parte de los lemas está estrechamente vinculada a la temática propuesta para la redacción del texto (la descripción de un viaje), con lo cual ya a raíz de esta selección se comprendería el asunto de los relatos. En efecto, contamos con la presencia de verbos (*ir*, *visitar*, *ver*, *llegar*) y sustantivos (*ciudad*, *viaje*, *lugar*) emblemáticos.

Dentro de este conjunto, los sustantivos componen casi la mitad de las unidades léxicas más frecuentes, el 48%. Los lexemas que tienen el mayor número de apariciones son *ciudad* con 115 y *viaje* con 107, seguidas por *día*, *año* y *amigo* que superan las 50 ocurrencias.

| **Rango** | **Vocablo** | **Frecuencia** |
|---|---|---|
| 1 | ciudad | 115 |
| 2 | viaje | 107 |
| 3 | día | 95 |
| 4 | año | 78 |
| 5 | amigo | 62 |
| 6 | semana | 39 |
| 7 | lugar | 35 |
| 8 | vez | 33 |
| 9 | familia | 32 |
| 10 | Londres | 29 |
| 11 | mañana | 28 |
| 12 | gente | 25 |
| 13 | avión | 23 |
| 14 | casa | 23 |
| 15 | isla | 21 |
| 16 | novio | 20 |
| 17 | museo | 20 |
| 18 | noche | 20 |
| 19 | playa | 20 |
| 20 | compañero | 19 |

| | | |
|---|---|---|
| 21 | cosa | 19 |
| 22 | escuela | 18 |
| 23 | hermano | 18 |
| 24 | aeropuerto | 17 |
| 25 | experiencia | 16 |

Tabla 77. Sustantivos más frecuentes del corpus general de RL.

Los verbos cubren el 24% de las palabras más frecuentes. Dentro de esta cifra, excluyendo el auxiliar *ser* (292 apariciones), encontramos una amplia gama de verbos entre los que *ir* es el más utilizado (166 apariciones) antes de *hacer*, *gustar*, *tener*, *estar*, *visitar*, *pasar* que presentan más de 50 ocurrencias cada uno. Otros verbos, si bien considerados significativos en este ámbito (por ejemplo, *llegar*, *salir*, *viajar*) se manifiestan en un número reducido de ocurrencias. Apreciamos, asimismo, la actualización de algunas unidades pluriverbales con el verbo *tomar* como *tomar el avión*, *tomar el tren*, *tomar el autobús*, *tomar el sol* (16 ocurrencias sobre un total de 26) y apariciones de la perífrasis verbal *ir + a + infinitivo*, como *ir a bailar*, *ir a nadar*, *ir a visitar*.

| **Rango** | **Vocablo** | **Frecuencia** |
|---|---|---|
| 1 | ser | 292 |
| 2 | ir | 166 |
| 3 | hacer | 110 |
| 4 | gustar | 93 |
| 5 | tener | 83 |
| 6 | estar | 83 |
| 7 | visitar | 82 |
| 8 | pasar | 60 |
| 9 | ver | 47 |
| 10 | llegar | 38 |
| 11 | conocer | 29 |
| 12 | poder | 28 |
| 13 | tomar | 26 |

| | | |
|---|---|---|
| 14 | quedar | 24 |
| 15 | vivir | 21 |
| 16 | comer | 20 |
| 17 | encantar | 20 |
| 18 | encontrar | 17 |
| 19 | salir | 15 |
| 20 | divertir | 14 |
| 21 | viajar | 14 |
| 22 | haber | 14 |
| 23 | decidir | 12 |
| 24 | dar | 12 |
| 25 | hablar | 11 |

Tabla 78. Verbos más frecuentes del corpus general de RL.

Por su parte, los adjetivos ocupan el 20% de la lista de las palabras temáticas más frecuentes. *Primero* es la unidad más repetida, sobre todo en la expresión *la primera vez* (repetida 19 veces en 34 ocurrencias totales del ítem). Otros lexemas muy utilizados son *maravilloso*, *grande*, *pequeño*, que superan las 20 apariciones. Como muestra la tabla, y como es lógico teniendo en cuenta la tipología de texto, la mayoría de los adjetivos son calificativos: expresan cualidades o propiedades del objeto designado, incluso si alargamos el análisis al entero listado.

| **Rango** | **Vocablo** | **Frecuencia** |
|---|---|---|
| 1 | primero | 34 |
| 2 | maravilloso | 24 |
| 3 | grande | 22 |
| 4 | pequeño | 20 |
| 5 | típico | 19 |
| 6 | mejor | 18 |
| 7 | diferente | 17 |
| 8 | nuevo | 16 |

| | | |
|---|---|---|
| 9 | estupendo | 15 |
| 10 | último | 15 |
| 11 | bueno | 14 |
| 12 | español | 13 |
| 13 | junto | 12 |
| 14 | famoso | 11 |
| 15 | siguiente | 11 |
| 16 | solo | 10 |
| 17 | importante | 9 |
| 18 | lleno | 9 |
| 19 | particular | 9 |
| 20 | amable | 7 |
| 21 | fantástico | 6 |
| 22 | largo | 6 |
| 23 | verde | 5 |
| 24 | único | 5 |
| 25 | vario | 5 |

Tabla 79. Adjetivos más frecuentes del corpus general de RL.

Finalmente, los adverbios son los que menos aparecen entre las palabras semánticas (8%). Los de mayor uso son *más*, *muy* y *mucho* (incluso en la forma superlativa *muchísimo*) que tienen 95, 92 y 79 repeticiones. Los demás adverbios incluidos en la tabla siguiente presentan un número bastante inferior de ocurrencias, hasta bajar a 4 (*ahora*).

| **Rango** | **Vocablo** | **Frecuencia** |
|---|---|---|
| 1 | más | 95 |
| 2 | muy | 92 |
| 3 | mucho | 79 |
| 4 | allí | 58 |
| 5 | también | 53 |

| | | |
|---|---|---|
| 6 | no | 47 |
| 7 | después | 28 |
| 8 | cerca | 18 |
| 9 | siempre | 18 |
| 10 | nunca | 15 |
| 11 | así | 14 |
| 12 | además | 12 |
| 13 | bien | 10 |
| 14 | ya | 8 |
| 15 | luego | 7 |
| 16 | casi | 6 |
| 17 | solo | 6 |
| 18 | aquí | 6 |
| 19 | antes | 6 |
| 20 | poco | 5 |
| 21 | bastante | 5 |
| 22 | tanto | 5 |
| 23 | lejos | 5 |
| 24 | entonces | 5 |
| 25 | ahora | 4 |

Tabla 80. Adverbios más frecuentes del corpus general de RL.

Al examinar la lista entera de esta clase, los adverbios en -mente son los siguientes, en orden alfabético: *afortunadamente*, *contrariamente*, *desafortunadamente*, *estupendamente*, *exactamente*, *exageradamente*, *inmediatamente*, *magníficamente*, *mediamente*, *normalmente*, *obviamente*, *particularmente*, *perfectamente*, *prácticamente*, *precisamente*, *seguramente*, *solamente*, *temporáneamente*, *totalmente*, *tranquilamente*, *únicamente*, *verdaderamente*. Revelan una frecuencia que va de un máximo de 3 a un mínimo de 1, esto es, se trata en gran mayoría de palabras hápax.

### 4.2.2 Palabras clave

En esta última sección observamos las palabras clave del corpus a partir de la tabla que encierra las unidades extraídas mediante *Sketch Engine*. Este análisis contribuye a definir cuál es el tema principal de los textos analizando sus lexías y comparándolas con las que forman parte del corpus tomado como referencia, el *Spanish Web corpus 2018*.[44]

| **Rango** | **Vocablo** | **Frecuencia Nalesso (2019a)** | **Frecuencia *Spanish Web corpus 2018*** |
|---|---|---|---|
| 1 | vacación | 12 | 2.403 |
| 2 | estupendo | 15 | 113.919 |
| 3 | viaje | 107 | 1.198.008 |
| 4 | Londres | 29 | 317.393 |
| 5 | maravilloso | 24 | 330.841 |
| 6 | visitar | 82 | 1.264.694 |
| 7 | pasear | 10 | 175.041 |
| 8 | avión | 23 | 433.906 |
| 9 | monumento | 12 | 227.479 |
| 10 | típico | 19 | 378.160 |
| 11 | novio | 20 | 432.929 |
| 12 | gustar | 93 | 241.1475 |
| 13 | coger | 10 | 258.519 |
| 14 | aeropuerto | 17 | 458.292 |
| 15 | allí | 58 | 1.613.761 |
| 16 | encantar | 20 | 587.578 |
| 17 | divertir | 14 | 455.357 |
| 18 | paisaje | 10 | 342.996 |

[44] *Sketch Engine* ofrece el acceso a este recurso que hace parte de la familia *TenTen corpus*, un conjunto de corpus de más de 10.000 millones de palabras creado a partir de textos recogidos en línea. En el caso del español, contiene 17.500 millones de entradas.

| | | | |
|---|---|---|---|
| 19 | playa | 20 | 725.652 |
| 20 | isla | 21 | 773.341 |
| 21 | museo | 20 | 755.141 |
| 22 | tren | 10 | 374.200 |
| 23 | bar | 10 | 397.360 |
| 24 | verano | 16 | 657.445 |
| 25 | amigo | 62 | 2.631.598 |

Tabla 81. Palabras clave del corpus general de RL.

Sobresale la presencia masiva de sustantivos, seguidos por verbos y adjetivos, que concurren a establecer la temática abordada en los textos: se deduce que se trata de viajes, en particular gracias a las unidades *vacación*, *viaje*, *visitar*, *avión*, *monumento*, *típico*, *aeropuerto*, *playa*, *isla*, *museo*, *tren*. Percatamos, además, la presencia de lexemas que implican acciones, lugares, medios de transporte y sensaciones ligadas a la experiencia del viaje. Se señala que, aunque ciertos vocablos alcancen un índice de frecuencia bajo, se encuentran en posiciones altas en esta lista porque son ítems léxicos emblemáticos del tema.

A estas alturas, aprovechando del instrumento *Word Sketch* del programa, analizamos los distintos usos de algunos de estos vocablos mediante las siguientes figuras que ilustran cómo se combinan con otras unidades. Elegimos dentro del listado de las palabras clave, un ítem para cada una de las categorías morfológicas: un sustantivo, un verbo, un adjetivo y un adverbio.

El sustantivo *vacación* encabeza la lista, pues, es la primera palabra clave del corpus, aunque presenta solo doce ocurrencias totales. De la imagen se ve que se combina con: preposiciones de lugar, tiempo, modo y compañía; el posesivo *nuestro*; nombres comunes con los que forma unidades multipalabra; verbos que lo tienen como objeto; adjetivos que funcionan como atributos.


Figura 2. Uso del vocablo *vacación* en el corpus general de RL.

El primer verbo que aparece en la lista es *visitar*. Se utiliza en los textos junto a: sustantivos (nombres comunes y propios) con la función de objeto; adverbios demostrativos, de modo y temporales; la preposición de modo e instrumento *con*; la preposición *a* que desempeña la función de acusativo personal.

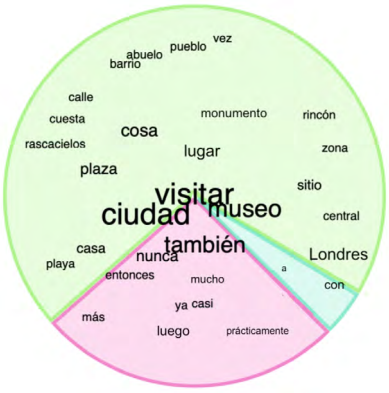


Figura 3. Uso del vocablo *visitar* en el corpus general de RL.

*Estupendo* es el adjetivo que se coloca primero en la lista y se emplea con: verbos copulativos (*ser* y *parecer*), con los que forma predicados nominales, teniendo la función de atributo; nombres comunes para los cuales es un adjetivo calificativo.


Figura 4. Uso del vocablo *estupendo* en el corpus general de RL.

Para terminar, observamos en la lista de las palabras clave el adverbio *allí* que se encuentra junto a: verbos, cuyo significado completa ejerciendo el papel de complemento; otros adverbios temporales y de modo.


Figura 5. Uso del vocablo *allí* en el corpus general de RL.

## Conclusiones

Este trabajo ha estudiado la competencia léxica en español de un grupo de estudiantes italianos por medio de un experimento transversal y, por primera vez, longitudinal finalizado a abrir el camino a aspectos no tratados con profundidad, o no tratados en absoluto, en otros proyectos enmarcados en el ámbito de las investigaciones sobre el aprendizaje del léxico en ELE. No es el primero que aborda este tema, pero se llevó a cabo con un enfoque original basado en la correlación de los datos procedentes de los análisis de la disponibilidad léxica, gracias a los cuales realizamos el estudio cuantitativo, cualitativo y comparativo del léxico disponible (plano paradigmático del léxico), y de los análisis de la riqueza léxica, que permiten examinar el vocabulario activo de los encuestados mediante el estudio de un corpus de textos escritos (plano sintagmático del léxico).

El volumen del corpus recopilado consiente extraer resultados concluyentes en cuanto a la competencia léxica del conjunto de aprendientes contemplado, tanto en el análisis y la comparación intramuestral como del contraste intermuestral. Todo esto posibilitó el alcance de los objetivos fijados: el análisis del léxico disponible y activo de la muestra y su evolución en el tiempo; la evaluación de la incidencia de los factores sociolingüísticos sexo, nivel de ELE y conocimiento de otras lenguas extranjeras en su bagaje léxico; el cotejo de nuestros datos con los de otros proyectos dedicados a estudiantes no nativos de español.

Para ello, el capítulo 1 se dedica al marco teórico y al estado de la cuestión, donde se presenta una revisión de los orígenes de la disponibilidad y de la riqueza léxica, hasta llegar al desarrollo de las propuestas más recientes en el campo de los estudios sobre ELE.

Después de esta sección introductoria, el capítulo 2 desarrolla la metodología de la investigación realizada a través del test de medición de

la competencia léxica suministrado a cien alumnos de la Universidad de Padua, aprendientes de ELE de nivel B1 y B2, organizando la recogida para comparar su conocimiento en cada uno de los dos niveles y en dos etapas distintas del aprendizaje. Se describen la selección de la muestra y su distribución conforme a las tres variables sociolingüísticas consideradas; las técnicas de elicitación y los procesos de recogida de los datos; los criterios de edición y de tratamiento del material; los tipos de análisis efectuados.

Destacamos que tras la recolección de los datos y antes de su procesamiento informático se editó el corpus para que resultase homogéneo y consintiese análisis fiables. Durante esta etapa no se detectaron errores graves, se trata en la mayoría de los casos de faltas ortográficas que no perjudican la competencia de los informantes y, sobre todo, de errores de colocación falsa probablemente causados por una distracción, ya que en general las palabras eran correctas. Asimismo, hay errores intralinguales e interlinguales (extranjerismos, préstamos, falsos amigos, palabras inventadas) debidos a la interferencia de la lengua materna y de otras lenguas conocidas, en particular del inglés y del francés al ser las dos LE más estudiadas junto al español en este entorno académico. Según el criterio etiológico (Torijano 2004) y la lingüística contrastivo-perceptiva (Di Gesù 2016) la organización del lexicón mental de un aprendiente extranjero favorecería la recuperación de material lingüístico de su idioma nativo o de cualquier otro idioma conocido como estrategia de aprendizaje. Todo esto refuta una de nuestras hipótesis según la cual la afinidad entre italiano y español llevaría a una importante incidencia de la LM en las actualizaciones por la poca distancia léxica que caracteriza los lexemas de estas lenguas (Calvi 1995: 86-90) puesto que en realidad no hay una gran aparición de italianismos. Al contrario, se corrobora otra hipótesis, ya que se localiza una cantidad apreciable de palabras del registro coloquial y de la variedad hispanoamericana. De la misma manera, en lo que atañe a las categorías morfosintácticas de las respuestas, comprobamos una presencia abrumadora de sustantivos y de pocas unidades multipalabras.

La parte empírica del trabajo se abre con el capítulo 3 que desarrolla el análisis de la disponibilidad léxica. En la primera sección planteamos la evaluación cuantitativa del léxico disponible partiendo de un total de 12.579 palabras y 1.490 vocablos. En este primer análisis, repartimos los dieciséis centros de interés en cuatro franjas en función de su mayor o menor productividad: cuatro campos superan las 1.000 unidades manifestándose como los más productivos, otros sobrepasan la media de 786 palabras, un conjunto no llega a este promedio y, por último, un grupo

no alcanza los 400 ítems quizá debido al poco interés suscitado por las temáticas y la poca atención que les otorgan los programas curriculares y los materiales de ELE. La cantidad de vocablos demuestra que no existe una relación directa con la productividad porque se manifiesta una distribución distinta de los campos más y menos ricos.

Conforme a los índices de cohesión y de densidad léxica algunos de los estímulos semánticos se presentan más cerrados (CI01, "partes del cuerpo"; CI03, "partes de la casa (sin muebles)"; CI04, "los muebles de la casa"; CI12, "medios de transporte") y otros más difusos (CI11, "el campo"; CI13, "trabajos del campo y del jardín"; CI16, "profesiones y oficios"). A este propósito, se confirma que ciertos centros consiguen un alto grado de cohesión porque posibilitan la activación de limitadas asociaciones mentales, mientras que un bajo índice conlleva una gran variedad temática que permite activar más conexiones (Echeverría *et al.* 1987: 67). Es más, puede que una mayor concreción implique un aprendizaje sistematizado a partir del cual los estudiantes no amplían sus conocimientos por encima de los propuestos en los cursos y, en el caso opuesto, es posible que cada sujeto conozca ítems distintos de los compañeros, ya que se trata de argumentos aprendidos o profundizados por su cuenta.

Antes de discutir la influencia de los condicionantes sociolingüísticos, recordamos que la población investigada se distribuía como sigue:

- variable sexo: 87 mujeres y 13 hombres,
- variable nivel de ELE: 50 aprendices de nivel B1 y 50 aprendices de nivel B2,
- variable conocimiento de otras LE: 19 conocen 2 LE y 81 conocen más de 2 LE.

Los datos sugieren que el factor sexo no influye significativamente en la cantidad de actualizaciones del grupo, las mujeres arrojan el 9,61% más de palabras con respecto al conjunto masculino, pero los hombres prevalecen en la producción de vocablos sobrepasando las compañeras del 66,35%. El nivel de cohesión revela que las respuestas de las mujeres son más dispersas con respecto a los varones que aportan un conjunto más cerrado.

En lo referente al nivel de ELE, como esperábamos, registramos un aumento en casi todos los centros del grupo avanzado que produce el 7,51% más de palabras y el 14,05% de vocablos. Sin embargo, este incremento no es tan significativo como para sugerir un desarrollo sobresaliente de los conocimientos del alumnado, esto es, el desnivel no es notable. Al examinar la concreción interna de cada variante, los estudiantes intermedios

alcanzan +8,47% en el IC y un +4,25% en la densidad léxica, con lo cual su léxico resulta más compacto.

Por último, el conocimiento de otras LE se demuestra influyente en la capacidad de activar más palabras para los informantes que conocen más lenguas (+10,54%), al contrario, los vocablos favorecen los que conocen dos (+50,84%). Asimismo, estos consiguen un más alto grado de cohesión que supone una mayor asociación conceptual del lexicón mental con respecto a los compañeros.

El análisis longitudinal revela que la competencia léxica ha mejorado tras la asistencia a un curso de ELE, no de manera sobresaliente: se calculan una mayor producción y riqueza a final del año. Los datos de la secunda suministración de la prueba indican un aumento del 30,29% en el total de palabras en todos los estímulos temáticos (pasan de 6.517 a 8.491) y del 7,50% en el total de vocablos (pasan de 1.161 a 1.244). En suma, parece que la acción docente ha tenido su eficacia, incluso porque los índices de cohesión y densidad revelan un conjunto de palabras más cerrado del 23,75%.

El siguiente apartado coteja los resultados de nuestros informantes con otros estudiantes de ELE para demostrar que la competencia léxica de nuestra muestra es mayor debido a la afinidad lingüística entre italiano y español. Contrariamente a lo esperado, esta hipótesis no se confirma: las cantidades de palabras y vocablos de los italianos son menores. En lo referente a la producción de palabras, los grupos se colocan como sigue: Sánchez-Saus (2016) con una media de 16,57; Hidalgo (2019) con 16,01; Šifrar Kalan (2014) con 15,73; Samper Hernández (2002) con 13,99; Carcedo (2000c) con 12,80; Del Barrio y Vann (2018) con 11,16 y Nalesso (2019a) con 9,94. En segunda instancia, los promedios de vocablos revelan el siguiente orden de riqueza: Samper Hernández (2002) con 4,83; Carcedo (2000c) con 3,39; Šifrar Kalan (2014) con 2,50; Sánchez-Saus (2016) con 2,48; Hidalgo (2019) con 1,95; Del Barrio y Vann (2018) con 1,91; Nalesso (2019a) con 1,11. Sin embargo, dejando los valores en sí mismos, todas las agrupaciones se comportan semejantemente escribiendo más palabras en el CI05, "alimentos y bebidas", y menos en los CI12, "medios de transporte", y CI02, "la ropa". El mayor número de vocablos se registra en los centros CI05, "alimentos y bebidas"; CI10, "la ciudad"; CI16, "profesiones y oficios"; CI15, "juegos y distracciones", y el menor en los CI01, "partes del cuerpo"; CI02, "la ropa"; CI12, "medios de transporte". Por su parte, el índice de cohesión pone de relieve que las respuestas de nuestros informantes presentan, en media, un valor de 0,88 colocándose entre los que obtienen

la mayor concreción (Samper Hernández 2002, Carcedo 2000c, Del Barrio y Vann 2018) y la menor (Šifrar Kalan 2014, Sánchez-Saus 2016, Hidalgo 2019). Si nos referimos al producto de la acción docente, los resultados sugieren que los estudiantes que alcanzan un nivel mayor de cohesión desarrollan un conocimiento léxico que es el producto de un aprendizaje organizado sobre un argumento bien trabajado en el aula, mientras que a los otros quizá no se le dedica la misma atención y el índice es más abierto.

En definitiva, se localizan carencias y lagunas en la competencia de nuestros encuestados posiblemente debidas al grado de atención que le otorgan los programas curriculares al componente léxico, al método didáctico empleado y al contexto de aprendizaje, con lo cual sería interesante tener en cuenta también el tipo de enseñanza como variable en futuros trabajos. Se confirma, una vez más, la importancia de plantear actividades explícitas para la adquisición del léxico, porque el aprendizaje incidental no parece suficiente. Centrarse en el vocabulario de la lengua objeto favorecería tanto la comprensión como la expresión de los discentes que, al contrario, es difícil sin una sólida base léxica. De esto se infiere que en Italia se descuida la enseñanza de este componente probablemente por el parentesco lingüístico del italiano con el español y la sensación de facilidad que causa la poca distancia tipológica entre LM y LE (al menos en nuestra Universidad, el profesorado confirma su mayor concentración en la gramática y en la contrastividad entre italiano y español, en detrimento de otros elementos). Justificaríamos, de esta manera, el inferior rendimiento de ambos grupos de informantes italófonos.

La secunda sección del capítulo coincide con el estudio cualitativo del material a partir de los índices de disponibilidad léxica extraídos de *Dispogen* con los objetivos de averiguar las temáticas predominantes y las categorías gramaticales más difundidas en las respuestas.

El análisis transversal demuestra que la oscilación de vocablos más disponibles (con ID ≥ 0,1) va de un mínimo de tres a un máximo de veinticinco ítems por centro de interés. Se califican como más productivos: el CI05, "alimentos y bebidas"; el CI01, "partes del cuerpo"; el CI10, "la ciudad". En contra, los que tienen el número menor de vocablos coinciden con los menos rentables en el análisis cuantitativo: el CI09, "iluminación y calefacción", y el CI13, "trabajos del campo y del jardín". Al desglosar los datos por variable, sobresale la supremacía de las mujeres, los estudiantes de nivel B2 y los que conocen más de dos LE.

En lo que se refiere a la activación de las asociaciones mentales, las unidades encajan perfectamente en los ámbitos prototípicos de cada estímulo, solo en posiciones más bajas o en CI poco rentables los informantes arrojan lexías referentes a asociaciones secundarias. Además, los sustantivos son la clase preponderante, los verbos aparecen en porcentajes escasos y los adjetivos se presentan aun menos. A raíz de estos datos, la cardinalidad del conjunto en ciertos ámbitos alcanza el 100% (en CI12, "los medios de transporte", según el factor sexo; en CI08, "la escuela: muebles y materiales"; CI09, "iluminación y calefacción"; CI12, "medios de transporte", conforme al nivel de ELE; en CI06, "objetos colocados en la mesa para la comida"; CI09, "iluminación y calefacción", con respecto al conocimiento de otras LE) y en otros, al contrario, baja hasta el 0%, en el caso de que no haya ninguna coincidencia (0% en el CI09, "iluminación y calefacción", según el género; 30% en el CI07, "la cocina y sus utensilios", en la comparación entre niveles; 37,50% en el CI07, "la cocina y sus utensilios", en función del número de LE conocidas).

El análisis longitudinal revela una gran variación del índice que va del 30% en el CI09, "iluminación y calefacción", al 100% en el CI16, "profesiones y oficios". De nuevo, los sustantivos son la clase dominante. Solo detectamos verbos en el CI08, "la escuela: muebles y materiales", (10%); el CI13, "trabajos del campo y del jardín", (55%); el CI16, "profesiones y oficios", (31%). Se confirma, asimismo, un desarrollo de la competencia léxica: los encuestados de nivel B2 son capaces de activar un número mayor de lexías con ID ≥ 0,1 ya que solo un campo queda invariado y dos presentan resultados inferiores. En otras palabras, tienen a disposición una cantidad mayor de léxico disponible, según la siguiente distribución: CI01, "partes del cuerpo", -4,55%; CI02, "la ropa", +38,18%; CI03, "partes de la casa (sin muebles)", +8,33%; CI04, "los muebles de la casa", +8,33%; CI05, "alimentos y bebidas", -6,67%; CI06, "objetos colocados en la mesa para la comida", +30%; CI07, "la cocina y sus utensilios", +22,22%; CI08, "la escuela: muebles y materiales", +38,89%; CI09, "iluminación y calefacción", +70%; CI10, "la ciudad", +22,73%; CI11, "el campo", ±0%; CI12, "medios de transporte", +16,67%; CI13, "trabajos del campo y del jardín", +42,86%; CI14, "los animales", +28,57%; CI15, "juegos y distracciones", +47,83%; CI16, "profesiones y oficios", +40%.

La comparación entre el léxico disponible de nuestros informantes con el CREA demuestra que cierto porcentaje pertenece al léxico más frecuente del español (el 20% de estas unidades coincide con aquellas de la muestra que presentan un ID ≥ 0,1). Por su parte, el contraste con el CAES

revela una correspondencia del 15,30%. Es preciso recordar que ambos corpus de referencia incluyen unidades funcionales del discurso, declinadas, conjugadas, y nombres propios, lo que implica que los vocablos coincidentes habrían aumentado en gran medida si no hubiésemos editado el material recogido.

Para concluir el análisis de la disponibilidad léxica cabe señalar que todo tipo de resultado comentado hasta ahora corrobora una de las hipótesis de partida según la cual los centros de interés producen cantidades mayores de palabras y de vocablos en función de la rentabilidad que tienen para los participantes al experimento. Algunos de estos no son adecuados para cumplir con las necesidades y los intereses culturales y comunicativos de los aprendices del siglo XXI. Algo que es de sobra conocido y que se trató en muchas investigaciones anteriores que abordan este aspecto metodológico con cierto detenimiento y de manera fundamentada.[45] De hecho, existen campos poco productivos, como el CI09, "iluminación y calefacción, el CI11, "el campo", y el CI13, "trabajos del campo y del jardín". Esto podría derivar del escaso interés que suscitan o de la poca atención que se le dedica debido a la realidad socioambiental en la que viven los informantes. A este propósito, Sánchez-Saus (2016) aboga por utilizar la lista de contenidos que el MCER recomienda para el nivel A1 en la selección de los centros de interés y Santos Díaz (2017) recomienda que la nómina debiera basarse en las nociones específicas del PCIC. Todo esto pone de manifiesto la urgencia de tomar nuevas decisiones que solventen la inadecuación de algunos de los temas establecidos como *input* para la elicitación hace más de setenta años. Habría que eliminar los estímulos pocos productivos e incluir nuevos como ya propuso Hidalgo (2019) para maximizar la posibilidad de expresión del alumnado.

El estudio de la riqueza léxica ocupa el capítulo 4 abriéndose con el análisis transversal. Los resultados generales demuestran que todos los relatos alcanzan un buen nivel de riqueza en lo que atañe a los índices de variedad léxica, en cada uno se detecta más del 50% de palabras diferentes (en promedio 63%). El índice de hápax también es bueno ya que presenta valores medio-bajos (en promedio 1,39) que se aproximan al ideal. Como planteaba una de nuestras hipótesis, considerado el nivel de ELE, los indicadores revelan una apreciable variación y el desarrollo de una adecuada

[45] Se respalda lo criticado ya en proyectos precedentes, incluso los publicados en un estadio temprano de los estudios (Juilland 1970, Mackey 1971, Benítez Pérez 1994, López Morales 1999, Bombarelli 2005, Moreno Fernández 2012, González Fernández 2014, Paredes García 2014, Tomé Cornejo 2015).

competencia léxica que conlleva un buen manejo de la lengua española. La densidad léxica, por su parte, revela resultados divergentes: el 90% de los textos se compone de una cantidad satisfactoria de *tokens* léxicos (la media es de 55%), pero los valores del IAT parten de un mínimo de 1,52 y llegan a un máximo de 2,33. El promedio de 1,84 supone la aparición de una palabra semántica cada dos gramaticales y baja el grado de riqueza detectado antes. Como este indicador es bastante alto en todo tipo de análisis, no vamos a discutirlo más. Esto corrobora que un texto con un apreciable índice de variedad de vocabulario no necesariamente presenta el mismo grado de densidad, por eso se opta a menudo por la aplicación de diferentes medidas finalizadas a obtener datos lo más completos posible (Read 2000, Johansson 2008, Jarvis 2013).

A la hora de repartir los resultados según variable, el sexo no incide en ninguno de los índices estudiados. Los relatos de ambas variantes igualan o sobrepasan el 50% de variación léxica y el 85% de cada grupo presenta un buen índice de hápax. La densidad marca una leve alteración de esta correspondencia, pero no de manera significativa, determinando un ligero influjo del factor ya que en toda la producción de los hombres el índice cubre o supera la mitad del texto mientras que el mismo resultado se detectó en el 89% de los escritos de la sección femenina.

La riqueza léxica contabilizada en función del nivel de ELE tampoco supone una gran influencia. El desnivel entre los relatos no es representativo estadísticamente: todos llegan al 50% de variación o lo superan. El índice de hápax manifiesta una pequeña diferencia: el 80% de los informantes intermedios y el 90% de los avanzados proporcionan valores que suben hasta 1,50 pero tampoco esta disparidad es discriminante. Lo mismo revela la densidad léxica, ya que el 92% de los estudiantes de nivel B1 tiene un índice equivalente o superior a 50% frente al 88% de los discentes de nivel B2. De ahí que no haya destacado la predominancia de ningún grupo: si bien el dominio lingüístico aumenta, los índices no ascienden ni bajan significativamente, parece casi que el desarrollo de la competencia léxica de un nivel al otro se ha ralentizado.

El conocimiento de otras LE es la variable que tiene más influencia. La variedad léxica revela que los informantes que conocen más de dos LE han obtenido resultados mejores del 9%. De la misma manera, alcanzan el mejor índice de hápax con un aumento del 16%. Se corrobora la *Hipótesis de la Interdependencia Lingüística* (Cummins 1979), según la cual un individuo que ya conoce una LE emplea sus competencias como elemento facilitador durante el aprendizaje de otra. Por eso, extraña el cambio de

tenencia en la aplicación del índice de la densidad porque presenta valores mejores en los textos de los informantes que saben dos LE.

El análisis longitudinal da cuenta de la evolución del conocimiento del alumnado tras la asistencia a los cursos anuales de español de la Universidad, junto al estudio de la variable nivel de ELE contribuye a verificar o rechazar que el dominio lingüístico produce sus efectos en la competencia léxica llevando a un aumento. Contradiciendo nuestra hipótesis, llama la atención el índice de la diversidad léxica: los informantes han alcanzado un rendimiento mejor a principio del año académico debido a que los datos son más altos. En cambio, el índice de hápax y de densidad se asemejan. Contabilizamos un buen uso de las palabras semánticas en el 88% de la primera muestra y el 96% de la segunda. Sin embargo, los resultados difieren de poco entre los dos momentos de la administración de la encuesta. Hecho este que sorprende porque parece demostrar que no se desarrolla el conocimiento de los participantes como se ha hipotizado, es más, casi se produce un bloqueo o, incluso, un bajón.

En cuanto al análisis comparativo, la correlación de los resultados valida que aprendices no nativos de distinta procedencia que comparten el nivel de español producen en sus textos una riqueza semejante. El grado de variedad léxica es homogéneo según los siguientes valores de TTR: 0,70 en Berton (2014); 0,64 en Cuba Vega y Cuba (2004); 0,63 en Nalesso (2019a); 0,60 en Wang (2016); 0,60 en Basso (2017). Igualmente, el IAT pone de manifiesto valores parecidos, si excluimos el trabajo de Basso (2017) que llega a 2,35: Wang (2016) contabiliza un promedio de 1,80; Cuba Vega y Cuba (2004) presentan 1,81; Nalesso (2019a) consigue una media de 1,84. Parece que todos los estudiantes saben escribir composiciones aceptables y pueden desempeñar una tarea de producción simple y no extensa. Sin embargo, la uniformidad detectada en los datos de la *Type/Token Ratio* confirman buenos resultados en la capacidad expresiva del estudiantado si consideramos el nivel de ELE de los sujetos participantes.

La sección final del capítulo se plantea como acercamiento a un estudio de tipo cualitativo de la riqueza léxica, por eso las conclusiones no son definitivas puesto que se trata de una propuesta dirigida a promover nuevas investigaciones, de la que este es un experimento piloto. El 93% de las palabras más frecuentes se compone de unidades de contenido nocional, no obstante, en los primeros rangos se detecta una mayoría de unidades funcionales del discurso, necesarias para la redacción de un texto. La categoría más presente es la de los sustantivos, entre ellos *ciudad* y *viaje* son los más utilizados. A continuación, los verbos *ser* e *ir* son los

más frecuentes y constatamos también el uso de unidades pluriverbales formadas por *tomar* y de la perífrasis *ir + a + infinitivo*. Por último, observamos la presencia de adjetivos calificativos y de varios tipos de adverbios (de tiempo, lugar, modo, cantidad). En lo que se refiere a la extracción de las palabras clave del corpus, averiguamos si las unidades léxicas definidas tales pueden determinar el tema abordado en los relatos. Para ello, observamos cuatro vocablos representativos de cada clase gramatical contemplada (*vacación*, *visitar*, *estupendo*, *allí*) y analizamos los usos más comunes en el corpus.

Los resultados más destacables de esta sección atañen a las posibilidades de explotación del material recabado que podrían aplicarse en este campo y, pese a las conclusiones limitadas, nos permiten afirmar que es posible realizar un análisis diferente de los datos, finalizado a la forma más que a la cantidad, que deje de lado índices y números. Creemos que merece la pena subrayar la importancia de seguir planteando proyectos de esta naturaleza con el objetivo de fijar pautas comunes que otorguen la elaboración de comparaciones detalladas porque posibilitarían un nuevo protocolo de estudio de la riqueza léxica y más extensamente de la competencia léxica. De nuevo, apuntamos que se trata de un trabajo introductorio que deja el camino abierto, quizá, a una ampliación de las propuestas y de estudios empíricos.

En definitiva, los datos de la disponibilidad léxica apuntan que los resultados de los informantes son buenos, no sobresalientes como se esperaba por la afinidad que une el italiano al español, en particular si consideramos su escaso rendimiento en el cotejo con otros aprendientes. De todos modos, registramos un aumento del léxico disponible de un nivel a otro y entre los dos momentos de ejecución del test. La medición de la riqueza léxica, por su parte, prueba que no necesariamente la capacidad expresiva se acompaña con el avance del nivel de español. En efecto, los resultados son apreciables según el dominio lingüístico de los encuestados (incluso en el análisis comparativo), pero no se produce el mismo desarrollo de la competencia entre los dos niveles y tampoco en el análisis longitudinal. Esto contradice, en parte, la idea de que existe una relación asociativa entre el nivel de ELE y la cantidad de lexías activadas y utilizadas en las dos partes de la prueba: aunque el léxico disponible aumenta, no sucede lo mismo al vocabulario activo (al discurso). Esto demuestra la utilidad de combinar disponibilidad y riqueza léxica, cuya correlación crítica confluye en el estudio de la competencia léxica desde una perspectiva nueva y más detallada.

A modo de conclusión, logramos corroborar o refutar las hipótesis planteadas a principios de la investigación por lo que consideramos conseguido el objetivo del trabajo y esperamos haber contribuido a poner los cimientos de un nuevo sistema capaz de mejorar la ejecución de los estudios sobre la competencia léxica en ELE. En efecto, una cuestión a la que todavía no se ha encontrado una respuesta unívoca es «¿cómo se puede valorar la competencia léxica?» (Morante Vallejo 2005: 50-53): los análisis de disponibilidad resultan indispensables para diagnosticar la cantidad y la calidad de vocabulario que conocen los informantes, y el grado de asociación cognitiva de lo aprendido; al mismo tiempo, los análisis de riqueza permiten examinar sus eductos y averiguar si saben una palabra y cómo la utilizan, esto es, si el conocimiento pasivo-receptivo entra a formar parte del activo-productivo.

Como colofón, sabemos que ninguna metodología está exenta de críticas y que cada investigación surge de la anterior para mejorarla. De ahí que reiteremos que nuestro principal objetivo no perseguía aportar respuestas prescriptivas a los interrogantes marcados, sino acercarnos a explicar el complejo fenómeno de la competencia léxica mediante un aporte original.

## Referencias bibliográficas


Aabidi, L. (2019). 'La disponibilidad léxica en español como lengua extranjera: dos décadas de investigación científica', *MarcoELE* 28.

Alvar Ezquerra, M. (2003). *La enseñanza del léxico y el uso del diccionario.* Madrid: Arco/Libros, S. L.

Andrés Pérez, B. (1997). *Riqueza léxica en textos escritos de tres niveles de EGB* (Memoria de licenciatura, Universidad de Alcalá).

Anthony, L. (2018) [programa informático]. *AntConc (Versión 3.5.2).* Tokyo: Waseda University. URL http://www.laurenceanthony.net/software/antconc/

Ávila, R. (1986). 'Léxico infantil de México: palabras, tipos, vocablos', en J. D. Moreno de Alba (ed.), *Actas del Congreso del II Congreso Internacional sobre el español de América.* México: Universidad Nacional Autónoma de México, 510-517.

Ávila Muñoz, A. M. (2016). 'Can speakers' virtual lexical richness be calculated? Individual and social determining factors', *Spanish in Context* 13 (2): 285-307.

Ávila Muñoz, A. M. (2017). 'The available lexicon: A tool for selecting appropriate vocabulary to teach a foreign language', *Iranian Journal of Language Teaching Research* 5 (1): 71-91.

Ávila Muñoz, A. M. y Sánchez Sáez, J. Mª. (2010). 'La disponibilidad léxica. Antecedentes y fundamentos', en A. M. Ávila Muñoz y J. A. Villena Ponsoda (eds.), *Variación social del léxico disponible en la ciudad de Málaga.* Málaga: Editorial Sarriá, 35-81.

Ávila Muñoz, A. M. y Sánchez Sáez, J. Mª. (2011). 'La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga',

*ELUA. Estudios de Lingüística Universidad de Alicante* 25: 45-74.

Baerlocher Rocha, C. (2013). *Los errores léxicos en textos escritos en español por alumnos universitarios brasileños en formación como profesores de Español Lengua Extranjera* (Tesis, Universidad de Barcelona).

Bakonyi, H. (1933). *Die gebräuchlichsten Wörter der deutschen Sprache, für den Fremdsprachenunterricht stufenmässig zusammengestellt.* Beiträge zur Methodik des deutschen Sprachunterrichts im Ausland. Goethe-Institut zur Fortbildung ausländischer Deutschlehrer, Deutsche Akademie, Heft 1. München: Reinhardt

Bartol Hernández, J. A. (2010). 'Disponibilidad léxica y selección del vocabulario', en R. Castañer Martín y V. Lagüéns Gracia (eds.), *De moneda nunca usada. Estudios filológicos dedicados a José Mª. Enguita Utrilla.* Zaragoza: Instituto Fernando el Católico, 85-107.

Bartol Hernández, J. A. *et al.* (s.f.) [programa informático]. *DispoLex.* Universidad de Salamanca. URL http://www.dispolex.com

Basso, V. (2017). *La riqueza léxica en la producción escrita de aprendices italófonos de E/LE* (Tesis de máster inédita, Università degli Studi di Padova).

Benítez Pérez, P. (1994). 'Léxico real/irreal en los manuales de español para extranjeros', en S. Montesa Peydró y A. Garrido Moraga (eds.), *Actas del Segundo Congreso Nacional de ASELE. Español para extranjeros: didáctica e investigación.* Málaga: ASELE, 325-333.

Berton, M. (2014). *La riqueza léxica en la producción escrita de estudiantes suecos de ELE* (Tesis de máster, Stockholms Universitet).

Berton, M. (2020). *Riqueza léxica y expresión escrita en aprendices suecos de ELE. Proficiencia general, competencia léxica pasiva, tipo y complejidad de la tarea* (Tesis doctoral, Stockholms Universitet).

Bombarelli, Á. (2005). *La disponibilidad léxica como herramienta didáctica: una propuesta de selección del vocabulario para un nivel umbral de ELE* (Memoria de máster inédita, Universidad de Salamanca).

Buchanan, M. A. (1927). *A Graded Spanish Word Book.* Toronto: University of Toronto Press.

Caggiula, S. (2013). *El español como lengua extranjera: un estudio de disponibilidad léxica y su aplicación a la enseñanza.* Carolina del Norte: Lulu Press.

Callealta Barroso, F. J. y Gallego Gallego, D. (2016). 'Medidas de disponibilidad léxica: comparabilidad y normalización', *Boletín de filología* 51 (1): 39-92.

Calvi, M. V. (1995). *Didattica di lingue affini. Spagnolo e italiano.* Milano: Guerini scientifica.

Capsada Blanch, R. y Torruella Casañas, J. (2017). 'Métodos para medir la riqueza léxica de los textos. Revisión y propuesta', *Verba* 44: 347-408.

Carcedo González, A. (1998). 'Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE', *RILCE. Revista de Filología Hispánica* 14 (2): 205-224.

Carcedo González, A. (1999a). 'Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad', *Pragmalingüística* 5-6: 75-94.

---- (1999b). 'Análisis de errores léxicos del español en la interlingua de los finlandeses', en T. Jiménez Juliá, M. C. Losada Aldrey, J. F. Márquez Caneda y S. Sotelo Docío (eds.), *Español como Lengua Extranjera: Enfoque Comunicativo y Gramática.* Actas del IX Congreso Internacional de ASELE (Santiago de Compostela, 23-26 de septiembre de 1998). Santiago de Compostela: Universidad de Santiago de Compostela, 465-472.

---- (1999c). 'Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas', *Documentos de Español Actual* 1: 73-87.

Carcedo González, A. (2000a). 'La lengua como manifestación de otredad cultural (o convergencia intercultural)', *Espéculo*, Monográfico "Cultura e intercultural en la enseñanza del español como lengua extranjera".

---- (2000b). 'Índices léxico-estadísticos y graduación del vocabulario en la enseñanza de E/LE', en M. Franco Figueroa, C. Soler Cantos, J. de Cos Ruiz, M. Rivas Zancarrón y F. Ruiz Fernández (eds.), *Nuevas Perspectivas en la Enseñanza del Español como Lengua Extranjera.* Actas del X Congreso de ASELE (Cádiz, 22-25 de septiembre de 1999). Cádiz: Servicio de Publicaciones Universidad de Cádiz, 175-183.

----- (2000c). *Disponibilidad léxica en español lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición).* Turku: Turun Yliopisto.

Castañeda-Jiménez, G. y Jarvis, S. (2014). 'Exploring lexical diversity in second language Spanish', en K. Geeslin (ed.), *The handbook of Spanish second language acquisition.* Chichester: Wiley-Blackwell, 498-513.

Cintrón Serrano, F. (1992). *Índices de riqueza léxica en escolares de Barranquitas* (Tesis de maestría, Universidad de Puerto Rico).

Consejo de Europa (2001). *Marco común europeo de referencia para las lenguas: Aprendizaje, Enseñanza, Evaluación.* (Traducción al español del Instituto Cervantes en 2002). Madrid: MECD-Anaya.

Cuba Vega, L. E. y Cuba, Y. M. (2004). 'Las pruebas de riqueza léxica y

su aplicación en Español como Lengua Extranjera (ELE)', en A. Díez Mediavilla (ed.), P. Couto Cantero, F. Vieito Liñares y E. Aradas Carollo (coords.), *Actas VIII Congreso Internacional Sociedad Española de Didáctica de la Lengua y la Literatura (La Habana, 5-9 de diciembre de 2004): Homenaje a María Zambrano y Alejo Carpentier*, 815-830.

Cummins, J. (1979). 'Linguistic interdependence and the educational development of bilingual children', *Review of educational research* 49 (2): 222-251.

Del Barrio de la Rosa, F. (2016). 'Algunas observaciones sobre la disponibilidad léxica en estudiantes italófonos de español', en E. Sainz González, I. Solís García, F. Del Barrio de la Rosa e I. Arroyo Hernández (eds.), *Geométrica explosión. Estudios de lengua y literatura en homenaje a René Lenarduzzi* . Venezia: Edizioni Ca' Foscari, 127-143.

Del Barrio de la Rosa, F. (2017a). 'Los estudiantes italófonos de ELE y los estudios de disponibilidad léxica en español', *Revista Nebrija de Lingüística Aplicada* 22: 52-57.

---- (ed.). (2017b). *Palabras. Vocabulario. Léxico. La lexicología aplicada a la didáctica y a la diacronía*. Venezia: Edizioni Ca' Foscari – Digital Publishing.

Del Barrio de la Rosa, F. (2018). 'Pares léxicos en el léxico disponible de estudiantes italianos', *Lingue e linguaggi* 26: 173-196.

Del Barrio de la Rosa, F. y Mae Vann, M. (2018). *Disponibilidad léxica de los estudiantes de español en Italia. Estudio y diccionarios*. Canterano: Aracne.

Di Gesù, F. (2016). *Linguistica contrastivo-percettiva di lingue tipologicamente affini: italiano e spagnolo*. Palermo: University press.

Dimitrijévic, N. (1969). *Lexical Availability. A new aspect of the lexical availability of secondary school children*. Heidelberg: Julius Gross Verlag.

Echeverría, M. S. *et al.* (1987). 'Disponibilidad léxica en Educación Media. Resultados cuantitativos', *RLA. Revista de lingüística teórica y aplicada* 25: 55-115.

Echeverría, M. S. *et al.* (1992). 'Evaluación de la Riqueza Léxica en Estudiantes de Último Año de Enseñanza Media', *Estudios Filológicos* 27: 59-72.

Echeverría, M. S. *et al.* (2005) [programa informático]. *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción de Chile: Universidad de Concepción.

Echeverría, M. S. *et al.* (2008). 'DispoGrafo: una nueva herramienta

computacional para el análisis de relaciones semánticas en el léxico disponible', *RLA. Revista de lingüística teórica y aplicada* 46 (1): 81-91.

Fernández Juncal, C. y Hernández Muñoz, N. (2018). 'Vías de transformación en la enseñanza de lenguas con mediación tecnológica', *CLAC. Círculo de Lingüística Aplicada a la Comunicación* 76: 3-12.

García Hoz, V. (1953). *Vocabulario usual, vocabulario común y vocabulario fundamental.* Madrid: CSIC.

García Marcos, A. (2019). *Adquisición del español como segunda lengua. El caso de la integración lingüística de escolares inmigrantes en Almería.* La Cañada de San Urbano, Almería: Editorial Universidad de Almería.

García Rosas, M. P. (1996). *Riqueza léxica en textos de estudiantes de español como lengua extranjera* (Memoria de máster inédita, Universidad de Alcalá).

Gómez Devís, Mª. B. (2019). 'A propósito de las redes semánticas en el léxico disponible de escolares de primero de Educación Primaria', *Ogigia. Revista electrónica de estudios hispánicos* 25: 165-183.

Gómez Molina, J. R. y Gómez Devís, Mª. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos. Estudio de estratificación sociolingüísticas.* Valencia: Universitat de València.

Gómez Sánchez, M. E. y Guerra Salas, L. (2004). 'Disponibilidad y fines específicos: análisis del centro de interés prensa', en I. Sanz Sainz; Á. M. Felices Lago (coords.), *Las nuevas tendencias de las lenguas de especialidad en un contexto internacional y multicultural. III Congreso Internacional de la Asociación Europea de Lenguas para Fines Específicos (AELFE).* Granada: Universidad de Granada, 695-703.

González Fernández, J. (2013). 'La disponibilidad léxica de los estudiantes turcos de español como lengua extranjera', *MarcoELE* 16.

González Fernández, J. (2014). 'Idoneidad de los centros de interés clásicos en los estudios de disponibilidad léxica aplicados al español como lengua extranjera', *Revista Nebrija de Lingüística Aplicada* 16: 41-53.

Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. (1964). *L'élaboration du français fundamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base.* Paris: Didier.

Guerra Salas, L. y Gómez Sánchez, M. E. (2004). 'Español de los medios de comunicación: aspectos de disponibilidad léxica', en H. Perdiguero; A. Álvarez (ed.), *Medios de comunicación y enseñanza del español como lengua extranjera.* Actas del XIV Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera (Burgos, 2003). Burgos: Servicio de publicaciones Universidad de Burgos, 356-371.

Guiraud, P. (1954). *Les caractères statistiques du vocabulaire. Essai de méthodologie*. Paris: Presses Universitaires de France.

Haché, A. M. (1991). 'Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna', en H. López Morales (ed.), *La enseñanza del español como lengua materna*. Río Piedras: Universidad de Puerto Rico, 47-60.

Hallebeek, J. (1986). 'Las palabras funcionales del español', *Boletín AEPE* 34-35: 205-216.

Ham Chande, R. (1979). 'Del 1 al 100 en lexicografía', en L. Fernando Lara, R. Ham Chande y Mª. I. García Hidalgo (eds.), *Investigaciones lingüísticas en lexicografía*. México: El Colegio de México, 110-132.

Henmon, V. A. C. (1924). A *French Word Book Based on a Count of 400.000 running Words*. Madison (Wisconsin): Bureau of Educational Research, University of Wisconsin.

Hernández Muñoz, N. (2005). 'La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología', en E. Díez Villoria, B. Zubiauz de Pedro y Mª. Á. Mayor Cinca (coords.), *Estudio sobre la adquisición del lenguaje*. Salamanca: Ediciones Universidad de Salamanca, 942-953.

Hernández Muñoz, N. (2014). 'Categorías en el léxico bilingüe perspectivas desde el priming semántico interlenguas y la disponibilidad léxica', *RAEL: Revista Electrónica de Lingüística Aplicada* 13 (1): 19-38.

Hernández Muñoz, N. (2015). 'La evaluación de la competencia léxica adulta una aproximación a través de la disponibilidad léxica y la especialización académica en preuniversitarios', *Revista de Filología de la Universidad de La Laguna* 33: 79-99.

Hernández Muñoz, N., Izura, C. y Tomé Cornejo, C. (2014). 'Cognitive Factors of Lexical Availability in a Second Language', en R. Mª. Jiménez Catalán (ed.), *Lexical Availability in English and Spanish as a Second Language*. New York: Springer, 169-188.

Hidalgo Gallardo, M. (2017). 'Sobre la disponibilidad léxica en ELE', *Boletín de ASELE* 56: 83-94.

Hidalgo Gallardo, M. (2019). *Factores y fuentes que inciden en el léxico disponible de estudiantes sinohablantes de ELE*. Monografías ASELE.

Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes-Biblioteca Nueva.

Instituto Cervantes (2021). El español: una lengua viva. Informe 2021. URL https://cvc.cervantes.es/lengua/espanol_lengua_viva/pdf/espanol_lengua_viva_2021.pdf

Instituto Cervantes. Banco de datos (CAES) [en línea]. *Corpus de aprendices del español*. URL http://galvan.usc.es/caes

Izquierdo Gil, M. C. (2005). *La selección del léxico en la enseñanza del español como lengua extranjera*. Málaga: ASELE.

Jarvis, S. (2013). 'Capturing diversity in lexical diversity', *Language Learning* 63: 87-106.

Jiménez Catalán, R. Mª. (2017). 'Estudios de disponibilidad léxica en español y en inglés: revisión de sus fundamentos empíricos y metodológicos', *Revista Nebrija de Lingüística Aplicada* 22: 16-31.

Johansson, V. (2008). 'Lexical diversity and lexical density in speech and writing: a developmental perspective', *Working Papers* 53: 61-79.

Juilland, A. y Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words*. London, The Hague: Mouton.

Juilland, A. (ed.). (1970). *Frequency dictionary of French Words*. London, The Hague: Mouton.

Käding, J. W. (1897). *Häufigkeitswörterbuch der Deutschen Sprache*. Steiglitz bei Berlin: der Herausgeber.

Kilgarriff, A. *et al.* (2004) [programa informático]. *The Sketch Engine*. URL http://www.sketchengine.eu

Laufer, B. (1991). 'The development of L2 lexis in the expression of the advanced language learner', *The Modern Language Journal* 75 (4): 440-448.

Laufer, B. y Nation, P. (1995). 'Vocabulary Size and Use: Lexical Richness in L2 Written Production', *Applied Linguistics* 16 (3): 307-22.

Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. London: Language Teaching Publications.

Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. London: Language Teaching Publications.

Lewis, M. (2000). *Teaching collocations. Further developments in the Lexical Approach*. London: Language Teaching Publications.

Linnarud, M. (1986). *Lexis in Composition: A Performance Analysis of Swedish Learners' Written English*. Malmö: CWK Gleerup.

López Chávez, J. (1992). 'Alcances panhispánicos del léxico disponible', *Lingüística* 4: 26-124.

López Chávez, J. (1995). 'Léxico fundamental panhispánico: realidad o utopía', en A. Matus Oliver (ed.), *Actas del III Congreso Internacional sobre el Español de América* (vol. 2). Santiago de Chile: Universidad Católica de Chile, 1006-1014.

López Chávez, J. y Strassburger Frías, C. (1987). 'Otro cálculo del índice de disponibilidad léxica', en *Presente y perspectiva de la lingüística*

*computacional en México*. Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada. México: Universidad Nacional Autónoma de México, 91-111.

López Morales, H. (1973). *Disponibilidad léxica en escolares de San Juan.* MS.

López Morales, H. (1984). *La enseñanza de la lengua materna. Lingüística para maestros de español.* Madrid: Editorial Playor.

López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros, S. L.

López Morales, H. (2011). 'Los índices de «riqueza léxica» y la enseñanza de lenguas', en J. de Santiago Guervós, H. Bongaerts, J. J. Sánchez Iglesias y M. Seseña (eds.), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE.* Actas del XXI Congreso Internacional de la ASELE (Salamanca, 29 de septiembre-2 de octubre de 2010) (vol. 1). Salamanca: Imprenta Kadmos, 15-28.

Lucas Puerta, J. (2006). *La competencia léxica en el discurso escrito en e/le de aprendices francófonos de ascendencia hispanófona* (Memoria de máster, Universidad de Barcelona).

Lucha Cuadros, R. y Díaz, L. (2016). 'El efecto positivo del trabajo de la expresión escrita en un curso de ELE general: un estudio empírico longitudinal con aprendices multilingües', *MarcoELE* 23: 1-16.

Luján García, C. I. y Bolaños Medina, A. (2014). 'Disponibilidad léxica y anglicismos informáticos en los centros de interés: internet, software y hardware', *Odisea* 15: 101-126.

Mackey, W. F. (1971). *Le vocabulaire disponible du Français*, 2 vols. Paris: Didier.

Madrigal-Melchor, J., Rivera-Juárez, J. M., Enciso-Muñoz, A. y López-Chávez J. (2012). 'Disponibilidad léxica para medir el crecimiento conceptual de electricidad', *Latin-American Journal of Physics Education* 6 (4): 648-651.

Michéa, R. (1950). 'Vocabulaire et culture', *Les Langues Modernes* 44: 188-189.

Michéa, R. (1953). 'Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage', *Les Langues Modernes* 47: 338-344.

Morante Vallejo, R. (2005). *El desarrollo del conocimiento léxico en segundas lenguas.* Madrid: Arco/Libros, S. L.

Moreno Fernández, F. (2012). 'Disponibilidad léxica: cuestiones metodológicas. A propósito de disponibilidad léxica de los estudiantes hispanos de Redwood City, CA', *Revista Nebrija de Lingüística Aplicada* 11: 53-61.

Müller, C. (1968). *Estadística lingüística*. Madrid: Gredos.

Nalesso, G. (2018a). 'El desarrollo de la competencia léxica de estudiantes italianos universitarios de ELE', *Orillas. Rivista d'Ispanistica* 7: 381-394.

---- (2018b). 'La eficacia de actividades específicas para el aprendizaje del léxico en ELE. Estudio de caso', en Mª. Bargalló Escrivá, E. Forgas Berdet y A. Nomdedeu Rull (eds.), *Léxico y cultura en LE/L2: corpus y diccionarios*. Actas del XXVIII Congreso Internacional ASELE (Tarragona, 6-9 de septiembre de 2017). Tarragona: Copysan, 505-514.

Nalesso, G. (2019a). *Disponibilidad y riqueza léxica en un grupo de estudiantes universitarios italianos de ELE* (Tesis doctoral, Università degli Studi di Padova).

---- (2019b). 'La eficacia de actividades para el aprendizaje del léxico especializado del arte en ELE. Estudio de caso', en C. Ramos Méndez y P. Salamanca Fernández (eds.), *Competencias en el aula de español como lengua extranjera en contextos universitarios y profesionales.* Actas del Congreso Internacional UniPro 2018 Múnich (Múnich, 2 de febrero de 2018). Múnich: Hochschule für Angewandte Sprachen, 100-116.

Nalesso, G. (2020). 'Disponibilidad léxica terminológica en ELE: una propuesta de análisis', *Orillas. Rivista d'Ispanistica* 9: 609-631.

Nalesso, G. (2022). 'Disponibilidad léxica y ortografía en ELE: un estudio para la enseñanza de la lengua', *RILEX. Revista Sobre Investigaciones léxicas* 5 (I): 7-36.

Navarro Marrero, Y. (2010). 'Terminología especializada en el área de fisioterapia: acercamiento desde la metodología de la disponibilidad léxica específica', *Interlingüística* 20.

Njock, P. E. (1979). *L'univers familier de l'enfant africain*. Québec: CIRB.

Paolini, A. (2017). *Estudio sobre disponibilidad léxica en alumnos de ELE en la Universidad de Padua* (Tesis de máster, Università degli Studi di Padova).

Paredes García, F. (2014). 'A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE', *Revista Nebrija de Lingüística Aplicada* 16: 54-59.

Paredes García, F. (2015). 'Disponibilidad Léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula', *Linred* 13.

Paredes García, F. (2017). 'La comparabilidad de los trabajos de disponibilidad léxica', *Revista Nebrija de Lingüística Aplicada* 22: 71-

77.
Pérez Serrano, M. (2009). *Estudio de disponibilidad léxica en estudiantes de E/LE en los centros de interés "medios de transporte" y "profesiones y oficios"* (Memoria de máster, Instituto Cervantes-UIMP).
Pfeffer, J. A. (1964). *Grundeutsch Basic (spoken) German Word list Grundstufe*. New Jersey: Prentice Hall Inc. Englewood Cliffs.
Portela, C. (1992). *Índices de riqueza léxica en estudiantes del primer año universitario* (Tesis de maestría, Pontificia Universidad Católica Madre y Maestra).
Read, J. A. S. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
Real Academia Española. (2005) [en línea]. *Diccionario panhispánico de dudas*. URL http://www.rae.es/recursos/diccionarios/dpd
Real Academia Española. (2009). *Nueva Gramática de la Lengua Española, vol. I: Morfología y Sintaxis*. Madrid: Espasa.
Real Academia Española. (2010). *Ortografía de la lengua española*. Madrid: Espasa.
Real Academia Española. (2017) [en línea]. *Diccionario de la Lengua Española*, 23.1ª edición. URL http://dle.rae.es
Real Academia Española. Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. URL http://corpus.rae.es/creanet.html
Reyes Díaz, M. J. (2007). 'Apuntes para la enseñanza del vocabulario', *Revista de Filología de la Universidad de La Laguna* 25: 529-538.
Roberto, J. A., Martí, M. A. y Salamó Llorente, Mª. (2012). 'Análisis de la riqueza léxica en el contexto de la clasificación de atributos demográficos latentes', *Procesamiento de Lenguaje Natural* 48: 97-104.
Rodríguez Bou, I. (dir.). (1952). *Recuento de vocabulario español*. San Juan de Puerto Rico: Organización de Estados Americanos.
Rubio Lastra, M. (2018). 'Disponibilidad léxica de 52 estudiantes taiwaneses universitarios de ELE A1', *MarcoELE* 27.
Rubio Sánchez, R. (2015). 'Estudio de disponibilidad léxica en aprendices italianos de español: análisis cuantitativo', en A. Gordejuela Senosiáin, D. Izquierdo Alegría, F. Jiménez Berrio, A. de Lucas Vicente y M. Casado Velarde (eds.), *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra, 429-441.
Rubio Sánchez, R. (2017). 'Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de Español como Lengua Extranjera', en F. Del Barrio de la Rosa (ed.), *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía*. Venezia:

Edizioni Ca' Foscari, 143-161.

Ruiz Basto, A. (1987). *Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades, Plantel Naucalpan* (Tesis doctoral, Universidad Nacional Autónoma de México).

Salcedo P., y del Valle, Mª. (2013). *La disponibilidad léxica matemática en estudiantes de enseñanza media de la ciudad de Concepción de Chile. Investigación lexicométrica.* Universidad de Concepción: Concepción de Chile.

Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera.* Málaga: ASELE.

Samper Padilla, J. A. (1998). 'Criterios de edición del léxico disponible: sugerencias', *Lingüística* 10: 311-333.

Samper Padilla, J. A., Bellón Fernández, J. J. y Samper Hernández, M. (2003). 'El proyecto de estudio de la disponibilidad léxica en español', en R. Ávila, J. A. Samper y H. Ueda (eds.), *Pautas y pistas en el análisis del léxico hispano (americano).* Madrid: Vervuert-Iberoamericana, 27-140.

Samper Padilla, J. A. y Samper Hernández, M. (2006). 'Aportaciones recientes de los estudios de disponibilidad léxica', *Lynx: Panorámica de estudios lingüísticos* 5: 5-95.

Sánchez-Saus Laserna, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas.* Sevilla: Editorial Universidad de Sevilla.

Santos Díaz, I. C. (2017). 'Selección del léxico disponible: propuesta metodológica con fines didácticos', *Porta Linguarum* 27: 122-139.

Šifrar Kalan, M. (2009). 'Disponibilidad léxica en español lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Finlandia', *Verba hispanica* 17: 165-182.

Šifrar Kalan, M. (2012). 'Análisis comparativo de la disponibilidad léxica en español como lengua extranjera (ELE) y lengua materna (ELM)', *MarcoELE* 15.

Šifrar Kalan, M. (2014). 'Disponibilidad léxica en diferentes niveles de español/lengua extranjera', *Studia Romanica Posnaniensia* 41 (1): 63-85.

Šifrar Kalan, M. (2018). 'La influencia de estancia Erasmus en el léxico disponible de ELE: el caso de los centros de interés sobre España', en *International conference "Exploring the lexicon of bilingual and plurilingual learners: lexical availability and vocabulary acquisition", Universidad de La Rioja* (Logroño, 4-5 de octubre de 2018). Logroño: Universidad de La Rioja, Servicio de Publicaciones, 44.

Tesitelová, J. (1992). *The main areas of quantitative linguistics*. New York: Planum Press.

Thorndike, E. L. (1921). *The Teacher's Word Book*. New York: Teachers College, Columbia University.

Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE* (Tesis doctoral, Universidad de Salamanca).

Tomé Cornejo, C. (2016). 'Vocabulario de la informática y las nuevas tecnologías. Caracterización desde la disponibilidad léxica', *Caracteres: estudios culturales y críticos de la esfera digital* 5 (1): 112-139.

Tomé Cornejo, C. (2018). 'Estrategias de recuperación del léxico disponible en español como lengua materna y como lengua extranjera', en *International conference "Exploring the lexicon of bilingual and plurilingual learners: lexical availability and vocabulary acquisition", Universidad de La Rioja* (Logroño, 4-5 de octubre de 2018). Logroño: Universidad de La Rioja, Servicio de Publicaciones, 42-43.

Toniolo, Mª. T. y Zurita, Mª. E. (2019). 'Léxico vitícola y vitivinícola de uso en Córdoba, Argentina. Estudio léxico de especialidad', en Mª. L. Perassi; M. Tapia Kwiecien (comp.), *Palabras como puentes: Estudios lexicológicos, lexicográficos y terminológicos desde el Cono Sur*. Córdoba: Buena Vista Editores, 85-100.

Torijano Pérez, A. (2004). *Errores de Aprendizaje, Aprendizaje de Los Errores*. Madrid: Arco/Libros, S. L.

Torres González, A. N. (1999). 'Incidencia de las variables sociales en los índices producción léxica de estudiantes del último curso de la enseñanza no universitaria', en J. de Las Cuevas y D. Fasla Fernández (eds.), *Contribuciones al estudio de la lingüística aplicada*. Castellón: Asociación Española de Lingüística Aplicada, 393-401.

Torres González, A. N. (2003a). 'Riqueza léxica en textos narrativos escritos por estudiantes de Tenerife', en F. Moreno Fernández, F. Gimeno Menéndez, J. A. Samper, Mª. L. Gutiérrez Araus, M. Vaquero y C. Hernández Alonso (coords.), *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*. Madrid: Arco/Libros, S. L., 435-449.

---- (2003b). 'La producción léxica de estudiantes no universitarios de Tenerife', en C. Díaz Alayón, M. Morera y G. Ortega (eds.), *Estudios sobre el español de Canarias*. Actas del I Congreso Internacional sobre Español de Canarias. Islas Canarias: Academia Canaria de la Lengua, 989-1009.

Tracy-Ventura, N. (2017) 'Combining corpora and experimental data to investigate language learning during residence abroad: A study of

lexical sophistication', *System* 71: 35-45.
Urzúa, P., Sáez, K. y Echeverría, M. S. (2006). 'Disponibilidad léxica matemática. Análisis cuantitativo y cualitativo', *Revista Lingüística Teórica y Aplicada* 44 (2), II Sem.: 59-76.
Van der Beke, G. E. (1935). *French Word Book*. New York: The Maximilian Company, Publications of the American and Canadian Committees on Modern Languages, vol. XV.
Wang, X. (2016). *Riqueza léxica en muestras de lengua escrita de estudiantes sinohablantes de pregrado y postgrado* (Tesis de maestría inédita, Universidad de La Habana).
West, M. (1953). *A General List of English Words*. London: Longman.
Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.
Zipf, G. K. (1935). *Psycho-biology of languages*. Cambridge: MIT Press.

La collana *Incipit* accoglie due serie distinte: le *Tesi*, selezionate fra quelle discusse all'interno del Dottorato in Scienze Linguistiche, Filologiche e Letterarie dell'Università di Padova e/o sotto la supervisione di docenti del Dipartimento di Studi Linguistici e Letterari dell'Università di Padova (DiSLL); i *Colloqui*, gli atti dei convegni organizzati annualmente da allievi e allieve del Dottorato.

El presente volumen estudia la competencia léxica de un grupo de aprendientes italófonos de español lengua extranjera (ELE). Para ello, enmarcándose en el ámbito de investigación sobre la didáctica del léxico, introduce cambios metodológicos con respecto a los estudios realizados en este marco hasta la fecha con el objetivo de profundizar el conocimiento de un fenómeno tan complejo como el dominio del vocabulario, ya que se emplean –conjuntamente por primera vez– dos metodologías lexicoestadísticas que lo describen cuantitativa, cualitativa y comparativamente. De un lado, la disponibilidad léxica, colocándose en el plano paradigmático de la lengua, analiza no solo de qué y cuántas palabras dispone un aprendiente, sino cómo se organizan conceptualmente en su lexicón mental (léxico potencial); por su parte, la riqueza léxica, ubicándose en el plano sintagmático, averigua qué y cuántas palabras se manejan en un educto (léxico activo). Asimismo, se realiza la recogida de datos en dos momentos distintos para examinar la evolución diacrónica de la competencia de los informantes. Este tipo de análisis –transversal y longitudinal– pretende obtener las claves necesarias para observar el desarrollo temporal de la adquisición del léxico y, por tanto, la efectividad de la acción didáctica.

ISBN 978-88-6938-331-1


**€ 20,00**