



Universitat de Lleida

Disponibilidad léxica de estudiantes de ELE en Tianjin, China: estudio comparado entre lengua extranjera y lengua materna

Volumen I: Estudio

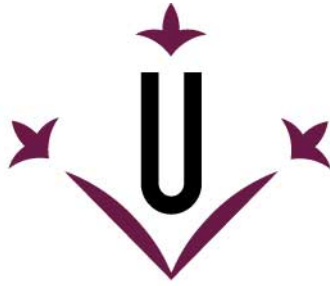
Lei Chen

<http://hdl.handle.net/10803/673254>

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



Universitat de Lleida

TESIS DOCTORAL

**Disponibilidad léxica de estudiantes de ELE
en Tianjin, China:
estudio comparado entre
lengua extranjera y lengua materna**

Volumen I: Estudio

Lei Chen

Memoria presentada para optar al grado de Doctor por la
Universitat de Lleida
Programa de Doctorado en Territorio, Patrimonio y Cultura

Directora y tutora
M^a Ángeles Calero Fernández

2021

RESUMEN

Esta tesis doctoral tiene como finalidad contribuir a los estudios de disponibilidad léxica en estudiantes de español como lengua extranjera, tomando como objeto analítico a alumnos sinohablantes en su país de origen y, concretamente, en Tianjin, China. La originalidad de nuestra investigación consiste en analizar el léxico disponible de los informantes tanto en español como lengua extranjera (ELE) como en chino como lengua materna (CLM), para comparar los resultados en las dos lenguas. Tomando como referencia la metodología establecida por el *Proyecto Panhispánico de Léxico Disponible*, determinamos nuestros procedimientos metodológicos de recogida de datos y su edición. El primer propósito del presente trabajo es recopilar el léxico disponible en español y en chino de los informantes procedentes de los centros educativos en Tianjin (288 estudiantes en el corpus de español y 172 en el de chino).

El segundo objetivo de la tesis es analizar cuantitativamente las influencias de los centros de interés (los dieciséis clásicos del *Proyecto Panhispánico*) y de las variables extralingüísticas (nivel de estudio, sexo, procedencia, nivel de inglés y tipo de centro) en los resultados del léxico disponible. Como hemos adelantado, la novedad de nuestro trabajo es examinar las diferencias y similitudes en los resultados de las dos lenguas contempladas. En cuanto a las variables, se intenta describir las tendencias que muestran los subgrupos según las variantes en lo referente a la producción léxica en las dos lenguas. Así se puede tener una idea de cuánto se aleja el conocimiento léxico en español de los estudiantes de su competencia léxica en chino.

El tercer punto de que se ocupa la tesis es un análisis cualitativo que consiste en revisar las convergencias y divergencias conceptuales que existen entre las listas de palabras en español y en chino de cada centro de interés. Se comparan los vocablos altamente disponibles de los alumnos en las dos lenguas en busca de los homólogos conceptuales de los vocablos de una

lengua en la otra. De este modo, si los alumnos aprenden a actualizar en español los conceptos que solamente son capaces de hacer en chino, se compensa considerablemente su deficiencia léxica en la lengua española.

RESUM

Aquesta tesi doctoral té com a finalitat contribuir als estudis de disponibilitat lèxica en estudiants d'espanyol com a llengua estrangera, prenent com a objecte analític alumnes sinohablants al seu país d'origen i, concretament, a Tianjin, Xina. L'originalitat de la nostra investigació consisteix en analitzar el lèxic disponible dels informants tant en espanyol com a llengua estrangera com en la seva llengua materna, per comparar els resultats en les dues llengües. Prenent com a referència la metodologia establerta pel Projecte Panhispànic de Lèxic Disponible, vam determinar els nostres procediments metodològics de recollida de dades i la seva edició. El primer propòsit del present treball és recopilar el lèxic disponible en espanyol i en xinès dels informants procedents de centres educatius a Tianjin (288 estudiants en el corpus d'espanyol i 172 en el de xinès).

El segon objectiu de la tesi és analitzar quantitativament les influències dels centres d'interès (els setze clàssics del Projecte Panhispànic) i de les variables extralingüístiques (nivell d'estudi, sexe, procedència, nivell de l'anglès i tipus de centre) en els resultats del lèxic disponible. Com hem avançat, la novetat del nostre treball és examinar les diferències i similituds en els resultats de les dues llengües contemplades. Pel que fa a les variables, s'intenta descriure les tendències que mostren els subgrups segons les variants pel que fa a la producció lèxica en les dues llengües. Així es pot tenir una idea de quant s'allunya el coneixement lèxic en espanyol dels estudiants de la seva competència lèxica en xinès.

El tercer punt del qual s'ocupa la tesi és una anàlisi qualitativa consistent a revisar les convergències i divergències conceptuals que hi ha entre les llistes de paraules en espanyol i en xinès de cada centre d'interès. Es comparen els vocables altament disponibles dels alumnes en les dues llengües a la recerca dels homòlegs conceptuals dels vocables d'una llengua a l'altra. D'aquesta manera, si els alumnes aprenen a actualitzar en espanyol els conceptes que només són capaços de fer en xinès, es compensa considerablement la seva

deficiència lèxica en la llengua espanyola.

ABSTRACT

The aim of this doctoral thesis is to contribute to the studies of lexical availability in students of Spanish as a foreign language, taking as analytical object Chinese-speaking students in their country of origin and, specifically, in Tianjin, China. The originality of our research consists in analyzing the available lexicon of the informants both in Spanish as a foreign language and in their mother tongue, in order to compare the results in the two languages. Taking as a reference point the methodology established by the Pan-Hispanic Project of Available Lexicon, we determine our methodological procedures for data collection and edition. The first purpose of this work is to compile the available lexicon in Spanish and Chinese from informants of educational centers in Tianjin (288 students in the Spanish corpus and 172 in the Chinese corpus).

The second objective of the thesis is to analyze quantitatively the influences of the centers of interest (the sixteen traditional centers of the Pan-Hispanic Project) and of the extralinguistic variables (level of study, sex, origin, level of English and type of center) on the results of the available lexicon. As we have mentioned, the novelty of our work is to examine the differences and similarities in the results of the two languages. Regarding the variables, an attempt is made to describe the tendencies shown by the subgroups according to the variants in terms of lexical production in the two languages. Through this way, we can get an idea of how far the students' lexical knowledge in Spanish differs from their lexical competence in Chinese.

The third point that the thesis deals with is a qualitative analysis which consists of reviewing the conceptual convergences and divergences that exist between the lists of words in Spanish and in Chinese of each center of interest. The highly available words of the students in the two languages are compared in search of the conceptual homologues of the words of one language in the other. In this way, if the students learn to update in Spanish the concepts that they are only capable of using in Chinese, their lexical deficiency in the Spanish language will be considerably compensated.

AGRADECIMIENTOS

Merece la pena recordar que esta tesis doctoral ha sido el resultado del respaldo y enseñanza que he obtenido de mis familiares, profesores y amigos.

Tantos motivos tengo para expresar mi gratitud a mi tutora y directora de tesis, M^a Ángeles Calero Fernández. Es ella quien me introdujo en el mundo de los estudios de disponibilidad léxica, hablándome, en diversas ocasiones en 2016, de lo que era y de sus aplicaciones a la sociolingüística y a la didáctica del español como lengua extranjera. Es ella quien me hizo ver que esta línea de investigación no estaba muy lejos de mí, animándome a asistir a las *Jornadas de Investigación en Disponibilidad Léxica* de 2017 en Salamanca, donde pude conocer a los grandes autores que he leído y citado en la tesis, y a participar en las *II Jornadas de Investigación en Disponibilidad Léxica* de 2019 en Lleida, donde me permití probar la primera dulzura de mis intentos de investigación. Es ella quien, durante los cuatro años de la elaboración de esta tesis, me consoló y me dio seguridad para avanzar, ya que en la realización del trabajo se me plantearon problemas que no había encontrado nunca en mi vida. También es ella quien revisó esta tesis apartado por apartado, dándome sugerencias valiosas de mejoramiento. Es imposible enumerar, entre unas cuantas líneas, todos los favores y beneficios que me ha brindado porque han sido de todos los aspectos y en todo momento, y todavía casi en silencio. Aunque nos hemos ido a juntar dos personas tímidas —como ella me dijo una vez—, he podido sentir en ella una persona siempre dinámica y un espíritu siempre tierno. Precio su actitud y modo de hacer las cosas, de los que he aprendido mucho.

Estoy agradecido por el cuidado que he recibido de Javier Terrado Pablo y Olga Fransi González, que me ayudaron, en las primeras etapas, a adaptarme a las nuevas situaciones que suponía mi vida en un país lejano y una cultura distinta y que nunca dejaron de ofrecerme apoyo durante los cuatro años de mis estudios.

Transmito mis sinceros agradecimientos a Maribel Serrano Zapata y M^a Begoña Gómez Devís, dos investigadoras importantes de la disponibilidad léxica, por las instrucciones que he obtenido de ellas unas veces hablando con ellas y otras escuchado sus discursos.

Dirijo también mi gratitud al equipo del *Proyecto Panhispánico de Léxico Disponible*, sobre todo a José Antonio Bartol Hernández, Natividad Hernández Muñoz y Carmela Tomé Cornejo, que me concedieron acceso al programa informático *Dispolex* y que contestaron mis dudas sobre su uso.

Además, este trabajo no sería posible sin el pleno apoyo de los alumnos y profesores de las universidades a que fui a pasar las encuestas. Ellos participaron desinteresadamente en la recogida de los datos, a pesar del inconveniente de las horas de las encuestas y del poco conocimiento del sentido de la investigación.

Finalmente, agradezco al China Scholarship Council la beca predoctoral que me ha concedido para que pueda dedicarme a la investigación; a la Tianjin Foreign Studies University, donde trabajo, por facilitarme la licencia para estar cuatro años en la Universidad de Lleida; y a la Universidad de Lleida por homologarme como personal docente e investigador en formación.

ÍNDICE

	Páginas
1. INTRODUCCIÓN	16
2. PRESUPUESTOS TEÓRICOS	27
2.1. La ciencia estadística y sus aplicaciones	27
2.1.1. <i>Las estadísticas y la estadística como ciencia</i>	27
2.1.2. <i>Orígenes históricos y nacimiento de la estadística</i>	28
2.1.3. <i>Aplicación de la estadística en las ciencias sociales</i>	31
2.1.3.1. <i>Condiciones científicas para aplicar la estadística en las ciencias sociales</i>	31
2.1.3.2. <i>Contacto de la estadística con variadas áreas de las ciencias sociales</i>	33
2.1.3.3. <i>Lingüística estadística: aplicación de la estadística en la lingüística</i>	35
2.1.4. <i>Metodología general de una investigación estadística</i>	37
2.2. Léxico-estadística: estudios estadísticos sobre el léxico	39
2.2.1. <i>La noción de la léxico-estadística</i>	39
2.2.1.1. <i>Características del léxico al encuentro de la estadística</i>	39
2.2.1.2. <i>Evolución de la medición estadística del léxico</i>	41
2.2.2. <i>De las listas de frecuencia léxica al léxico básico</i>	43
2.2.3. <i>Nacimiento de la disponibilidad léxica</i>	51
3. ESTADO DE LA CUESTIÓN	56
3.1. Estudios pioneros de disponibilidad léxica	56
3.2. Estudios de base de disponibilidad léxica en el mundo hispánico	82
3.2.1. <i>El Proyecto Panhispánico de Disponibilidad léxica</i>	82
3.2.2. <i>Estudios fundamentales de disponibilidad léxica</i>	85
3.2.2.1. <i>Estudios fundamentales en Hispanoamérica</i>	85
3.2.2.2. <i>Estudios fundamentales en España</i>	105
3.3. Aplicaciones de la disponibilidad léxica	116

4. METODOLOGÍA	129
4.1. Entorno de la investigación	129
4.1.1. <i>Tianjin</i>	129
4.1.2. <i>El español en China</i>	130
4.2. La muestra y la obtención de las encuestas	133
4.2.1. <i>Los centros de enseñanza seleccionados</i>	133
4.2.2. <i>Recogida de las encuestas y composición de la muestra para el análisis de la disponibilidad léxica en ELE</i>	134
4.2.3. <i>Recogida de las encuestas y composición de la muestra para el análisis de la disponibilidad léxica en CLM</i>	140
4.3. Diseño de la encuesta y procesamiento de los datos	144
4.3.1. <i>El cuestionario y la prueba</i>	144
4.3.2. <i>Procesamiento informático</i>	152
4.3.2.1. <i>Inclusión o exclusión de las encuestas cumplimentadas</i>	152
4.3.2.2. <i>Uso del programa Dispolex</i>	155
4.4. Las variables extralingüísticas	159
4.4.1. <i>Variable “nivel de estudio”</i>	162
4.4.2. <i>Variable “sexo”</i>	165
4.4.3. <i>Variable “procedencia”</i>	168
4.4.4. <i>Variable “nivel de inglés”</i>	174
4.4.5. <i>Variable “tipo de centro”</i>	180
4.5. Criterios de edición	182
4.5.1. <i>Criterios generales</i>	183
4.5.2. <i>Características específicas en los centros de interés</i>	198
5. ANÁLISIS CUANTITATIVO DE LOS DATOS	226
5.1. Resultados cuantitativos generales	226
5.1.1. <i>Índices globales en ELE y en CLM</i>	226
5.1.2. <i>Índices por centro de interés en ELE y en CLM</i>	237

5.1.2.1. Número de palabras.....	237
5.1.2.2. Promedio de palabras por informante.....	243
5.1.2.3. Número de vocablos	249
5.1.2.4. Frecuencia acumulada, índice de disponibilidad y aparición	254
5.1.2.5. Densidad léxica e índice de cohesión.....	268
5.2. Resultados cuantitativos de las variables extralingüísticas	275
5.2.1. Incidencias cuantitativas de la variable "nivel de estudio".....	276
5.2.2. Incidencias cuantitativas de la variable "sexo".....	306
5.2.3. Incidencias cuantitativas de la variable "procedencia".....	335
5.2.4. Incidencias cuantitativas de la variable "nivel de inglés".....	364
5.2.5. Incidencias cuantitativas de la variable "tipo de centro".....	391
6. ANÁLISIS CUALITATIVO DE LOS DATOS.....	421
6.1. Resultados cualitativos generales	423
6.1.1. Partes del cuerpo humano.....	426
6.1.2. Ropa y complementos.....	427
6.1.3. Partes de la casa (sin muebles).....	428
6.1.4. Los muebles de la casa.....	429
6.1.5. Alimentos y bebidas	430
6.1.6. Objetos colocados en la mesa para la comida.....	431
6.1.7. La cocina y sus utensilios.....	432
6.1.8. La escuela: sus muebles y materiales	433
6.1.9. Iluminación y calefacción.....	434
6.1.10. La ciudad	435
6.1.11. El campo	436
6.1.12. Medios de transporte.....	437
6.1.13. Trabajos del campo y del jardín	438
6.1.14. Los animales.....	439
6.1.15. Juegos y distracciones	440
6.1.16. Profesiones y oficios	441
6.2. Incidencias cualitativas de la variable "nivel de estudio"	442

6.2.1. Partes del cuerpo humano.....	444
6.2.2. Ropa y complementos.....	450
6.2.3. Partes de la casa (sin muebles).....	456
6.2.4. Los muebles de la casa.....	462
6.2.5. Alimentos y bebidas	468
6.2.6. Objetos colocados en la mesa para la comida	474
6.2.7. La cocina y sus utensilios.....	478
6.2.8. La escuela: sus muebles y materiales	484
6.2.9. Iluminación y calefacción.....	490
6.2.10. La ciudad	496
6.2.11. El campo	502
6.2.12. Medios de transporte.....	507
6.2.13. Trabajos del campo y del jardín	511
6.2.14. Los animales.....	516
6.2.15. Juegos y distracciones	522
6.2.16. Profesiones y oficios	528
7. CONCLUSIONES	535
8. BIBLIOGRAFÍA	564

APÉNDICES

	Páginas
APÉNDICE 1: EL CUESTIONARIO.....	599
APÉNDICE 2: LÉXICO DISPONIBLE GENERAL	606
2.1. Disponibilidad léxica general en ELE.....	607
2.2. Disponibilidad léxica general en CLM.....	762
APÉNDICE 3: LÉXICO DISPONIBLE SEGÚN LA VARIABLE	

"NIVEL DE ESTUDIO"	925
3.1. Disponibilidad léxica en ELE según la variable "nivel de estudio"	926
3.2. Disponibilidad léxica en CLM según la variable "nivel de estudio"	1195

1. INTRODUCCIÓN

El léxico, junto con la fonología, la morfología y la sintaxis, forma parte de los elementos esenciales de la lengua. Sin embargo, a diferencia de los demás componentes lingüísticos que presentan estructuración e integridad, el léxico parece suelto e inestable, puesto que se organiza en torno a unos campos semánticos que jamás llegan a ser finitos y que pueden ser cambiantes. Esta particularidad del léxico plantea problemas a los investigadores, quienes ven difícil hacer inventario integral de las unidades innumerables que componen el léxico para analizarlo. De ahí se deriva la aplicación de la estadística a los estudios del léxico. Proporciona una aproximación a dicho inventario de unidades léxicas, por medio de los bancos de datos que los programas informáticos son capaces de procesar en sustitución de las labores humanas.

Según López Morales (1986), la medición estadística del léxico atraviesa cuatro fases, que son, de la más antigua a la más moderna, la de los diccionarios normativos, la de las listas de frecuencia léxica, la del léxico básico y la de la disponibilidad léxica. Sintéticamente, los diccionarios normativos recogen el mayor número posible de vocablos que existen en la lengua. Las listas de frecuencia muestran las palabras con vitalidad, poniéndolas en orden según el número de veces que se repiten en un corpus cualquiera. El léxico básico soluciona la limitación de la fase anterior, recaudando el léxico de diferentes tipos de fuentes y combinando la frecuencia con la dispersión. Finalmente, la disponibilidad léxica estudia el léxico disponible de una comunidad —que seguidamente definiremos—, atendiendo a los campos semánticos y recogiendo palabras temáticas que quedan fuera del contenido del léxico básico. La presente tesis constituye un trabajo que aborda el tema del léxico desde la perspectiva de la disponibilidad léxica.

La disponibilidad léxica es una línea de investigación que surgió en Francia en los años cincuenta, cuando un equipo de investigadores intentó seleccionar del caudal léxico aquellos vocablos más habituales para

enseñarlos a los aprendientes que necesitaban adquirir el vocabulario fundamental de la lengua francesa. Los estudiosos franceses, al observar que en sus listas de palabras frecuentes abundaban aquellas de significado vacío o genérico —palabras gramaticales o adjetivos y adverbios de uso general, por ejemplo— y faltaban otras muy habituales —palabras concretas dependientes de la circunstancia comunicativa: *metro*, *auto* etc.—, propusieron una prueba asociativa de vocabulario basada en distintos centros de interés. Estos sirvieron como estímulo para que los informantes escribieran las palabras —sobre todo, sustantivos— en torno a cada temática determinada. De modo que el equipo francés pudo obtener esas palabras temáticas que denominó “palabras disponibles”. La suma del léxico básico y el léxico disponible da lugar al léxico fundamental que los investigadores se habían propuesto recolectar.

En el marco de esta línea de investigación, se han llevado a cabo gran cantidad de estudios del español como lengua materna, como segunda lengua y como lengua extranjera. La introducción de la disponibilidad léxica en el español se debe a López Morales (1973), quien recopiló el léxico disponible del área metropolitana de San Juan de Puerto Rico. Más tarde, su enfoque en la enseñanza de la lengua materna condujo a la creación del *Proyecto Panhispánico de disponibilidad léxica*, el cual tiene como objetivo elaborar diccionarios de léxico disponible del español como lengua materna para las diversas zonas del mundo hispánico. Mientras tanto, los primeros trabajos de disponibilidad léxica sobre el español como lengua extranjera se atribuyeron a Carcedo González (1998a, 1998b, 1999, 2000), quien analizó el léxico disponible de los dieciséis centros de interés empleados por el *Proyecto Panhispánico* en 350 estudiantes fineses de bachillerato y de universidad.

Nuestro trabajo sigue los métodos establecidos en el *Proyecto Panhispánico* y se adhiere a los estudios de disponibilidad léxica aplicados a la enseñanza del español como lengua extranjera (ELE). Enfocamos la cuestión desde la perspectiva de la enseñanza de ELE a los aprendices sinohablantes, puesto que va adquiriendo mucha importancia tanto en China —país de origen de los estudiantes— como en España e Hispanoamérica —

zonas con situación de inmersión lingüística—. Esta tesis contribuye a la línea de investigación en la recogida y análisis del léxico disponible de los alumnos chinos de ELE en su país de origen.

Como la inmensa extensión geográfica de China y la presencia de los centros educativos en los que se aprende ELE en casi todas las provincias del país requieren un enorme volumen de muestra, que una tesis doctoral sola no es capaz de alcanzar, optamos por tomar como sitio de muestreo una de sus provincias, que resulta ser Tianjin —municipio bajo jurisdicción central que tiene un nivel administrativo provincial—. Tianjin, como zona objeto de estudio, cuenta con una población estadística íntegra que incluye a estudiantes de ELE de todos los perfiles existentes en el país. Así pues, estadísticamente, la muestra en Tianjin puede representar el conjunto del alumnado de ELE en China. Concretamente, entre las universidades chinas que tienen un departamento de español y que, por consiguiente, pueden aportar informantes que componen la muestra, hay tanto públicas como privadas, tanto generalistas como especializadas en lenguas extranjeras, tanto prestigiosas como menos reconocidas y tanto antiguas como más recientes. Los integrantes de nuestra muestra en Tianjin disponen de todos los rasgos arriba citados, lo cual no es habitual si se toma la muestra en alguna otra provincia. De esta manera, siguiendo el modo del *Proyecto Panhispánico* de elaborar diccionarios de disponibilidad léxica para las diversas zonas del mundo hispánico, pretendemos ofrecer un modelo para la descripción del léxico disponible en estudiantes chinos de ELE para las diversas zonas de China.

La innovación de nuestro trabajo estriba en recoger y analizar el léxico disponible de los informantes tanto en español como lengua extranjera (ELE) como en chino como lengua materna (CLM), para luego comparar los resultados en las dos lenguas. De ahí viene el objetivo principal de esta tesis, que es recopilar las listas de vocablos disponibles en ELE y en CLM de los estudiantes de ELE en Tianjin. Para ello, creamos un corpus de léxico disponible para cada una de las dos lenguas en cuestión. El corpus de ELE lo constituyen 288 encuestas en español completadas por 288 informantes

universitarios procedentes de diferentes centros educativos de Tianjin durante el mes de noviembre del año 2017. El corpus de CLM se compone de otras 172 encuestas en chino completadas por 172 de los 288 informantes anteriores en el mismo mes que el de las encuestas en ELE. En ambos corpus, atribuimos a los sujetos valores de acuerdo con las siguientes variables: nivel de estudio, sexo, procedencia, nivel de inglés y tipo de centro, para poder estratificar a los informantes según dichos factores y analizar sus incidencias.

La cumplimentación de las encuestas implica dos partes: una, rellenar las informaciones personales relativas a las variables arriba comentadas y dos, realizar la prueba asociativa del vocabulario para recolectar el léxico disponible en los dieciséis centros de interés clásicos —los mismos empleados en muchos de los estudios del *Proyecto Panhispánico* y los más habituales en los estudios de disponibilidad léxica de ELE—. Los centros de interés son áreas nocionales que estimulan a los encuestados a evocar palabras relacionadas con el área. En cada centro, los informantes cuentan con dos minutos para escribir las unidades léxicas que les vienen a la mente en forma de lista, lo que facilita ver cuáles son las palabras más disponibles, es decir, cuáles son las que se les ocurren primero y ocupan las primeras posiciones de la lista. En cuanto al número de palabras requeridas, utilizamos las listas abiertas, no estableciendo un límite de palabras en cada centro de interés. El vaciado de los datos obtenidos de las encuestas se lleva a cabo en el programa informático *Dispolex*, que permite hacer cálculos como 1) el índice de disponibilidad, 2) la frecuencia relativa, 3) el porcentaje de aparición, y 4) la frecuencia acumulada. Además, también aporta índices cuantitativos como 1) el número de palabras, 2) el número de vocablos, 2) el promedio de palabras por informante, y 3) el índice de cohesión.

Al examinar y analizar los índices cuantitativos extraídos del programa, pretendemos comprobar las hipótesis y conseguir los objetivos que siguen a continuación.

Con respecto a las diferencias que se hallan entre los centros de interés,

suponemos que:

- 1) Los centros de interés, como factores lingüísticos que influyen en el léxico disponible, muestran resultados de producción léxica muy diferentes. Examinaremos cuáles son los centros más productivos y cuáles tienen una producción léxica menor.
- 2) De igual modo, la riqueza léxica también varía según el centro de interés. Intentaremos ver qué centros revelan una mayor variedad léxica y cuáles son menos ricos en vocabulario.
- 3) Los centros de interés influyen en el comportamiento léxico de los alumnos e implican distintos grados de densidad léxica —el carácter compacto o difuso de un campo léxico—. Procuraremos saber cuáles son los centros más compactos —esto es, los que muestran una homogeneidad en las respuestas por los informantes— y cuáles son los más difusos —es decir, aquellos en los que las repuestas por diferentes informantes manifiestan heterogeneidad.

En cuanto a las diferencias que se encuentran entre las dos lenguas, creemos que:

- 1) La productividad léxica de los estudiantes es menor en ELE que en CLM, tanto globalmente como por centro de interés. Intentaremos averiguar cuánto dista la primera de la segunda y si existe alguna excepción donde el resultado de ELE supera el de CLM.
- 2) Igualmente, la riqueza léxica de los informantes es menor en ELE que en CLM, tanto globalmente como por centro de interés. Trataremos de descubrir en qué grado se manifiesta la diferencia entre los resultados de las dos lenguas y si hay alguna excepción.
- 3) La producción léxica de los alumnos en CLM influye positivamente en la que se detecta en ELE: cuanto más productivo es un centro de interés en CLM, más productivo es también en ELE, y viceversa. Por lo tanto, el rango de los centros según la productividad léxica en una lengua puede resultar parecido a aquel en la otra lengua.
- 4) De modo similar, la riqueza léxica de los alumnos en ELE mantiene una correlación positiva con la riqueza en CLM: cuanto más rico es un

centro de interés en CLM, más rico es también en ELE, y viceversa. Por eso, bajo cada lengua, el rango de los centros según la riqueza léxica en una lengua se presenta más o menos como es el rango en la otra lengua.

- 5) Aunque la densidad léxica de un centro de interés en una lengua difiere de la del mismo en la otra lengua, el rango de los centros según la densidad en una lengua puede ser más o menos igual que el rango en la otra lengua, porque el carácter compacto o difuso de un centro con respecto a otro es parecido en ELE y en CLM.

En relación con las incidencias de las variables sociales que se contemplan, pensamos que:

- 1) Los distintos estratos según cualquier variable tienen sus propios centros de interés más productivos. A pesar de las diferencias, los centros más productivos de cada estrato son, más o menos, los mismos. Lo mismo sucede con los centros menos productivos. Es decir, la productividad léxica de un centro con respecto a otro se ve afectada de manera limitada por las variables.
- 2) Además de lo anterior, los distintos estratos según cualquier variable coinciden considerable o parcialmente en sus centros más productivos en ELE y en CLM. Esto también sucede con los centros menos productivos. Es decir, las variables influyen limitadamente en la productividad léxica de un centro con respecto a otro en ambas lenguas.
- 3) En cada estrato según cualquier variable, la producción léxica en ELE es inferior a aquella en CLM en general y en la mayoría de los centros de interés. Si hay excepciones, averiguaremos qué estrato en cada variable implica una ventaja de resultados de ELE sobre los de CLM.
- 4) En lo referente a la variable nivel de estudio, los estudiantes de niveles de estudio altos exhiben una mayor producción léxica que los de niveles bajos. Y, a medida que los alumnos ascienden de nivel de estudio, generan una producción léxica en ELE cada vez mayor y más aproximada a la que muestran en CLM.

- 5) En lo que se refiere a la variable sexo, no existe una superioridad notoria de la producción léxica de un género sobre el otro, aunque los chicos o las chicas pueden tener sus propios centros de interés más desarrollados. Procuraremos descubrir cuáles son los centros que suponen alguna ventaja para cada estrato según el sexo.
- 6) En cuanto a la variable procedencia, la producción léxica de los alumnos puede tener que ver con el grado de desarrollo de su zona de procedencia. Examinaremos si los resultados favorecen a alumnos que proceden de zonas cuyo grado de desarrollo es alto, medio o incluso bajo.
- 7) Con relación a la variable nivel de inglés, los estudiantes de niveles de inglés altos tienen una mayor producción léxica que los de niveles bajos. Y a medida que incrementa el nivel de inglés, los informantes generan una producción léxica en ELE cada vez mayor y más aproximada a la que manifiestan en CLM.
- 8) En lo relativo a la variable tipo de centro, los estudiantes de centros públicos muestran una mayor producción léxica que los de centros privados, dadas las características que tienen estos tipos de centros en China.

Con la finalidad de alcanzar los propósitos y confirmar o refutar las hipótesis, la presente tesis se estructura en seis capítulos principales, además de este que corresponde a la introducción. Empezamos con los presupuestos teóricos, de los que se ocupa el segundo capítulo. Explicamos de qué tratan las estadísticas y la ciencia de la estadística. Repasamos brevemente las bases históricas de la ciencia estadística y el nacimiento y uso de ella en la sociedad. Reflexionamos sobre las razones científicas que favorecen la aplicación de la estadística a las ciencias sociales y presentamos el contacto de la estadística con variadas áreas de las ciencias sociales, sobre todo, con la lingüística. También damos a conocer los métodos generales de una investigación estadística. Más adelante, nos centramos en la aplicación de la estadística a los estudios del léxico, que da lugar a la léxico-estadística. Analizamos las características del léxico que avalan su procesamiento

estadístico y revisamos la evolución de la medición estadística del léxico: primero, averiguamos los primeros intentos de análisis estadísticos sobre el léxico, después, examinamos las fases de estudio posteriores, desde las listas de frecuencia léxica hasta la elaboración del léxico básico; y, por último, nos centramos en el nacimiento de la disponibilidad léxica.

En el tercer capítulo que habla del estado de la cuestión, primero presentamos las obras pioneras de la disponibilidad léxica, aportando un resumen de sus metodologías y resultados más destacados. Son obras importantes ya que abren el camino de la línea de investigación, proporcionando el patrón del que se sirven los estudios posteriores. Concretamente, nos referimos a los trabajos de Gougenheim et al. (1964) sobre el léxico en el francés de Francia; de Pfeffer (1964), en alemán; de Ó Huallacháin (1966), en irlandés e inglés; de Dimitrijević (1969, 1981), en inglés como lengua materna y como lengua extranjera respectivamente; de Mackey et al. (1971), en el francés de Canadá; de Victory (1971), en español e inglés, y de Njock (1979), en el francés de Camerún. Luego, pasamos a los estudios de disponibilidad léxica en el mundo hispánico, repasando diversos trabajos fundamentales realizados, por un lado, en Puerto Rico, la República Dominicana, México, Chile y Costa Rica, para Hispanoamérica, y, por otro lado, en el País Vasco, Madrid, Canarias, Andalucía, Castilla y León, Castilla-La Mancha, Asturias, Valencia, Cataluña, Galicia y Aragón, para España. En estas investigaciones se van estableciendo los métodos de estudio de la disponibilidad léxica, desarrollándose el cálculo del índice de disponibilidad, los centros de interés, las pruebas asociativas, las variables, etc. Entre tanto, destacamos el papel coordinador del *Proyecto Panhispánico* de disponibilidad léxica en todos los territorios de habla hispánica. Al final de este capítulo, dedicamos un apartado a las aplicaciones de la disponibilidad léxica a diferentes ámbitos de la lingüística, donde ponemos de relieve los trabajos de excelencia llevados a cabo en el marco de la dialectología, la sociolingüística, la psicolingüística y, sobre todo, la enseñanza de la lengua.

El cuarto capítulo de la tesis es donde explicamos detalladamente la

metodología de nuestra investigación. En primer lugar, presentamos brevemente el panorama del lugar de muestreo —el municipio de Tianjin— y la situación de la enseñanza del español como lengua extranjera en China. Luego, señalamos cuáles son los centros educativos —o sea, universidades— que consideramos para ir a pasar las encuestas y qué rasgos tiene cada uno. Damos a conocer el modo de selección de los informantes que constituyen la muestra de ELE y la de CLM y exhibimos la composición final de la muestra de cada lengua. Exponemos con precisión los procedimientos de las pruebas asociativas del vocabulario en ambas lenguas. En lo referente al cuestionario, aclaramos qué partes incluye y para qué sirve cada una. Más tarde, especificamos cómo tratamos las encuestas recogidas, puesto que hemos excluido algunas que hemos dado por no válidas para el análisis posterior. Luego, indicamos cómo aprovechamos el programa informático *Dispollex* para el procesamiento de los datos recabados y alertamos de una limitación que el programa tiene para el análisis de las variables. En una sección posterior, presentamos las variables que se contemplan en la investigación. Explicamos cómo se determinan las variantes de cada variable y exponemos la estratificación de los informantes según ellas para analizar y comparar la proporción de los distintos estratos en la muestra de ELE y de CLM. Al final del capítulo, indicamos los criterios generales para la edición de los datos y los rasgos específicos que presenta cada centro de interés en lo referente a la edición.

El quinto capítulo se dedica al análisis cuantitativo de los datos. Primero, observamos y comparamos los índices globales que se muestran debajo, con el fin de analizar las diferencias entre las muestras de ELE y de CLM con relación a la productividad y riqueza léxicas del conjunto:

- 1) total de palabras,
- 2) promedio de palabras por informante,
- 3) promedio de palabras por informante y por centro,
- 4) total de vocablos,
- 5) porcentaje del total de vocablos sobre el de palabras,
- 6) total de vocablos que suman la frecuencia acumulada del 75%,

- 7) total de vocablos con índice de disponibilidad superior a 0.02,
- 8) total de vocablos con índice de disponibilidad superior a 0.1,
- 9) total de vocablos mencionados por más del 25% de los informantes,

Luego, por centro de interés, extraemos los siguientes índices de la muestra de ELE y de CLM, con la finalidad de analizar las coincidencias y diferencias entre las áreas nocionales y entre las dos lenguas con respecto a la productividad, riqueza y densidad léxicas.

- 1) número de palabras en cada centro de interés,
- 2) promedio de palabras por informante en cada centro de interés,
- 3) número de vocablos en cada centro de interés,
- 4) número de vocablos en cada centro de interés que suman la frecuencia acumulada del 75%,
- 5) número de vocablos en cada centro de interés con índice de disponibilidad superior a 0.02,
- 6) número de vocablos en cada centro de interés con índice de disponibilidad superior a 0.1,
- 7) número de vocablos en cada centro de interés mencionados por más del 25% de los informantes,
- 8) densidad léxica de cada centro de interés,
- 9) índice de cohesión de cada centro de interés.

Por último, observamos la producción léxica —el índice más fiable— de los estratos según cada variable extralingüística para informar de las incidencias cuantitativas de cada una. Sobre todo, vemos si se confirman o se refutan las hipótesis planteadas más arriba sobre cada una de las variables.

En el sexto capítulo, ya recopilados los diccionarios de disponibilidad léxica en las dos lenguas, comparamos cualitativamente los vocablos altamente disponibles en ELE y en CLM tanto en general como según la variable del nivel de estudio, con el objetivo de saber el déficit léxico de los alumnos en ELE en referencia a su conocimiento del vocabulario en CLM. Pretendemos que el análisis cualitativo sirva para, por un lado, informar de

cómo se manifiesta dicho déficit y, por otro, contribuir a compensarlo y mejorar el vocabulario disponible en ELE de los alumnos.

El séptimo capítulo es donde enumeramos las conclusiones más importantes a las que llegamos en los diversos apartados anteriores. Las partes finales son las referencias bibliográficas —que consultamos a lo largo del desarrollo de la tesis y que sostienen la elaboración de este trabajo— y los apéndices —que engloban el modelo del cuestionario y los diccionarios de disponibilidad léxica en las dos lenguas: primero, el general y, segundo, el obtenido según el nivel de estudio—.

2. PRESUPUESTOS TEÓRICOS

2.1. La ciencia estadística y sus aplicaciones

2.1.1. Las estadísticas y la estadística como ciencia

No es posible que una persona no haya oído hablar nunca de la estadística, dado que es muy frecuente la aparición del término “estadística” en nuestra vida cotidiana por sus relaciones directas o indirectas con muchas áreas de la sociedad. Generalmente esta palabra nos trae a la mente imágenes de cifras, gráficos, diagramas, tablas, etc. Realmente son abundantes las estadísticas en nuestros días, presentadas en diversas publicaciones a través de alguna de las representaciones de datos mencionadas. Estas estadísticas, en plural, son un conjunto de datos numéricos, que son recogidos, clasificados y presentados de manera sistemática y ordenada.

Mientras tanto, la estadística, en singular, es una ciencia y se ha hecho hoy en día imprescindible para los investigadores de áreas muy variadas. No hay una única definición de estadística. Por ejemplo, Vélez Ibarrola et al. la definen como “la ciencia que estudia, mediante métodos cuantitativos, las poblaciones que se obtienen como síntesis de la observación de unidades estadísticas” (Vélez Ibarrola, Ramos Méndez, Hernández Morales, Carmena Yañez, & Navarro Fernández, 2004: 8); Mingue, en cambio, afirma que la estadística es “la ciencia que tiene por objeto aplicar las leyes de la cantidad a los hechos sociales para medir su intensidad, deducir las leyes que los rigen y hacer su predicción próxima” (Ruiz Muñoz, 2004: 3); Barnett, por su parte, dice que “la estadística es la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre” (Gutiérrez Cabria, 1994: 23).

De estas definiciones y de otras similares se desprende que la estadística hace referencia a la obtención de informaciones prácticas por medio de leyes de la cantidad o métodos cuantitativos y a partir de la observación, medición,

deducción o predicción sobre unas unidades dispersas que son llamadas datos.

Lo que resulta evidente es que los investigadores utilizan la ciencia estadística para procesar, observar, analizar y estudiar los datos numéricos con el fin de describir los fenómenos o las reglas que están escondidas detrás de las frías cifras, resumiendo los resultados en unos cuadros ilustrativos que facilitan la visualización y comprensión —e incluso, en algunos casos, el pronóstico— del tema estudiado, y es frecuente que lo hagan.

2.1.2. Orígenes históricos y nacimiento de la estadística

El origen de la estadística puede datar de tiempos muy remotos, porque, desde los comienzos de la humanidad, el hombre ya siente la necesidad informativa de llevar cuenta de sus posesiones —bien de sus ganados, bien de sus pieles de animales cazados—. Este tipo primitivo de recuento y registro se debe considerar, sin lugar a dudas, la primera manifestación de la estadística en la historia humana.

Esta necesidad informativa es intrínseca del hombre. Cuanto más compleja sea la estructura social en que vive el hombre, más necesidad sentirá éste de hacer recuentos y registros para conocer mejor su situación y entorno, lo que contribuye a un buen funcionamiento del grupo social. A medida que avanza la forma de organización de los grupos sociales, las manifestaciones de la estadística pasan de los simples recuentos de las posesiones de una persona a los censos sobre la población o la riqueza de un estado.

Así que, en el antiguo Egipto, en el año 3050 a.C., los faraones lograron realizar registros relativos a la riqueza y al número de personas, con el objeto de planificar la construcción de las pirámides (Ruiz Muñoz, 2004). Después de las inundaciones del Nilo, realizaban censos sobre las tierras y sus propietarios para medir y demarcar las posesiones (Gutiérrez Cabria, 1994). Se da en el libro *Números* del *Pentateuco* del *Antiguo Testamento* otro de los primeros censos, de parte de Moisés, en el antiguo Israel, quien recopiló datos

estadísticos de la población hebrea describiendo los nombres por familias y por linaje (Ruiz Muñoz, 2004; Vélez Ibarrola et al., 2004). Los romanos, a su vez, sabían emplear bien los métodos estadísticos, con los cuales establecieron el cimiento de la organización social y política de las ciudades. Cada cinco años, hacían un censo de la fortuna de los ciudadanos¹, en el cual los cabezas de familia declaraban la calidad de las tierras, la cantidad de los ganados, la diversidad de los cultivos, etc. También anotaban a los miembros de cada familia, distinguiendo entre varones y mujeres y entre adultos y menores de edad, con el objeto de conocer el número de ciudadanos capaces de cumplir el servicio militar y el pago de impuestos (Gutiérrez Cabria, 1994; Vélez Ibarrola et al., 2004).

Durante la Edad Media se redujeron estas prácticas porque las técnicas y métodos quedaron prácticamente olvidados. Sin embargo, podemos aún citar, por suerte, unos casos excepcionales en España, Francia e Inglaterra. En la España musulmana, se realizaban registros de los jefes de familia que pagaban impuestos, al estilo romano. Además, conforme a la *Crónica del moro Rasis*, escrita por Ahmad ibn Muhammad al-Razi, Bayan Almorgerg realizó en Córdoba el primer censo de edificios y viviendas conocido de España, según el cual había en Córdoba 113,000 casas y 300 mezquitas (Gutiérrez Cabria, 1994). En Francia, los reyes Pepino el Breve y Carlomagno hicieron, respectivamente, censos sobre las tierras pertenecientes a la Iglesia en los años 758 y 762 (Gutiérrez Cabria, 1994; Ruiz Muñoz, 2004). En Inglaterra, Guillermo el Conquistador recopiló, en el año 1068, los datos de las tierras del país respecto a la propiedad, la extensión y el valor de las mismas, con la finalidad de estimar los impuestos que se podían recaudar. Los datos recogidos están registrados en el *Domesday Book*. Este es un documento bastante importante en la historia de la estadística, porque ya no se trata de un simple recuento de los objetos, sino que describe muy detalladamente las relaciones entre los propietarios, las condiciones y las valoraciones de las tierras (Ruiz Muñoz, 2004; Vélez Ibarrola et al., 2004).

¹ Recuérdese que solo los ciudadanos romanos (ciues), que constituían los estratos sociales más altos de la sociedad de Roma, eran los que tenían obligaciones fiscales.

A partir del siglo XVI, empezó en Europa la costumbre de registrar periódicamente nacimientos, fallecimientos, matrimonios y bautismos. El comienzo de esta costumbre tuvo lugar en Inglaterra, donde mandaban a registrar el número de muertes causadas por la peste debido al temor general por ésta. La iniciativa continuó en Europa y los contenidos avanzaron tan considerablemente que años después ya empezaban a realizarse inscripciones de nacimientos y fallecimientos, diferenciados por sexo, con las cuales se hicieron tablas de mortalidad por sexo de los habitantes. A partir de los métodos estadísticos en desarrollo, la tradición cristiana de Europa fomentó los registros de matrimonios y bautismos, lo que influyó notablemente en la evolución de la estadística (Ruiz Muñoz, 2004; Vélez Ibarrola et al., 2004).

En estas épocas iniciales de la estadística, la recogida de datos se caracteriza por registros o censos descriptivos con finalidades administrativas. Más tarde, a partir del siglo XVII, la estadística pudo profundizarse y surgió como una nueva ciencia gracias al alemán Hermann Conring —quien usó por primera vez el término “statistik” en un curso que impartía en la Universidad de Helmstedt—, así como a su discípulo Godofredo Achenwall —quien utilizó el mismo término en una obra publicada² y fue considerado por eso inventor de la estadística—. Estos estudiosos alemanes y sus continuadores aplicaban la metodología estadística —métodos cuantitativos de las ciencias matemáticas— en la ciencia política con el objeto de describir los fenómenos políticos, económicos y sociales del estado. De ahí consiguieron construirse los fundamentos de la estadística descriptiva, y en lo sucesivo aparecieron investigaciones innovadoras sobre la base de esta ciencia. Citamos aquí las investigaciones del inglés John Graunt, continuador de aquellos alemanes, quien, estudiando los documentos publicados sobre la mortalidad en Londres desde 1592 a 1603, intentó hacer estimaciones, e incluso pronósticos sobre las proporciones entre varones y mujer, sobre el número de personas que

² Esta obra en alemán de Godofredo Achenwall, la primera en que aparece el término “estadística”, se titula *Abriß der neuen Staatswissenschaft der vornehmen Europäischen Reiche und Republiken*, y se publicó en 1749.

morirían por enfermedades, la esperanza de vida de la población, la incidencia de cada una de las cuatro estaciones en los fallecimientos, etc. Dichas investigaciones están presentadas en su obra del año 1662 titulada *Natural and Political Observations Made upon the Bills of Mortality*. Fue un libro novedoso de análisis estadístico, ya que no sigue siendo una simple recogida de datos descriptivos, sino que se trata de análisis de informaciones para dar conclusiones relativas a la población. Fue un gran avance en la evolución de la estadística (Gutiérrez Cabria, 1994; Vélez Ibarrola et al., 2004).

Las estimaciones y pronósticos de John Graunt se entienden como conjeturas porque aún no estaban claras en aquel entonces las nociones de probabilidad. El cálculo de probabilidades tiene su origen en la afición del hombre por los juegos de azar, pero científicamente permite hacer inferencias estadísticas con mayor exactitud, lo cual constituye las bases matemáticas para sostener firmemente la ciencia estadística. Como una nueva disciplina paralela al posterior desarrollo de la estadística, la teoría de probabilidades también amplía el alcance de la aplicación de la ciencia estadística. De este modo, los avances en lo sucesivo de la estadística giran, en muchos de los casos, alrededor del cálculo de probabilidades.

2.1.3. Aplicación de la estadística en las ciencias sociales

2.1.3.1. Condiciones científicas para aplicar la estadística en las ciencias sociales

Según la clasificación de las ciencias generalmente aceptada y ampliamente utilizada, están divididas en ciencias fácticas —llamadas también ciencias empíricas— y ciencias formales. Las primeras tienen como objeto de estudio los hechos que se producen en el mundo de la realidad; mientras que las últimas estudian las formas creadas por el hombre en su mente, son abstractas y se sustentan en la razón. Las fácticas tratan de describir y explicar los hechos de la realidad, y las formales estudian y sacan conclusiones de las leyes de la lógica.

La investigación científica es una actividad con el objetivo de obtener nuevos conocimientos, en la que se aplican unos métodos capaces de resolver problemas y dar explicaciones a ciertas observaciones.

Las ciencias sociales son un subgrupo de las ciencias fácticas —o empíricas—, de modo que estudian las distintas áreas del mundo real relativo al ser humano como un ser social. Así pues, una investigación de las ciencias sociales debe partir de algún problema basado en determinadas observaciones de la realidad social del hombre y trata de hacer una descripción, interpretación, comprensión y explicación de los hechos reales a través de unos métodos empírico-analíticos.

No es fácil que los investigadores de las ciencias sociales encuentren alguna metodología empírico-analítica apta para el estudio de sus respectivas áreas, y es verdad que llevan tiempo buscándola. Se sabe que entre los diferentes métodos aplicados de la metodología empírico-analítica se cuenta el llamado método de la medición, el cual se apoya y se caracteriza por sus procedimientos estadísticos. Precisamente son los análisis estadísticos, propios de la ciencia estadística, los que van aportando instrumento metodológico para las investigaciones de las ciencias sociales, un instrumento auxiliar que con el tiempo se ha hecho imprescindible y que ha abierto nuevas perspectivas para los investigadores de las ciencias sociales.

De todas maneras, se atribuye tanta importancia a la estadística en las ciencias sociales porque las estudia y les aporta una metodología de análisis. Sus métodos estadísticos, entre otras contribuciones, permiten una representación más exacta en nuestra mente de los hechos reales de naturaleza social, posibilitan una clara interpretación de las diversas observaciones complejas mediante simples gráficos, facilitan unas descripciones más precisas a través de los datos cuantitativos, y sirven como base para extraer conclusiones convincentes utilizando procedimientos matemáticos.

2.1.3.2. Contacto de la estadística con variadas áreas de las ciencias sociales

Lambert Adolphe Jacques Quetelet, considerado como padre de la estadística moderna, fue el primero en introducir los métodos estadísticos en las ciencias sociales. Tiene plena conciencia de la complejidad de los fenómenos sociales y ve una enorme necesidad de someter a la medición la gran cantidad de variables de las ciencias sociales. Con la aplicación de la estadística en variados campos, tiene como objeto de estudio la búsqueda de leyes estadísticas que subyacen en los fenómenos observados como, por ejemplo, la tasa de criminalidad en la criminología. Con Lambert Adolphe Jacques Quetelet, se inician nuevas perspectivas de la estadística que dan lugar a la sociometría, la antropometría, la biometría, etc.

Hoy en día, no hay ninguna discrepancia en considerar que la estadística se da como una ciencia instrumental cuyo ámbito de aplicación comprende muchos de los campos de las ciencias sociales. Citamos a continuación algunos de ellos.

- En la demografía, como se ha comentado en los apartados sobre los orígenes históricos de la estadística, la descripción de las poblaciones humanas se apoya en los censos con el fin de poder analizar estadísticamente cifras de población y características humanas como nacionalidad, religión, profesión, nivel de estudios y estado civil. Además, las tasas de nacimiento y mortalidad, los cálculos de movimiento migratorio, entre muchos otros, son de índole estadística.
- En la administración pública, también como lo comentado en los apartados anteriores, la estadística es indispensable para conocer, por ejemplo, la distribución de los habitantes, el reparto de la riqueza en distintos sectores, la organización de los servicios públicos, etc. Todo esto induce a tomas de decisiones adecuadas que contribuyen

al bienestar público.

- En la economía, la estadística ha encontrado un campo sumamente fértil para su aplicación, porque, con un frecuente manejo de los datos numéricos, es natural y necesario el uso de los métodos estadísticos. Por ejemplo, se estudia la evolución de los precios por medio de números índice o los hábitos de consumo a través de encuestas destinadas a una muestra.
- En la sociología, los fenómenos y relaciones sociales se caracterizan por ser colectivos y, por lo tanto, la estadística constituye una herramienta apropiada para hacer descripciones, análisis y comparaciones de las numerosas propiedades individuales de las unidades que forman parte de los grupos sociales.
- En la psicología, las investigaciones se ocupan de los casos individuales o de grupo, lo que requiere, entre otros procedimientos, la recolección y análisis de los datos, y el diseño y evaluación de una intervención de forma cuantitativa. El dominio de la estadística es indispensable en estas operaciones.
- En la educación, la estadística proporciona métodos efectivos para hacer entender los múltiples aspectos de la macro-estructura educativa y sus interrelaciones con el sistema social general, lo que favorece, por ejemplo, el planeamiento educativo y el desarrollo de proyectos educativos.
- En las humanidades, la aplicación de la estadística es todavía reciente, pero ciertamente se han iniciado nuevas metodologías innovadoras en investigaciones de, por ejemplo, geografía, historia, literatura y lingüística, las cuales han empezado a recurrir a métodos estadísticos y han dado lugar a corrientes como la geografía, historia, literatura o lingüística cuantitativas (Pérez Juste, García Llamas, Gil

Pascual, & Galán González, 2009; Vélez Ibarrola et al., 2004).

2.1.3.3. Lingüística estadística: aplicación de la estadística en la lingüística

La lingüística estadística, llamada también lingüística cuantitativa, es la aplicación directa de la estadística en la lingüística. Utiliza métodos estadísticos para la resolución de problemas lingüísticos. Suele tener objeto de estudio la formulación de leyes lingüísticas, es decir, leyes de la lengua.

Desde etapas tempranas del estudio sobre las lenguas, ya hay una parte de la lingüística que trata de explicar la manera como las lenguas se estructuran y hacen posible la comunicación. Plantean como solución a este problema que la lengua tienen unas normas morfosintácticas para estructurar unidades y expresiones comunicativas. De ahí en adelante, las investigaciones lingüísticas en este campo se esfuerzan por detallar normas gramaticales de la lengua cada vez más estrictas y formales. Sin embargo, los investigadores también se dan cuenta de que es común que la gente flexibilice, deforme e incluso ignore las normas para satisfacer necesidades comunicativas concretas (Manning & Schütze, 2000).

Tal y como los modelos morfosintácticos dirigen la estructura de la lengua y, por consiguiente, son detectables repetidas veces en las expresiones lingüísticas, algunas —más bien muchas— de las llamadas flexibilizaciones de normas también se caracterizan por sus repeticiones inconscientes en el uso de la lengua, las cuales no pueden simplemente considerarse mal estructuradas, dado lo creativa que es la lengua y por lo común y frecuente que son ellas mismas. Así que, por un lado, están los modelos gramaticales y, por otro, los patrones usuales. Ambos forman parte de las leyes de la lengua y deben ser objeto de estudio de la lingüística, porque es cierto que el uso de la lengua es un comportamiento automático y, por tanto, las repeticiones inconscientes y casi automáticas de modelos morfosintácticos, así como de patrones usuales, tienen la misma importancia. Ahora bien, desde un punto de

vista científico, cabe preguntar cuántos hay y cuáles son. La estadística es el mejor instrumento para contar, medir y describir, por ejemplo, la frecuencia, dispersión y uso de estos modelos y patrones lingüísticos.

Todo lo dicho es aplicable a cualquiera de los niveles de la lengua, esto es, no solo a la morfosintaxis, sino también a los fonemas y sus pronunciaciones, a las unidades suprasegmentales, al vocabulario y la fraseología, así como a las unidades pragmáticas de la lengua.

La lingüística, desde una perspectiva empírica, simple, que no se centre en estudios puramente introspectivos, necesita estudiar la lengua tal y como aparece en el mundo real, pero no es capaz de hacerlo a partir de una observación exhaustiva sobre la totalidad de usos de la lengua en todo el mundo humano. Así que, asimilando la metodología de la estadística, la lingüística estadística toma una muestra concreta de textos escritos o transcritos —un corpus, por ejemplo— como sustituto de la realidad del contexto humano. Analizando la frecuencia y uso de unos modelos o patrones en determinada muestra de textos orales o escritos, la estadística ofrece un nuevo enfoque de la lingüística que permite describir cuáles son los modelos y patrones más frecuentes y usuales de las lenguas. Este nuevo enfoque recién emergente es la lingüística estadística.

No obstante, por lo emergente que es la lingüística estadística, a pesar de la importancia que va adquiriendo, su ubicación en el conjunto de las ciencias todavía queda imprecisa e, incluso, confusa en lo que se refiere a su distinción a la lingüística matemática. Los trabajos llevados a cabo en una investigación de la lingüística estadística, debido a sus métodos relacionados con unas nociones matemáticas o con otras formas de expresión basadas en las matemáticas, son incorrectamente considerados pertenecientes a la lingüística matemática. En esta tesis que utiliza métodos estadísticos, pero que tiene su eje en el estudio lingüístico, considero necesario darle a la lingüística estadística una localización adecuada que la distingue de la lingüística matemática, porque esta última no es la que abordamos. La línea divisoria

entre estas dos disciplinas reside en lo siguiente: se incluyen en la *lingüística matemática* “los trabajos que se hacen para crear modelos matemáticos del lenguaje utilizando el instrumental algebraico, el de la lógica matemática y el de la teoría de los algoritmos” (Gladkiĭ & Mel’čuk, 1972: 35); mientras que la *lingüística estadística* está alejada de los modelos matemáticos y pretende delimitar la observación sobre los datos lingüísticos a partir de los procedimientos estadísticos y dicha observación está basada en modelos de análisis lingüístico (Mateo García, 1997). Así pues, si intentamos atribuir alguna condición de semejanza a estas dos disciplinas, será que ambas forman parte de la *lingüística computacional* —llamada también *lingüística informática*—, ciencia interdisciplinaria que surge de la interrelación entre la informática y el lenguaje y “estudia las aplicaciones y aportaciones de la informática al análisis del lenguaje” (Payrató Giménez, 1998: 108).

Al fin y al cabo, en la lingüística estadística, los investigadores se fijan en las tareas propias de la lingüística, no en la matematización del lenguaje ni de la lengua y, mientras tanto, se sirven de la estadística, la cual hace posible el empleo del elemento técnico de su metodología en cuanto al procesamiento y análisis de masas de información, inaccesibles en una investigación tradicional de la lingüística (Mateo García, 1997). Es la estadística un instrumento potente al servicio de los lingüistas. Así que cuando decimos, en un principio, que la lingüística estadística trata de una aplicación de la estadística en la lingüística, nos referimos a que la lingüística es el objeto al que se aplica la estadística y no por el contrario.

2.1.4. Metodología general de una investigación estadística

Todas las investigaciones científicas comprenden tres fases iniciales que consisten en el planteamiento del problema, la propuesta de las hipótesis o preguntas de investigación y la definición de los objetivos. Una vez superadas estas tres etapas de primer acercamiento a la investigación, se debe pensar en una metodología adecuada —o más de una, si los objetivos perseguidos dentro de una misma investigación son distintos— que guíe todo el proceso de

la misma.

La ciencia estadística tiene una metodología fundamental y ampliamente aceptada para muchas áreas de la investigación científica, tal y como se ha comentado en los apartados anteriores. El primer paso de la metodología en un estudio de estadística aplicada es determinar la población y las unidades de la investigación en cuestión. La población es el conjunto de elementos — seres u objetos— de características comunes de los que se quiere conseguir cierta información; mientras una unidad de estudio es cada uno de los elementos de esa misma población.

Conforme a cómo delimitan y tratan las unidades de estudio, las investigaciones estadísticas pueden ser *de observación*, *de experimento* o *de muestreo*. En las investigaciones de observación, el investigador observa directamente un conjunto de unidades dadas. En las investigaciones de experimento, el investigador divide las unidades en subgrupos y asigna diferentes tratamientos a cada uno de los subgrupos separados. Por último, en las investigaciones de muestreo, el investigador selecciona cierta cantidad de unidades de la población objetivo mediante algún procedimiento, las cuales constituyen una muestra del estudio. Están bien vistas tanto la elección de uno de los tres tipos de método como la aplicación combinada de dos de ellos o incluso de los tres. Dado que la población y las unidades en una investigación estadística suelen ser inmensas y, a veces, infinitas, es bastante recomendable y utilizado el método de muestreo, porque de una muestra representativa se estudian las unidades que permiten generalizar lo que caracteriza a toda la población objetivo.

Una vez determinadas las unidades por estudiar, hay que tener en cuenta qué propiedades o características de las unidades son de interés para cada estudio. Las propiedades y características que fluctúan y adoptan distintos valores cuya observación y medición es posible con la estadística son llamadas *variables*. Conforme a la medición, existen dos tipos de escala en las variables: las *cuantitativas*, que toman valores numéricos; y las *cualitativas*, que toman

valores de categoría. El siguiente paso que se debe seguir es obtener y recopilar datos mediante la medición de las unidades según diferentes variables. Así que los datos se entienden también como resultados de la medición sobre las variables de las unidades.

Cuando ya están los datos recopilados, se debe emprender la sistematización y el registro de los mismos, sometiéndolos a algún proceso de organización, corrigiendo errores de codificación, etc., para que sean informaciones fiables y accesibles. Es imprescindible una sistematización rigurosa de los datos obtenidos a fin de facilitar el análisis de los mismos. Hoy en día casi no es posible el análisis de los datos sin el apoyo de algún programa estadístico computacional. Así que, en la actualidad, se puede servir de la tecnología informática y disponer de un conjunto de medios de procesamiento que se aplican en los datos por analizar. En esta fase, por ejemplo, se puede hacer análisis de forma descriptiva recurriendo a procedimientos de recuento, así como a elaboración de tablas de frecuencias, porcentajes, etc. Posteriormente, de la interpretación de los análisis y del examen de los resultados se sacan conclusiones de la investigación estadística en cuestión.

También son posibles los análisis probabilísticos, que usan la información que aportan los datos recopilados para sacar conclusiones generalizables, esto es, para predecir el comportamiento de cada una de las unidades en que se manifiesta la variable.

2.2. Léxico-estadística: estudios estadísticos sobre el léxico

2.2.1. La noción de la léxico-estadística

2.2.1.1. Características del léxico al encuentro de la estadística

La lingüística del pasado, sobre todo, las primeras fases de su aplicación a la enseñanza y aprendizaje de lenguas, llevaba largo tiempo centrada en aspectos de la lengua como la gramática y la fonética, olvidando y dejando

aparte el léxico. Tal circunstancia se debía a que bastantes estudiosos de aquella época —en su mayoría los de métodos de enseñanza de lenguas— tenían el concepto de que las lenguas se estructuraban sobre la base de la gramática y que el sistema del léxico organizado en torno a campos semánticos —o incluso sencillamente al azar— era mucho menos regularizado que el de la gramática y la fonética (Terrádez Gurrea, 2001).

De hecho, el léxico da la impresión de obedecer muy poco a la estructuración y es distinto y más complejo que casi todos los demás elementos de la lengua. Mientras la fonología y la morfosintaxis tienen series constituidas por unidades de estudio finitas que permiten hacer inventarios completos de las mismas, el léxico de una lengua parece tener elementos que nunca llegan a ser estrictamente finitos. Por otra parte, es habitual pensar que el léxico es muy inestable en el sentido de que suele variar de un hablante a otro y, aun en el mismo hablante, de un momento a otro, lo que resulta poco común en la fonología y la morfosintaxis. Es estos niveles de la lengua, en cambio, los fenómenos de estudio en un período de tiempo de observación determinado muestran una estabilidad notable (Müller, 1973). Estas particularidades que manifiesta el léxico obstaculizan su estudio en una investigación de lingüística tradicional. Así es por qué la lingüística aplicada a la enseñanza y aprendizaje de lenguas llega mucho más tarde y difícilmente al léxico en comparación con el modo como estudia otros aspectos de la lengua.

Por lo tanto, ha habido una necesidad de diseñar un nuevo método en las investigaciones sobre el léxico que superara su carácter rebelde —infinidad e inestabilidad de sus unidades de estudio—. Afortunadamente, la aplicación de la estadística como herramienta en los estudios del léxico —conocida como *léxico-estadística* o *lexicometría*— ha aportado significativas novedades. Estos estudios del léxico a partir de su medición, o sea los trabajos cuantitativos sobre el léxico, hacen posible análisis estadísticos que llevan enormes ventajas sobre los métodos clásicos. En realidad, desde el inicio de la introducción de la estadística en la lingüística, el estudio sobre léxico, aprovechando el empleo de ese instrumento estadístico, ha tenido una

ponderable porción de presencia en la totalidad de investigaciones pertenecientes a las variadas ramas o subdivisiones de la lingüística, es decir, el nivel de la lengua que la lingüística estadística ha atendido de preferencia es el léxico. Esta preferencia por el léxico es comprensible, porque las unidades que componen el léxico —las palabras—, como elementos de estudio, son muy fáciles de obtener en cuanto a su extracción y delimitación dentro de un conjunto de texto escritos u orales. Asimismo, el ordenador, de manera automática, distingue y reconoce mejor las palabras —elementos que se quedan naturalmente entre dos espacios en blanco— que los demás elementos lingüísticos como fonemas, morfemas, sintagmas, oraciones, etc. Se puede afirmar que, a partir de la estadística, el léxico ha ido ganando la atención que le faltaba en el pasado.

2.2.1.2. Evolución de la medición estadística del léxico

Los estudios de léxico-estadística “se ocupan de realizar cálculos léxicos mediante procedimientos estadísticos; dicho de otro modo, tratan de establecer, por medio de diferentes métodos, índices estadísticos para las distintas unidades léxicas de una lengua” (Serrano Zapata, 2014: 39). A juicio de López Morales (1986), la medición estadística del léxico ha experimentado cuatro fases, las cuales son, de más antigua a más moderna, los diccionarios normativos, las listas de frecuencia léxica, el léxico básico y la disponibilidad léxica. Los diccionarios normativos, primeros trabajos sobre el léxico, tienen el empeño de hacer lo posible para recoger el mayor número de vocablos existentes en una lengua. Las listas de frecuencia léxica ponen las palabras en orden según su respectivo número de veces que se repiten en un corpus, tratando de mostrar las palabras con vitalidad. El léxico básico, repartiendo el léxico en diversos tipos de fuentes y combinando la frecuencia con la dispersión, soluciona las limitaciones de las listas de frecuencia, en las que se calcula la frecuencia a partir de un corpus unitario e indiscriminado. La disponibilidad léxica, finalmente, pone atención en aquellas palabras muy habituales y temáticas, pero que quedan fuera de las listas de léxico básico porque su presencia es limitada cuando se recogen palabras sin atender al

tema del texto.

Creemos que los verdaderos intentos estadísticos de la lexicometría empiezan a darse solo con la aparición de los índices de frecuencia léxica, porque es de aquí cuando comienzan los trabajos laboriosos que deben basarse en métodos estadísticos utilizando un corpus muy extenso. A continuación, respecto a la evolución de la lexicometría nos fijamos solo en las últimas tres fases citadas por López Morales.

Si observamos estas tres fases de la medición estadística del léxico, veremos que todos estos trabajos de recuento léxico —estudios sobre el léxico frecuente, básico y disponible— tienen en un principio una utilidad muy práctica en lo relativo a la pedagogía. En función de la frecuencia, dispersión o disponibilidad de las palabras en un corpus representativo y, generalmente, voluminoso, se puede obtener una información objetiva de las mismas sobre, por ejemplo, cuáles de ellas se repiten o se usan más que otras y, por consiguiente, se puede conocer la importancia de cada una de ellas. De esta manera, se supone que en la enseñanza de una lengua se puede poseer fundamentos al decidir qué palabras deben enseñarse en cada nivel de estudio y al determinar cuáles de ellas son las que deben enseñarse para que los alumnos las conozcan y cuáles para que las dominen. Entre los interesados por la elaboración de materiales de enseñanza, se piensa que los estudios de recuento léxico pueden servir para juzgar el grado de dificultad que tienen los materiales de aprendizaje y para seleccionar el léxico adecuado que debe aparecer en los mismos. Aparte de las simples palabras, los resultados estadísticos también posibilitan la selección de los grupos de palabras —unidades fraseológicas, por ejemplo— que deben enseñarse en combinación.

Aunque surgen de cara a la enseñanza de lenguas, los estudios estadísticos del léxico han sido intentos de conocer cuál es el vocabulario usual de una lengua. No disponemos de muchos medios para obtener las palabras de uso frecuente hasta que se recurre a la léxico-estadística, ya que los primitivos intentos de medición estadística, esto es, los diccionarios normativos

—método convencional para la recogida del vocabulario existente en una determinada comunidad de habla—, tratan de incluir, lo más exhaustivamente posible, todo el caudal léxico de una lengua (López Morales, 1986). Dicho de otra manera, en un diccionario normativo, todos los vocablos y acepciones recogidos aparecen por igual, sin establecer diferencias entre una palabra y otra ni subrayar las de uso frecuente que debe manejar un hablante común. Recogen vocablos poco conocidos, algunos conocidos pero que no se utilizan en las conversaciones cotidianas, u otros de cuyo significado suele dudar un hablante común. También están presentes los arcaísmos, regionalismos, tecnicismos, cultismos, etc., que están muy distantes de las palabras habituales. Así que, a pesar de la enorme riqueza de contenido que tienen los diccionarios normativos, no nos brindan informaciones útiles sobre el vocabulario usual ni nos sirven para la enseñanza del léxico. Este es otro factor que distingue los diccionarios normativos de los estudios sobre el léxico frecuente, básico y disponible que serán tratados a continuación.

2.2.2. De las listas de frecuencia léxica al léxico básico

La frecuencia, en lingüística, es el número de apariciones de una unidad lingüística en un conjunto amplio de enunciados escritos u orales. Como se ha comentado en el apartado anterior, con la finalidad de descubrir cuáles son las palabras usuales y de facilitar la selección del léxico adecuado en la enseñanza de lenguas, empiezan las confecciones de listas de frecuencia léxica basadas en corpus. El conocimiento de éstas sirve para repasar algunos de los métodos a menudo usados en la léxico-estadística. En la década de los años 20, los estudios estadísticos sobre el léxico ya llegaron a las lenguas más analizadas por la lingüística —nos referimos a continuación al alemán, el español, el inglés y el francés—.

Los orígenes de las investigaciones estadísticas léxicas se suelen atribuir a Friedrich Wilhelm Kaeding, quien, en 1898, aplicó por primera vez métodos estadísticos de frecuencia de palabras en la selección del vocabulario usual del alemán. Con ello elaboró un diccionario de frecuencias de la lengua

alemana —titulado *Häufigkeitwörterbuch der deutschen Sprache* ('Diccionario de frecuencia de la lengua alemana')— basado en materiales escritos como libros de historia y literatura, textos comerciales, de medicina y teológicos, debates parlamentarios, diarios y revistas (Mateo García, 1997; Pfeffer, 1964).

Las recopilaciones de esta índole para el español se remontan a 1920, año en que Hayward Keniston sacó a la luz *Common Words in Spanish*, obra destinada a la enseñanza del español a extranjeros, la cual incluía palabras ordenadas por frecuencia extraídas de cuentos, novelas, obras de teatro, periódicos, revistas, etc. Además, de esta primera obra, el mismo autor sacó una segunda versión en el año 1929, titulada *Spanish Idiom List* (Ávila Martín, 2010; Mateo García, 1998).

Un año después de *Common Words in Spanish*, es decir, en 1921, Edward Lee Thorndike recopiló *The Teacher's Word Book* para el inglés. Era un listado por orden alfabético de palabras de alta frecuencia tomadas de una amplia gama de textos, que incluían cuarenta y una fuentes que variaban desde obras clásicas inglesas, literatura para niños o libros de texto de primaria, hasta periódicos, correspondencia epistolar, libros de cocina, costura, agronomía, comercio, etc. Cabe destacar que Thorndike fue el primero en introducir, metodológicamente, el *rango* para combinarlo con la *frecuencia* en el cálculo de los resultados, aportación relevante que orientó las investigaciones posteriores. La noción de rango —traducido del *range* en inglés— refiere la presencia de una palabra en las diversas fuentes del corpus, o, lo que viene a ser lo mismo, el número de fuentes en que aparece una palabra. Por lo tanto, mientras la *frecuencia* indica cuán a menudo se usa la palabra, el *rango* señala con qué amplitud es utilizada (Pfeffer, 1964: 2). La combinación de ambos hizo que los resultados fueran más fiables.

La fórmula empleada en *The Teacher's Word Book*, que resultó de la combinación de la frecuencia y el rango, fue la siguiente:

$$\frac{\textit{Frecuencia}}{10} + \textit{Rango}.$$

La había diseñado Thorndike junto con otros colegas, entre los cuales estaba Vivian Allen Charles Henmon. Fue precisamente éste, quien, un poco más tarde, en el año 1924, publicó en los Estados Unidos *A French Word Book Based on a Count of 400.000 Running Words*, listado de frecuencia de palabras que marcó el primer intento de esta índole centrado en la lengua francesa. De las 400.000 palabras que entraron en el corpus del investigador, 80.000 provenían de materiales editados para uso en escuelas o universidades, 20.000 de prosa filosófica, 50.000 de artículos de revista, 30.000 de artículos de periódico, 140.000 de obras de escritores franceses modernos, 15.000 de poesía, 25.000 de ensayos políticos y militares, 15.000 de cartas y telegramas, y, al final, 25.000 de literatura de ficción popular. Como primer paso, este investigador redujo el corpus entero a 9.187 palabras, de las cuales publicó luego 3.905. Los 3.905 palabras finales estaban organizadas en tres niveles: el primero contenían 650 palabras cuya frecuencia superaba el 50; el segundo nivel lo constituían 1.250 palabras de frecuencias comprendidas entre 25 y 50; y, en el último nivel, estaban las palabras cuyas frecuencias oscilaban entre 5 y 25 (Pfeffer, 1964).

En la misma década, en el año 1927, salió a la luz, en Toronto (Canadá), el segundo enfoque de la lexicometría sobre la lengua española, *A Graded Spanish Word Book*, de Milton Alexander Buchanan. Esta publicación proporcionó una lista de 18.331 palabras diferentes, sobre la base de un conjunto de 1.200.000 palabras, provenientes de cuarenta textos pertenecientes a siete ámbitos temáticos: teatro, novela, poesía, folklore, prosa, textos técnicos y científicos y publicaciones periódicas. De las 18.331 palabras de la lista, Buchanan designó 5.324 con frecuencia mayor de 10 como esencia del español. Y de estas, estableció una lista de las primeras 2.000 palabras más importantes probando tres tipos de cálculo: 1) una prueba de solamente *frecuencia*, 2) una de solamente *rango* y 3) otra en que combinó los dos y cuyo resultado le pareció más fiable (Pfeffer, 1964: 4). Así es como, por primera vez en el ámbito del español, se combinaron, en la clasificación de las palabras obtenidas, la *frecuencia* con el *rango*; aunque el efecto del rango en esta obra sólo era notorio en las palabras de baja frecuencia, mientras en las de alta

frecuencia el rango resultó un factor casi nulo (Ávila Martín, 2010; Ezquerro, 1974; Mateo García, 1998).

Otra obra importante que aporta avances metodológicos es la publicada por George Emil Vander Beke en los Estados Unidos en 1932. Era la segunda investigación de nuestra disciplina que se hizo sobre el francés, con el título de *French Word Book*. El autor siguió los métodos generales de las publicaciones anteriores, pero expandió el volumen del corpus hasta recopilar 1.147.748 palabras corrientes. Estas fueron extraídas de novelas e historias escritas por una serie de autores franceses eminentes como Alexandre Dumas, Marcel Proust y sus contemporáneos, de publicaciones periódicas de los años 1925 y 1926, de las biografías, la crítica, la filosofía y la ciencia del primer cuarto del siglo XX, etc., excluyendo, a su vez, la poesía. El número final de palabras seleccionadas en el libro fue 19.253, las cuales, a diferencia de las publicaciones anteriores, fueron ordenadas en función de su *rango* en lugar de su *frecuencia*. Así ordenadas las palabras, se puso de manifiesto que, del total de 19.253 palabras, solo 6.067 tenían un *rango* que superaba 5 (Pfeffer, 1964: 3). Además de ponderar la combinación de la noción de *rango* con la de *frecuencia*, consideraba Vander Beke que una palabra que aparecía una vez en diversos textos tenía más peso que otra que aparecía varias veces en un mismo texto (Mateo García, 1998).

La siguiente contribución importante a este tipo de estudio sobre la lengua española después de *A Graded Spanish Word Book* se dio en 1952 en Puerto Rico. En ese año, Ismael Rodríguez Bou publicó el *Recuento del Vocabulario Español*, patrocinado por el Consejo Superior de Enseñanza de Puerto Rico. El autor utilizó el corpus más voluminoso hasta ese momento, el cual incluía 7.066.637 palabras. Rodríguez Bou estaba convencido de que el corpus limitado a 1.200.000 palabras que había utilizado Milton Alexander Buchanan para *A Graded Spanish Word Book* introdujo limitaciones, y se propuso superarlas ofreciendo más garantía con un recuento a partir de un corpus mucho más extenso. Para ello, Ismael Rodríguez Bou se sirvió del corpus de Milton Alexander Buchanan y empleó otras fuentes escritas y orales

muy amplias y de características innovadoras.

Destaca el uso de fuentes no solamente literarias, por lo que, además de literatura religiosa y textos escolares, Rodríguez Bou emplea periódicos, revistas y, novedosamente, composiciones escritas por alumnos portorriqueños. Como otras fuentes no escritas, el autor recogió una especie de vocabulario oral, esto es, palabras procedentes de textos orales y conversaciones transcritas, algunas de las cuales provienen de programas de radio. Por último, llama la atención que el autor, además de dichas fuentes, aprovechó, en la obtención del vocabulario, unas pruebas de asociación controlada y libre —técnicas cercanas a las utilizadas en la metodología de la disponibilidad léxica—. Las pruebas de asociación controlada, a su vez, fueron efectuadas a través de noventa palabras de estímulo, en función de las cuales los alumnos debían escribir cuantas palabras se les ocurrieran. A estas pruebas añadió otras de asociación libre como complemento, en las que exigió a los alumnos escribir cuantas palabras se les ocurrieran sin ningún estímulo, con la finalidad de evitar la limitación de la prueba anterior (Ávila Martín, 2010; Ezquerro, 1974; Serrano Zapata, 2014).

Un año después del *Recuento del Vocabulario Español*, en 1953, Víctor García Hoz dio a luz el *Vocabulario usual, común y fundamental*, otro enfoque de la disciplina sobre el español. A diferencia de Rodríguez Bou, Hoz utilizó un corpus restringido de 400.000 palabras. Él se apoyó en técnicas estadísticas más desarrolladas que las utilizadas en las investigaciones anteriores y aseguró, justificando el volumen reducido de su corpus, que el recuento de 400.000 unidades era suficiente para dar con el vocabulario de base que un adulto medio debía dominar. Formaban parte de su corpus cartas privadas procedentes de familias de clase modesta de la sociedad —con las cuales intentaba revelar las palabras de expresión espontánea de la vida familiar—; periódicos, publicaciones oficiales del estado, eclesiásticas y de organizaciones sindicales —con los que pensaba reflejar el vocabulario de la vida social—; y los libros más vendidos —como fuentes culturales y literarias—.

Una novedad interesante es que Víctor García Hoz acuñó, en esta obra suya, los conceptos de *léxico activo* y *léxico latente*. Le asignó al primero el significado de la totalidad de palabras que una persona actualizaba por lo común en su habla diaria y en su escritura espontánea; y por el último entendió aquellas palabras que una persona no usaba habitualmente pero que comprendía al oírlas o leerlas. El *vocabulario usual*, en opinión del autor, era la suma del *léxico activo* y el *latente* (Ávila Martín, 2010; Ezquerro, 1974). Estadísticamente, las 12.911 palabras cuyas frecuencias oscilaban entre 36.999 y 1, las calificó de *vocabulario usual*; de estas, seleccionó 1.971 que aparecían por lo menos una vez en cada tipo de fuente y las denominó *vocabulario común*; las palabras dentro de este último con una frecuencia mínima de 40 fueron designadas como *vocabulario fundamental* (Pfeffer, 1964).

A lo largo de esos años y por medio de estos trabajos y otros siguientes, se iba refinando la metodología en lo que se refiere al criterio renovado de selección de fuentes de texto, a la confección de un corpus de léxico más diversificado y representativo, al empleo de fórmulas estadísticas más evolucionadas, etc. Sin embargo, los problemas más delicados relativos a la frecuencia de una palabra comenzaron a vislumbrarse de todos modos. Se esperaba que las palabras más frecuentes también fueran las más usuales o habituales, pero, por un lado, las palabras de muy alta frecuencia eran realmente pocas y, por otro, los resultados no llegaron a coincidir con las observaciones empíricas sobre el discurso comunicativo, lo que condujo a la idea de que la importancia comunicativa de una palabra no dependía solo de su frecuencia, sino que también debían considerarse otros índices.

Por eso, en el año 1964, Alphonse Juilland y Eugenio Chang Rodríguez hicieron público el *Frequency Dictionary of Spanish Words*, mejorando considerablemente la metodología estadística. Su corpus se componía de 500.000 palabras provenientes de 25.000 oraciones. Los autores dieron razón al hecho de trabajar a partir de un corpus reducido comentando que datos adicionales dejaban de afectar a las propiedades matemáticas de los tipos estadísticos, esto es, los cinco tipos léxicos utilizados por los autores en los

que se repartían estadísticamente a partes iguales las 500.000 unidades. Estos cinco ámbitos léxicos —cada uno de los cuales contenía por igual 100.000 palabras— son 1) obras de teatro; 2) literatura de ficción, novelas y cuentos cortos; 3) prosa, ensayos y cartas; 4) textos periodísticos; y 5) textos técnicos y científicos. De esta manera bastante novedosa, además de calcular la frecuencia de una palabra en el corpus entero como conjunto, pudieron obtener también la frecuencia de una palabra en cada uno de los ámbitos léxicos e hicieron posible, por consiguiente, la comparación de la importancia de una palabra entre un ámbito y otro.

Se puede notar que el corpus en cuestión estaba compuesto exclusivamente por fuentes escritas y, además, solo una quinta parte de ellas la constituían textos no literarios —textos periodísticos—. A esta característica de las fuentes se le suma otra, que era la homogeneidad geográfica y temporal de las mismas: según los autores, solo fueron admitidos para formar parte del corpus textos del español peninsular, excluyendo producciones provenientes de Hispanoamérica o españolas de naturaleza dialectal, con la finalidad de garantizar la uniformidad espacial; además, al objeto de asegurar la homogeneidad cronológica, solo fueron aceptados textos producidos entre las dos guerras mundiales —entre el año 1920 y 1940, cuando se realizó al fin la recopilación de las fuentes—. Por último, los autores también intentaban controlar las variaciones socioculturales y estilísticas, solo reconociendo el vocabulario adulto estándar. Debido a estas características del corpus del *Frequency Dictionary of Spanish Words*, parece más propicio para estudios filológicos que didácticos y no es una investigación que tenga por objetivo la enseñanza del español, toda vez que la pretensión de la obra es simplemente averiguar cuál es el vocabulario que conoce un hablante medio, esto es, el léxico básico, concepto diferente y más evolucionado que el léxico frecuente, siendo este último el conjunto de palabras que se repiten más y que usamos constantemente.

Ahora pasamos a la aportación estadísticamente más importante en esta obra de Alphonse Juilland y Eugenio Chang Rodríguez. Ellos buscaban un

coeficiente que pudiera decidir la aceptación o rechazo de un vocablo en el léxico básico y finalmente llegaron a recurrir a tres parámetros estadísticos: *índice de frecuencia*, *índice de dispersión* e *índice de uso*. En una primera aproximación, los autores estudiaron la cantidad de textos en los que aparecía una palabra. Es una noción que hemos mencionado ya en los apartados anteriores: el *rango*, que se estrenó por Edward Lee Thorndike en *The Teacher's Word Book*, que introdujo Milton Alexander Buchanan en *A Graded Spanish Word Book* y que preconizó George Emil Vander Beke en *French Word Book*. A este concepto le llamaban Alphonse Juilland y Eugenio Chang Rodríguez *dispersión simple* y lo estudiaron luego más profundamente porque se dieron cuenta de que no diferenciaba dos palabras que tenían la misma frecuencia y que presentaban un mismo número de fuentes en las que aparecían; por ejemplo, en un corpus de 2 tipos de fuente, está claro que no es lo mismo, en cuanto a la importancia, una palabra que se repite 100 veces en un tipo de fuente y otras 100 veces en el otro tipo, por un lado, que, por el otro, otra palabra que se repite 199 veces en un tipo y solo 1 vez en el otro, aunque el *rango* de ambos casos es 2. Así, para equilibrar la frecuencia y relacionarla con los universos temáticos, los autores propusieron la *dispersión compleja*, midiendo la estabilidad de la frecuencia de una palabra en los cinco ámbitos léxicos utilizados; por ejemplo, si una palabra presentaba la misma frecuencia en cada uno de los cinco ámbitos, se deducía que su uso era muy estable en diversos universos temáticos. De esta manera, no solo tomaban en cuenta el número de textos en que aparecía la palabra, sino que también consideraban el modo como se repartía la frecuencia de esta. Sobre la base de la *dispersión compleja*, los autores produjeron un *índice de dispersión* y, multiplicando el *índice de frecuencia* por el de *dispersión*, obtuvieron el *índice de uso*, indicador en el que se basó la lista del *Frequency Dictionary of Spanish Words*. Al final, en esta obra se obtuvo el vocabulario usual —o, lo que viene a ser lo mismo, el vocabulario básico—, que incluía palabras frecuentes y estables y que resultaba de una medida más refinada que la del vocabulario meramente frecuente.

2.2.3. Nacimiento de la disponibilidad léxica

Las medidas de frecuencia léxica también eran utilizadas en Francia a mediados de los años cincuenta, cuando un equipo de investigadores intentaba diseñar algún manual de francés cuyo objetivo era facilitar el aprendizaje de dicha lengua a gran cantidad de inmigrantes que llegaban a Francia y, asimismo, a numerosos habitantes de varios países africanos que antes habían compuesto la Unión Francesa, donde el francés seguía siendo lengua oficial o lengua más influyente. Formaban parte en este equipo Georges Gougenheim, René Michéa, Paul Rivenc y Aurélien Sauvageot, quienes publicaron, en el año 1954, *Le français élémentaire*. Este manual es un material de enseñanza de este idioma a extranjeros, en el cual se propone presentar un francés de base, reducido a lo esencial del léxico y de la gramática y, así simplificado, útil para una adquisición rápida del idioma. En el área del léxico, la selección de los vocablos se basó exclusivamente en el criterio de la frecuencia, criterio comúnmente utilizado en aquel entonces, aun antes de la toma en consideración de la elaboración del léxico básico iniciado por Alphonse Juilland y Eugenio Chang Rodríguez. En el transcurso de los siguientes años, los autores continuaron refinando la obra original y le hicieron renovaciones y reediciones, sacando a la luz, respectivamente, en 1956, *L'élaboration du français élémentaire* y, en 1958, *Français fondamental (1^{er} degré)*.

Mientras avanzaban en esas continuas investigaciones, los investigadores franceses, al revisar el listado de las palabras obtenidas, se dieron cuenta de la deficiencia y parcialidad del criterio de frecuencia. Notaron que muchas palabras habitualmente utilizadas por un hablante francés no entraban en el listado de frecuencia, algunas de ellas figuraban en lugares muy posteriores y las palabras gramaticales predominaban en el listado ocupando casi la mitad del total de palabras obtenidas. Se hizo evidente que las palabras de alta frecuencia eran aquellas sin significado o de significado genérico (palabras gramaticales o verbos, adjetivos y adverbios de uso general); por otro lado, las palabras de significado concreto (enorme cantidad de sustantivos

que hacen referencia a realidades del universo) no formaban parte del listado de palabras frecuentes o, aunque aparecían en él, eran finalmente excluidas del léxico finalmente seleccionado por presentar muy baja frecuencia. La deficiencia y parcialidad de la medida de frecuencia se explica si nos fijamos en que el resultado está condicionado por la temática de los textos seleccionados para constituir el corpus, a partir del cual se calculan las frecuencias. Es decir, algunas palabras muy usuales no son en absoluto frecuentes en el marco del corpus tratado, puesto que éste no alberga todos los temas por muy extenso que sea. De ahí que René Michéa planteara la necesidad de diferenciar *palabras atemáticas* de *palabras temáticas*. Las primeras son aquellas que aparecen comúnmente en cualquier texto sin que importe el asunto de que se trata. Forman parte de esta categoría las palabras gramaticales, numerosos adjetivos y adverbios genéricos, bastantes verbos comunes y algunos nombres generales. Por otra parte, las palabras temáticas sólo se actualizan en el discurso cuando existe un estímulo de algún tema dado, ya que señalan seres u objetos y son palabras concretas. Por lo tanto, el criterio de la frecuencia dentro de un corpus dado no favorece la selección de las palabras temáticas, lo cual condujo a los investigadores franceses a pensar que no debían depender sólo del método de frecuencia y tenían que buscar otra medida complementaria.

René Michéa inició sus primeros experimentos partiendo de la idea de ofrecer estímulos temáticos. Leyó un fragmento literario de unas ochocientas palabras —extraído de *El Tonelero de Nuremberg*, de Hoffmann— ante diez alumnos y después les dijo que escribieran las palabras que les habían resultado de interés al escuchar el texto. Este primer experimento salió exitoso, dado que los alumnos anotaron palabras esperadas en vez de palabras gramaticales o verbos de significado genérico. Más tarde, Georges Gougenheim y René Michéa realizaron, cada uno, otro tipo de prueba. Pidieron a los alumnos que imaginaran un viaje en tren a partir de la llegada a la estación ferroviaria y escribieran luego las primeras veinte palabras que se les

ocurrieran con respecto a dicho viaje³. Ambos investigadores sacaron resultados favorables en cada una de sus pruebas, siendo sustantivos la mayoría de las palabras anotadas por los alumnos (López Morales, 1999b). Después de los primeros intentos, René Michéa tuvo la solución definitiva y se decidió por trabajar con pruebas asociativas. Así, creó una prueba asociativa basada en unos “centros de interés”, los cuales consistían en unos campos temáticos concretos que servían como estímulo para que los alumnos escribieran en una lista las palabras que les vinieran a la mente en torno a cada uno de ellos. A continuación, se presentan los dieciséis centros de interés empleados por los investigadores franceses:

- 1. *Les parties du corps / Partes del cuerpo*
- 2. *Les vêtements (peu importe que ce soient des vêtements d'homme ou de femme) / La ropa (no importa que sea de hombres o de mujeres)*
- 3. *La maison (mais pas les meubles) / La casa (pero sin los muebles)*
- 4. *Les meubles de la maison / Muebles de la casa*
- 5. *Les aliments et boissons des repas (à tous les repas de la journée) / Alimentos y bebidas (en todas las comidas del día)*
- 6. *Les objets placés sur la table et dont on se sert à tous les repas de la journée / Objetos colocados en la mesa y utilizados para todas las comidas del día*
- 7. *La cuisine, ses meubles et les ustensiles qui s'y trouvent / La cocina, sus muebles y los utensilios que se encuentran en ella*
- 8. *L'école, ses meubles et son matériel scolaire / La escuela, sus muebles y el material escolar*
- 9. *Le chauffage et l'éclairage / Calefacción e iluminación*
- 10. *La ville / La ciudad*
- 11. *Le village ou le bourg / El campo o el pueblo*
- 12. *Les moyens de transport / Medios de transporte*
- 13. *Les travaux des champs et du jardinage / Trabajos del campo y*

³ La instrucción original en francés que dieron a los alumnos es lo siguiente: *Pensez à un voyage en chemin de fer à partir du moment où vous vous trouvez à la gare et écrivez les 20 mots qui vous viennent les premiers à l'esprit* (Gougenheim, Michéa, Rivenc, & Sauvageot, 1964: 147).

del jardín

- *14. Les animaux / Animales*
- *15. Les jeux et distractions / Juegos y distracciones*
- *16. Les métiers (les différents métiers et non pas les noms qui se rapportent à un seul métier) / Oficios (los diferentes oficios y no los nombres que se refieren a un solo oficio)*

Estos centros de interés estaban pensados, por un lado, para ser universales, de modo que pudieran despertar asociaciones en cualquier tipo de encuestado, tuviera éste la condición que tuviera; y, por otro lado, para representar, lo más completamente posible, diversos compartimientos del mundo semántico bien diferenciados entre sí —aunque no es fácil abarcar todos los ámbitos del universo y, por tanto, queda margen para el desarrollo del contenido de los centros de interés—. Así las pruebas fueron realizadas en 904 alumnos —488 chicos y 416 chicas— entre 9 y 12 años de edad, residentes en cuatro regiones representativas de Francia. A cada uno de los alumnos le pidieron que escribiera veinte palabras por cada uno de los dieciséis centros de interés.

Los investigadores franceses llamaron a las palabras obtenidas de esta manera “palabras disponibles”. Aprovechando este nuevo “corpus” de palabras disponibles, empezaron a medir la frecuencia de cada una de ellas, es decir, las veces que había sido anotada cada palabra por los alumnos encuestados. De este modo, asignaron a cada palabra obtenida un “grado de disponibilidad” dependiente de la frecuencia, el cual simboliza, según los autores franceses, la presencia más o menos inmediata de una palabra en nuestra memoria. Como consecuencia, conseguieron confeccionar un listado de léxico disponible conforme a dicho grado de disponibilidad. De ahí, surge por primera vez la noción de léxico disponible. Según el investigador René Michéa, un vocablo disponible es aquel que, aunque no siempre actualizado en el discurso, existe en la mente del hablante y puede ser empleado de inmediato y naturalmente por éste en las situaciones discursivas que crea convenientes. Humberto López Morales (1999b: 11) define posteriormente el léxico disponible como “el

caudal léxico utilizable en una situación comunicativa dada”; teniendo en cuenta que, siempre que no se produzca alguna asociación, ciertas unidades léxicas no aparecen. En definitiva, el léxico disponible guarda relación con las palabras temáticas como se ha demostrado más arriba, las cuales son el objetivo profundo de los investigadores franceses en su afán por averiguar cuál es el léxico fundamental de su idioma. Como consecuencia de esas investigaciones de disponibilidad léxica, en el año 1964, se dio a la luz una publicación corregida de las conclusiones anteriores que se titula *L'élaboration du français fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*; y ésta volvió a ser reeditada y ampliada tres años después, en 1967.

Así es como nació la disponibilidad léxico, siendo una medida complementaria de la información que aportaban las listas de frecuencia o, más bien, de la información que aportaba el listado del léxico básico, dado que, en 1964, Alphonse Juilland y Eugenio Chang Rodríguez ya publicaron el *Frequency Dictionary of Spanish Words*, el cual refinó considerablemente el criterio de la frecuencia y dio lugar al léxico básico. Ahora, si volvemos a la iniciativa de indagar cuáles son las palabras usuales de una comunidad de habla, podemos darle una respuesta en el ámbito de la léxico-estadística: está, por un lado, el léxico básico, que constituyen unas actualizaciones léxicas efectivas, frecuentes y estables; y, por otro lado, el léxico disponible que consiste en otras actualizaciones léxicas potenciales e inestables. De la combinación de ambos, complementarios entre sí, resulta el léxico fundamental de una lengua.

3. ESTADO DE LA CUESTIÓN

3.1. Estudios pioneros de disponibilidad léxica

Pasando de las obras de frecuencia léxica y léxico básico a la primera obra en Francia sobre la disponibilidad léxica, llegamos, en este apartado, a los estudios de trascendencia que abordaron tempranamente la disponibilidad léxica y que, más tarde, dirigieron el desarrollo o sirvieron como ejemplo para las investigaciones posteriores. Presentaremos a continuación diacrónicamente la metodología adoptada y las conclusiones obtenidas en cada una de las obras que abrieron camino a la disciplina o que se dieron a conocer ampliamente en la fase inicial de la misma.

3.1.1. En primer lugar, volvemos a fijarnos en *L'élaboration du français fondamental*, concentrándonos, esta vez, en qué novedades ha aportado, cómo se ha estructurado, qué conclusiones parciales y finales ha sacado y qué resultados finales ha tenido. Como se ha mencionado en el apartado anterior, durante décadas se creyó que las unidades léxicas más habituales de una lengua eran aquellas más frecuentes en función de su número de aparición⁴ en algún corpus concreto bien diseñado, de modo que las investigaciones principales de la léxico-estadística se enfocaban en la frecuencia léxica, calculando solo las palabras estables del discurso y pasando por alto las demás, que René Michéa definía como *palabras temáticas*. Una vez que descubrió la limitación del criterio de frecuencia léxica a mitad de camino de su intento de seleccionar el léxico fundamental de la lengua francesa, el equipo investigador del que formaba parte René Michéa diseñó pruebas asociativas con el objetivo de obtener lo que llamaba *palabras disponibles*, en las cuales utilizaba centros de interés —campos semánticos que servían de estímulo

⁴ O bien, de su número de aparición junto con su dispersión, según otras investigaciones posteriores. Como hemos mencionado, Edward Lee Thorndike (1921), Milton Alexander Buchanan (1929), George Emil Vander Beke (1932) fueron los autores que estudiaron la dispersión, dándole el nombre de *rango*. Alphonse Juilland y Eugenio Chang Rodríguez (1964) desarrollaron la noción del *rango* y crearon los índices de *dispersión simple* y *dispersión compleja*.

temático— para incitar asociaciones de palabras temáticas. La obra resultante de estas investigaciones del equipo francés, *L'élaboration du français fondamental*, es el primer estudio que aborda la disponibilidad léxica.

En esa obra pionera —con el fin de facilitar la enseñanza de una lengua de base que incluía un vocabulario y una gramática simplificados—, la selección del léxico fundamental se basaba, a su vez, en tres factores. Estos eran la *frecuencia*, la *disponibilidad* y el *empirismo racional*, de los cuales el segundo factor constituía un parámetro innovador. En la primera parte de la obra, los autores franceses pusieron de manifiesto la parcialidad de aquellos estudios anteriores al suyo, en los que se intentaba realizar tal simplificación del vocabulario aplicando o bien un método empírico racional, o bien un procedimiento estadístico de frecuencia. La reflexión sobre los antecedentes de estudio léxico se dio cuando los autores franceses hicieron una revisión minuciosa de los resultados de esos estudios, que presentaban grandes inadecuaciones. Eran inadecuaciones que para nada favorecían, e incluso obstaculizaban, la adquisición y aprendizaje del léxico de base: carencia de palabras comúnmente usadas, originada por las limitaciones de la metodología adoptada; aparición de palabras impropias de la expresión cotidiana, debido a la toma, como referencia, de fuentes demasiado formales, literarias o anticuadas; y subjetividad de los autores en la recogida de textos para formar el corpus de base.

En la segunda parte de la obra, los autores, con el fin de contrarrestar los mencionados inconvenientes de las fuentes o del corpus de lengua escrita, optaron por otro método que consideraban reformado en cuanto a la recogida de datos. Grabaron muestras de lengua oral en entrevistas con informantes determinados y efectuaron transcripciones que servían de base para la elaboración de su lista de frecuencias. A pesar de ello, mientras repasaban sus propios resultados y comparaban estos con los de investigaciones anteriores, les llamó la atención la presencia predominante de palabras gramaticales y la ausencia o baja frecuencia de palabras muy usuales.

Así pues, en la tercera parte de la obra, los investigadores franceses propusieron la existencia de un vocabulario disponible frente al vocabulario frecuente y los distinguieron por la característica de ser temático o atemático, respectivamente. Detallaron, en esta parte, la metodología de la disponibilidad léxica, dando informaciones sobre los alumnos informantes y sus escuelas, los datos que se solicitaron a cada informante en el cuestionario y el procedimiento de las encuestas de prueba asociativa. Curiosamente, aprovechando las informaciones personales requeridas a cada encuestado, los autores llevaron a cabo un análisis sobre diferencias geográficas y sociológicas de los datos obtenidos. Como resultado, concluyeron que no había diferencias ostensibles en las palabras obtenidas de acuerdo con las cuatro áreas geográficamente dispares que eran tomadas para componer la muestra, si no se trataba de alguna palabra propia de una región específica. Al contrario, sociológicamente, sí había desigualdad notoria entre palabras de escuelas urbanas y las de escuelas rurales. El factor sociológico sexo no aportaba diferencias considerables, excepto en unos pocos centros de interés en los que los chicos y las chicas tenían preferencias esperadamente distintas: por ejemplo, en *La ropa*, los chicos aludieron a más indumentaria de chicos que de chicas, y viceversa; en *La casa*, los chicos se inclinaban por aparatos para almacenar comida y otros accesorios, mientras que las chicas mostraban preferencia por objetos de confort y hasta de lujo. Al final de la tercera parte, los autores incorporaron la lista de vocablos obtenidos en las pruebas sobre los centros de interés *Partes del cuerpo* y *La ropa*, enseñando sus grados de disponibilidad.

Por último, en la cuarta parte de la obra es donde los pioneros franceses expusieron su resultado final sobre el francés fundamental, integrando los tres índices citados más arriba: la *frecuencia*, la *disponibilidad* y el *empirismo racional*. Estaba compuesto de 1475 vocablos, los cuales incluían el vocabulario más frecuente del idioma —en su mayoría, palabras gramaticales, verbos, adjetivos y adverbios— y los sustantivos que se obtuvieron de las listas de disponibilidad —pero no todos ni exclusivamente, porque habían descontado y agregado a la lista algunos vocablos—. Explicaron que el criterio que habían seguido para este cribaje había sido el de que los vocablos del

francés fundamental debían permitir la expresión, en francés, de cualquier concepto básico de manera concisa (Bellón Fernández, 2011; Serrano Zapata, 2014). *L'élaboration du français fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, publicado en 1964, sentó las bases nocionales y metodológicas de nuestra disciplina y, de estas, brotaron, en lo sucesivo, numerosos estudios sobre el léxico disponible. Desde esta obra de base, la disponibilidad léxica se estrenó y se creó una nueva línea de investigación sobre el léxico que constituye una parcela importante de la léxico-estadística.

3.1.2. Siguiendo los pasos casi inmediatos de los investigadores franceses, en el mismo año —en 1964—, un autor alemán, Jay Alan Pfeffer, se inspiró en la noción de *disponibilidad* y la asimiló en una investigación léxica suya con finalidad didáctica. Los resultados de esta investigación los recopiló en *Basic (spoken) German word list: Grundstufe*. Su iniciativa era, de igual modo, elaborar el léxico fundamental del alemán que favoreciera el aprendizaje y enseñanza de este idioma. El criterio de disponibilidad léxica formó parte de los métodos que se utilizaron para seleccionar las palabras fundamentales. Las encuestas de Pfeffer para recoger las palabras disponibles las recogió en Alemania, Austria y Suiza. En las zonas germanófonas de estos países, empezó a realizar pruebas en 2.000 informantes —cifra estimada en un principio— procedentes de 82 escuelas pertenecientes a 48 ciudades. Había chicos y chicas en igual número cuya edad oscilaba entre 15 y 16 años

A diferencia de la metodología empleada en Francia, Pfeffer daba a los alumnos encuestados 21 centros de interés⁵ y les concedía un total de diez minutos para completar todos los campos semánticos. Además, quería que cada informante aportara 20 sustantivos, 12 verbos y 8 adjetivos en cada uno de los centros. Según explicaba el autor, la proporción de 20 sustantivos / 12 verbos / 8 adjetivos se basaba en el examen de varios recuentos de palabras

⁵ No detalla Pfeffer, en su propia obra, cuáles eran estos veintiún centros de interés que se utilizaron en las pruebas asociativas, sino que solo cita tres de ellos, que son: *Ein- und Verkaufen* (Compra y Venta), *Handwerk und Beruf* (Artesanía y Profesión) y *Kleidung und Wäsche* (Ropa y Lavandería) (Pfeffer, 1964: 11).

en alemán, en los cuales estaba demostrado que la proporción entre los sustantivos, los verbos y los adjetivos era aproximadamente 4:2:1. No obstante, al pensar en recabar diferentes tipos de palabras, Pfeffer preveía que, si se solicitaba a un alumno, al mismo tiempo, sustantivos y verbos, habría en las respuestas de éste bastantes derivaciones subconscientes de verbo a sustantivo y viceversa, lo que obstaculizaría la aparición de palabras diferentes. Por lo tanto, la decisión fue que, en algunas escuelas y en 1.000 estudiantes —la mitad de la muestra estimada—, solo se recogieron sustantivos, mientras que, en las otras escuelas y en los 1.000 estudiantes restantes, solo verbos y adjetivos. Pensada así la realización de la prueba, el investigador afirmó que, para interferir lo menos posible en el programa educativo de las escuelas, les ofreció la opción de escoger varias clases con el fin de dar suficientes respuestas; en consecuencia, el número de estudiantes que en realidad participaron en las pruebas se aproximó a los 5.400 (Pfeffer, 1964). De todas maneras, la investigación resultó exitosa y, al examinar los resultados obtenidos en las pruebas, la determinación definitiva del investigador fue incluir en el léxico fundamental aquellos vocablos disponibles cuyo índice de disponibilidad superara 100 (Serrano Zapata, 2014: 53).

3.1.3. Con la publicación sucesiva de los resultados sobre el léxico fundamental del francés y del alemán, la metodología de la disponibilidad léxica, como una técnica de selección de palabras usuales recién estrenada e innovadora, iba ganando la atención de los investigadores que estudiaban el léxico. Así pues, no tardó en ser retomada, esta vez, en Irlanda, en una investigación de índole didáctica sobre el irlandés. Fue una obra —*Buntús Gaeilge* ('Lengua Irlandesa')— de Colmán Ó Huallacháin, quien trataba de averiguar los elementos que se consideraban más importantes en los niveles iniciales de la enseñanza del irlandés. Para ello, el autor partía de tres aspectos a la vez: la estructura sintáctica, la morfología y el vocabulario donde se recogieron las palabras disponibles⁶. Los dos primeros aspectos eran tratados

⁶ Ó Huallacháin utilizaba en su estudio *utility words* para referirse a las palabras disponibles. Esta denominación es diferente de la que se utiliza actualmente en los trabajos en inglés sobre disponibilidad léxica —*available words*—.

a partir de un corpus de base constituido por grabaciones de conversaciones que hizo el autor en unas situaciones habituales de comunicación. Está claro que el último aspecto —las palabras disponibles, de las que aquí nos ocupamos— se inspiró en los trabajos de Gougenheim et al. y de Pfeffer.

La novedad de la investigación de Ó Huallacháin reside en que pretendía establecer al mismo tiempo el léxico disponible de dos lenguas en contacto, el irlandés y el inglés. Contaba, para cada una de las dos lenguas, con 260 informantes, entre estudiantes —desde primaria y secundaria hasta escuelas comerciales— y adultos, los cuales eran elegidos teniendo en cuenta la edad, el sexo, la residencia rural o urbana, la ocupación y la distribución geográfica. El autor utilizó veinte centros de interés⁷ y solicitó a los informantes veinte palabras en cada uno de ellos. Como resultado, en el año 1966, pudo elaborar las listas de palabras disponibles en irlandés y en inglés, alfabéticamente organizadas y en función de sus frecuencias y distribuciones en los centros de interés (Serrano Zapata, 2014).

3.1.4. Indudablemente, Ó Huallacháin no era el único a quien había atraído la noción de disponibilidad léxica. Solo tres años después, en 1969, apareció un libro que trataba específicamente la disponibilidad léxica en inglés. Era *Lexical availability: A new aspect of the lexical availability of secondary school children*, de Naum R. Dimitrijević, cuyos datos fueron recogidos en Edimburgo, Escocia. Los trabajos de Dimitrijević recibieron gran influencia de Michéa y Gougenheim; como decía el mismo autor, tomaba en parte la metodología de disponibilidad léxica de las investigaciones de ellos para realizar la suya propia. Antes de entrar en el tema principal, en las primeras

⁷ Estos veinte centros de interés que utilizó Ó Huallacháin son: 1. *The body and its parts*, 2. *Physical exercises, athletics, sports*, 3. *Recreation hobbies, indoor pastimes*, 4. *Illnesses (Kinds of sicknesses), injuries, cures*, 5. *Religion, the church*, 6. *Clothes of all Kinds (men, women, boys, girls)*, 7. *Food and drink*, 8. *Meals, utensils for each meal and for preparing the meals*, 9. *The home or house: furniture, fittings, heat, light*, 10. *Business: buying, selling*, 11. *Life in town and city*, 12. *Life and work in the country and the landscape*, 13. *Animals, birds, fish, insects*, 14. *Trees, plants, flowers, weeds, crops*, 15. *Trades and Kinds of occupation or work*, 16. *Travel, means of transport and things connected with them*, 17. *The family, relatives and people's feelings about themselves and others*, 18. *School life, activities, furniture*, 19. *The ruling of the country* y 20. *The weather, the time of day, seasons*.

páginas del libro, Dimitrijević expresó cómo entendía una palabra disponible, comentando que “una palabra está disponible si, sin ser particularmente frecuente, siempre está lista para ser utilizada y si viene a nuestra mente de forma inmediata y natural cuando la necesitamos”⁸ (Dimitrijević, 1969: 3).

La muestra de su estudio la constituían 185 alumnos de 14 años de edad, todos provenientes de centros de enseñanza urbanos. Las variables sociolingüísticas con las que Dimitrijević realizó análisis cuantitativos y cualitativos fueron: el *sexo* —chico o chica— y el *tipo de centro escolar* —*junior secondary school* o *senior secondary school*—. Hemos de recordar que esta segunda variable refleja el nivel de estudio —variable comúnmente empleada en las investigaciones actuales— de los informantes, puesto que una *junior secondary school* da educación inmediatamente inferior a la de una *senior secondary school*. En la aplicación de la prueba asociativa, el autor diseñó once centros de interés⁹, fijó en cinco minutos el tiempo de respuesta correspondiente a cada centro de interés y requirió que los alumnos escribieran las respuestas lo más rápido que pudieran.

Llama la atención, con respecto a las respuestas recabadas en su investigación, que solicitó a los informantes escribir *lexical item* o *lexical phrase*¹⁰ en vez de sólo palabras individuales. Aparte de ello, cabe destacar otra aportación metodológica que el autor hizo en la aplicación de la prueba: puesto que se tenía como objetivo recoger la mayor cantidad posible de palabras disponibles proporcionadas por los informantes, Dimitrijević no les pidió un número fijo de unidades léxicas, es decir, no se adhirió al uso de las listas cerradas —listas de veinte palabras en las pruebas de Gougenheim y

⁸ Ha sido del propio autor de la tesis esta traducción al español. El texto original en inglés es el siguiente: “A word is available if, without being particularly frequent, it is always ready to be used and if it comes to our mind immediately and naturally when we need it”.

⁹ Los once centros de interés que Dimitrijević ingenió modificando aquellos de los pioneros franceses, Gougenheim y Michéa, son: 1. *Animals*, 2. *Countryside*, 3. *Town*, 4. *Entertainment*, 5. *Jobs and Professions*, 6. *Science*, 7. *Means of transports*, 8. *Politics*, 9. *Parts of the house*, 10. *Food and drink* y 11. *Clothes*.

¹⁰ Se entienden por *lexical item* o *lexical phrase* unidades léxicas que abarcan las palabras individuales y las unidades compuestas, en contraposición con las palabras individuales exclusivamente.

sus colegas, por ejemplo— sino que facilitó una lista abierta correspondiente a cada uno de los once centros de interés. Su decisión de las listas abiertas constituye una pauta fundamental para los trabajos posteriores, y ellas siguen presentes en gran mayoría de las investigaciones sobre disponibilidad léxica de hoy en día (Ayora Esteban, 2006; Serrano Zapata, 2014).

A pesar de que partía de la finalidad de conocer la disponibilidad de las palabras inglesas en los escoceses, Dimitrijević pudo también hacer análisis cuantitativos según las variables. En cuanto a la variable *sexo*, el autor detectó en la mayoría de los centros de interés un mayor número de respuestas por parte de los chicos. Aun así, creía que el *sexo* era una variable menos importante que el *tipo de centro escolar*, cuyos resultados eran verdaderamente relevantes. Los estudiantes de *senior secondary school* mostraban una producción léxica mejor que los de *junior secondary school* en todos los centros de interés, con una media de 5,51 palabras más por centro. Ello es comprensible si tenemos en cuenta el nivel de estudio de estos dos grupos de informantes en función la etapa educativa en la que se encuentran.

Por otro lado, desde una perspectiva de análisis cualitativo, el autor propuso que se pusiera la atención en los contactos que pudiera tener la disponibilidad léxica con otros estudios lingüísticos o psicológicos, aunque generalmente la finalidad de la disciplina era didáctica. En este sentido, Dimitrijević hizo unas observaciones interesantes: parejas de palabras que frecuentemente aparecían juntas —*perro* y *gato*, por ejemplo—; casos en que los informantes empezaban una lista encadenando palabras relacionadas con alguna área específica y no pasaban a otra área sin terminar el recorrido completo de la anterior. Efectivamente, la propuesta de Dimitrijević de extender las investigaciones de disponibilidad léxica se ha hecho realidad, teniendo en cuenta que los estudios interdisciplinarios son cada vez más frecuentes en la disponibilidad léxica, convirtiéndose ésta en una herramienta no solo en la enseñanza de lenguas maternas y extranjeras, sino también en otros campos

educativos, lingüísticos, psicológicos o sociales¹¹ (Bellón Fernández, 2011).

3.1.5. Pero los trabajos de Dimitrijević sobre disponibilidad léxica no acaban aquí. Si echamos la mirada hacia adelante, veremos que los resultados estadísticos que él sacó en 1969 en Edimburgo volvieron a ser empleados, en 1981, en un artículo resumido de estudio comparado, titulado *A comparative study of the lexical availability of monolingual and bilingual schoolchildren*. Para este artículo, Dimitrijević había recabado datos de léxico disponible en Novi Sad, ciudad de la antigua Yugoslavia, entre 228 estudiantes que estudiaban inglés como lengua extranjera. Estos informantes estaban divididos en dos grupos en función de su lengua materna: 114, esto es, la mitad de ellos, eran monolingües cuya lengua materna era el serbocroata; la otra mitad eran informantes bilingües que tenían como lengua materna el húngaro y como segunda lengua el serbocroata. Estando así la muestra, el autor hizo, en esta nueva investigación, comparaciones cuantitativas y cualitativas según variables sociales entre los 114 alumnos monolingües y 114 bilingües de Novi Sad y los anteriores 185 informantes de Edimburgo.

A diferencia del estudio en Edimburgo, donde los alumnos informantes eran homogéneamente de 14 años, en Novi Sad, eran chicos y chicas de dos estratos de edad: 11 y 14 años. Esta diferencia en la edad se dio porque los estudiantes de 11 y 14 años fueron elegidos, respectivamente, del quinto y octavo cursos de la escuela primaria. El autor consideraba estos dos cursos clave como variable en su investigación, explicando que en el quinto curso empezaba la enseñanza formal de la lengua materna, mientras que el octavo curso es el último de la escuela primaria de Yugoslavia. Junto con la *lengua materna*, el *sexo* y la *edad*, Dimitrijević consideró como variable la *inteligencia*¹² de los informantes y sus *calificaciones* en la escuela. Además de estas

¹¹ Tratamos, en un apartado posterior, el contacto que tiene la disponibilidad léxica con otros campos de estudio.

¹² Debido al reducido volumen del artículo, el autor no explica cómo había evaluado la inteligencia de los informantes. Solamente alude, en una nota a pie de página, a que la variable *inteligencia* contenía tres variantes (Dimitrijević, 1981: 114), pero tampoco menciona cuáles eran.

variables que podían afectar, en general, a la disponibilidad léxica de los estudiantes, el autor también tomó otras que probablemente influían en el desarrollo de la lengua y en el comportamiento lingüístico: el *número de hermanos* del informante, su *estatus social* en la escuela y la *profesión y educación de sus padres*; aunque afirmó él mismo, en el apartado de los resultados obtenidos, que estas últimas variables apenas estaban en correlación con las respuestas proporcionadas por los informantes. Con respecto a los centros de interés, el autor agregó dos más que en el estudio de Edimburgo, esto es, *Parts of the body* y *War and peace*, como resultado de lo cual, los centros de interés en Novi Sad sumaban 13¹³.

Sobre la aplicación de la prueba asociativa, cabe mencionar que Dimitrijević había pensado que la prueba en forma escrita era una posible debilidad de la metodología, puesto que con pruebas orales podría obtener muchas más repuestas y palabras diferentes. Así pues, aunque, según decía el autor, era imposible hacer pruebas orales sobre la enorme cantidad de informantes que él tenía, hizo una sobre una muestra pequeña —un 10% del total de los informantes—. Grabó las respuestas orales y, al comparar estas con las obtenidas en las pruebas escritas, observó poca diferencia, tanto respecto al número de respuestas como a la diversidad de palabras. Todo lo contrario, lo que pasó con las pruebas orales era que algunas palabras se repitieron, en general, varias veces.

Como resultado de este estudio comparado, Dimitrijević aportó las siguientes conclusiones cuantitativas relevantes. En primer lugar y al respecto de la *lengua materna*, entre los informantes de Novi Sad, los monolingües —prácticamente, los de mayor edad (14 años)— proporcionaron más palabras que los bilingües¹⁴. De igual modo, los primeros sobrepasaban, sin importar la

¹³ Estos resultaron ser 1. *Animals*, 2. *Countryside*, 3. *Town*, 4. *Entertainment*, 5. *Jobs and professions*, 6. *Means of transportation*, 7. *Science*, 8. *Parts of the house*, 9. *Food and drinks*, 10. *Clothes*, 11. *Politics*, 12. *Parts of the body* y 13. *War and peace*.

¹⁴ Según explicaba el mismo autor, los informantes monolingües de menor edad (11 años) también aportaban más palabras que los bilingües, pero, estadísticamente, esta diferencia de respuestas no era muy significativa (Dimitrijević, 1981: 114).

edad, a los segundos en el número de palabras diferentes. Asombrosamente, en la comparación transnacional, los estudiantes de Edimburgo que tenían el inglés como lengua materna no mostraban su ventaja: los monolingües de Novi Sad dieron más palabras que ellos, mientras que los escoceses tampoco sobrepasaban significativamente a los bilingües de Novi Sad en este sentido.

En segundo lugar, con respecto a la *edad*, comprobó que los alumnos mayores aventajaban a los de menor edad, tanto en los grupos monolingües como en los bilingües.

En tercer lugar, en cuanto al *sexo*, las chicas aportaron más respuestas que los chicos; sin embargo, los chicos dieron más palabras diferentes que las chicas, resultado comprobable en todos los grupos de lengua materna inglesa, serbocroata o húngara. Pero, cuando pasó de los datos como conjunto a los datos de cada uno de los centros de interés, vio que había centros preferidos por los chicos y otros, por las chicas, es decir, en algunos centros de interés los chicos escribieron más palabras que las chicas, y viceversa.

En cuarto lugar, el investigador indicó que la variable *inteligencia* tenía una correlación positiva con el resultado cuantitativo de la disponibilidad léxica; al contrario, las *calificaciones escolares* solo ejercían influencia positiva en pocas comparaciones, de modo que no vio la importancia de las mismas. Por último, aseguró, como hemos comentado antes, que el *número de hermanos* del informante, su *estatus social* en la escuela y la *profesión y educación de sus padres* no eran variables válidas en su investigación.

Además de las conclusiones cuantitativas generales que se acaban de mencionar, Dimitrijević se interesó, en el sentido cualitativo, por comparar las cinco primeras palabras —o lo que viene a ser lo mismo, las cinco palabras de mayor grado de disponibilidad— en cada centro de interés. Al respecto de la *lengua materna*, las comparaciones transnacionales permitieron ver gran diferencia entre las primeras palabras dadas en Edimburgo y en Novi Sad, sobre todo, en los centros de interés *Politics*, *Food and drinks* y *Professions*,

lo que era imaginable, teniendo en cuenta los muy distintos contextos culturales y sociales que influían en sus maneras de vivir y pensar.

Por otra parte, entre los alumnos de Novi Sad, no se encontró mucha diferencia entre los monolingües y los bilingües; es decir, ambos grupos apuntaron, más o menos, las mismas cinco primeras palabras. Sin embargo, los alumnos bilingües escribieron ciertas palabras dialectales de su segunda lengua —el serbocroata—¹⁵, las cuales no aparecían en las listas de los hablantes monolingües cuya lengua materna era el mismo serbocroata.

En cuanto a la *edad*, poca diferencia se notó al comparar las listas en lengua materna y en segunda lengua; pero los alumnos de edad menor dieron menos palabras de alta disponibilidad que los mayores. Por último, chicos y chicas no difirieron mucho en la selección de palabras, con excepción de los centros de interés *Professions, Food and drinks* y *Clothes*.

Al final del artículo, Dimitrijević planteó unas investigaciones viables que se podían hacer en adelante. Entre otras, recomendó aplicar las pruebas en otros países de habla inglesa, haciendo cotejos con su muestra en Edimburgo para buscar efectos de los contextos culturales y sociales en la disponibilidad léxica de una misma lengua; sugirió que se estudiaran informantes bilingües cuyas lenguas fuera más cercanas, considerando que el serbocroata y el húngaro, en su propia muestra de Novi Sad, eran dos lenguas muy distantes; también esperaba extender la aplicación de la disponibilidad léxica a estudios neurolingüísticos, dado el intenso interés de los lingüistas por estos (Dimitrijević, 1981).

3.1.6. Ahora volvamos a la línea de tiempo según la que tratamos los estudios pioneros de la disponibilidad léxica. Después de la primera obra de Dimitrijević en 1969, aparecieron en Canadá otras investigaciones similares de

¹⁵ Aquí se entiende que los informantes bilingües húngaro-serbocroatas, al completar las encuestas sobre el léxico disponible en su segunda lengua —el serbocroata—, aportaron palabras dialectales.

gran influencia encabezadas por William Francis Mackey. Los estudios dieron como fruto, en el año 1971, *Le vocabulaire disponible du français* de dicho investigador canadiense con colaboraciones de Jean Guy Savard y Pierre Ardouin. En realidad, en torno a los años 1958 y 1959, con el objetivo de elaborar estudios comparativos, los lingüistas canadienses ya mostraban su interés por los estudios de disponibilidad léxica sobre el francés que entonces se estaban llevando a cabo en Francia. Poco después, algunos de ellos pusieron en marcha investigaciones de este tipo en regiones canadienses como Québec y Manitoba, pero modificando, por su parte, la metodología de la disponibilidad léxica. William Francis Mackey vio inconvenientes en el hecho de que cambiaran las pautas metodológicas, sobre todo, si habían partido de la idea de establecer comparaciones posteriores entre el léxico de una y otra zona de habla francesa. Considerando como riesgo la alteración de los métodos de trabajo, Mackey se ajustó, a su vez, a la metodología de los investigadores franceses, con la finalidad de buscar diferencias culturales en el léxico entre dos regiones del mundo de habla francesa: Francia y Acadia (Canadá) (Ayora Esteban, 2006; Mateo García, 1998).

Según Mackey, para conseguir esto, el motivo por el que recurrió al léxico disponible era evidente: dado que las palabras atemáticas que ofrecían las listas de frecuencias o el vocabulario básico eran compartidas por todos los hablantes de una misma lengua sin importar la zona geográfica, solo el léxico disponible, estrechamente relacionado con los contextos, insinuaba características socioculturales de la lengua propia de cada una de las regiones estudiadas (Bellón Fernández, 2011).

La obra resultante de estas investigaciones, *Le vocabulaire disponible du français*, tiene dos tomos. El primero se titula *Le vocabulaire concret usuel des enfants français et acadiens: étude témoin*¹⁶. En este tomo, el autor repasó en el primer capítulo los tres indicadores que, en la mayoría de los estudios precedentes, se usaban para elaborar listas de palabras. Estos eran la

¹⁶ La traducción al español del título de este primer tomo es: *El vocabulario concreto usuel de los niños franceses y acadianos: un estudio de muestra*.

frecuencia, la *dispersión* y la *valencia*. La *frecuencia* calcula el número de veces que aparece una palabra en el corpus tratado; la *dispersión* estudia el número de fuentes o ámbitos temáticos diferentes en que aparece una palabra; y la *valencia* mide la cualidad de polivalente de una palabra, esto es, si tiene alta posibilidad de combinación en el discurso, si contiene otros conceptos o si es apta para ejercer como sinónimo de gran cantidad de palabras (Serrano Zapata, 2014: 51). Concluyó Mackey que los tres parámetros no servían mucho si se trataba de seleccionar los sustantivos usuales muy concretos, tal y como Gougenheim y sus colegas afirmaron en su libro.

Así pues, tal y como los investigadores franceses buscaron la solución en la disponibilidad léxica, Mackey pasó a introducir la noción de esta en el segundo capítulo de su obra y explicó los instrumentos y técnicas que utilizó para obtener el léxico disponible. Como se ha comentado en el apartado anterior, al respecto de la metodología, el autor consideró que se debía establecer el léxico disponible de Acadia en función de unos procedimientos idénticos a los que habían producido los resultados de Francia, porque su voluntad era comparar dichos resultados.

El trabajo de Mackey formaba parte de una serie de investigaciones coordinadas por el Círculo Lingüístico de Québec, el cual se proponía extender un proyecto de disponibilidad léxica por toda Canadá. Ese instituto seguía de cerca la propuesta de Mackey y celebró reuniones con los lingüistas de otras regiones canadienses como Montréal y Laval con el fin de evitar diferencias metodológicas. En el tercer capítulo del primer tomo de *Le vocabulaire disponible du français*, Mackey expuso las decisiones metodológicas tomadas para ese proyecto ambicioso, con las que estaban de acuerdo todos los investigadores que realizaban estudios de disponibilidad léxica en varias zonas canadienses. Las directrices generales del proyecto eran las siguientes:

1. Se tenía como objetivo estudiar las diferencias de disponibilidad léxica entre el francés de Francia y el de Canadá.
2. A los dieciséis centros de interés que habían utilizado en Francia se

agregaron seis más: *L'église et la vie paroissiale*, *Les sports et l'équipement sportif*, *Les outils*, *L'automobile et ses parties*, *Les actions courantes* y *Les affaires*.

3. Al respecto de la edad, se pasaron encuestas a informantes entre 9 y 12 años, que era el segmento de edad utilizado en Francia, aunque se reservó la posibilidad de ampliar la encuesta hasta los primeros cursos de estudios secundarios.
4. Las formas y condiciones del cuestionario eran las mismas que las francesas.
5. Distribuyeron las investigaciones en 12 regiones canadienses desde el punto de vista geográfico y sociocultural, las cuales eran Gaspésie, Bas-du-Fleuve, Las Saint-Jean, Québec Nord-Ouest, Montréal, Richelieu, Sherbrooke, Trois-Rivières, Québec, Ontario, Ouest Canadien y Acadie (Bellón Fernández, 2011; Mateo García, 1998; Serrano Zapata, 2014).

Sin embargo, de estas 12 regiones demarcadas en el proyecto, solo una de ellas se comparó con Francia y esta única era la Acadia estudiada por Mackey. En el resto del tercer capítulo, Mackey describió los rasgos de sus propias encuestas en Acadia. El autor se limitó a analizar los dieciséis centros de interés clásicos que habían empleado en la encuesta francesa y se atuvo al grupo de informantes cuyas edades estaban comprendidas entre los 9 y 12 años. Las comparaciones de palabras disponibles proporcionaron resultados interesantes y constataron algunas diferencias socioculturales entre Europa y América. Por ejemplo, en cuanto a las primeras palabras más disponibles, se encontró coincidencia solo en cuatro centros de interés: *oreilles*, en *Partes del cuerpo*; *chemise*, en *La ropa*; *lampe*, en *Calefacción e iluminación*; y *chien, chat y vache*, en *Los animales*. Lo que llamó la atención era que algunas palabras que alcanzaban altas posiciones en la lista de una comunidad caían a posiciones muy bajas en la de la otra. Este es el caso de la palabra *cinéma*, que ocupaba el primer lugar en Francia, pero cayó al puesto 138 en Acadia. Encima, en Francia, la palabra *théâtre* aparecía en el tercer lugar en *Las distracciones*, mientras que en Acadia el tercer lugar lo ocupaba *hockey* y era

en el puesto 242 donde por fin aparecía *théâtre* (Bellón Fernández, 2011: 32; Mateo García, 1998: 40).

En la aplicación de las pruebas, Mackey dijo a los informantes que escribieran las palabras a su manera, de modo que las respuestas originales estaban repletas de errores, incluidos los ortográficos. Por eso, en el cuarto capítulo del libro, el autor explicaba los criterios de corrección de los datos. Entre los criterios de edición y además de los de tipo ortográfico, destacaban los siguientes:

- a) admitió no solo los sustantivos, sino también los verbos y adjetivos;
- b) mantuvo las palabras inglesas y las propias del francés de Canadá, esto es, los anglicismos y los dialectalismos canadienses;
- c) conservó aquellas respuestas que no correspondían a su centro de interés, aunque, por supuesto, no consiguieron una frecuencia alta y se quedaron al final de la lista;
- d) incluyó las expresiones formadas por un verbo y un sustantivo en la categoría del sustantivo, es decir, *jouer au tennis* (jugar al tenis) se redujo a *tennis* (tenis) (Bellón Fernández, 2011: 33).

En el quinto capítulo, el autor describió en detalle los procedimientos informáticos en el tratamiento de los datos. Con respecto al cálculo del índice de disponibilidad, empleó el porcentaje de informantes que habían aportado un vocablo. Como resultado de la investigación, el autor publicó, en este primer tomo de su obra, los índices de disponibilidad en Francia y en Acadia, las palabras de las listas francesas que no aparecían en las correspondientes listas canadienses y viceversa, y un diccionario general del léxico disponible de la lengua francesa.

El segundo tomo de *Le vocabulaire disponible du français*, titulado *Vocabulaire disponible des enfants acadiens*, estaba dedicado, específicamente, al análisis de las encuestas que realizó el autor en Acadia. Incorporó Mackey en este tomo las listas del vocabulario disponible del francés

en esa región canadiense, junto con los porcentajes de disponibilidad. Como comentó el autor en la obra, la situación lingüística de Acadia se caracterizaba por el bilingüismo, y precisamente era el efecto del bilingüismo sobre el léxico disponible el que Mackey quería analizar. Así pues, este investigador optó por detectar unos fenómenos particulares que los bilingües a menudo producían en la lengua, esto es, los préstamos en las listas de palabras. La razón de detectar los préstamos en las listas de disponibilidad la dio Mackey en su libro:

*“Puisque la disponibilité est une manifestation de la langue et non seulement de la parole, c’est l’instrument tout à fait indiqué pour déceler les emprunts, non seulement parce que le sujet bilingüe fait l’inventaire de ses connaissances d’une langue à la fois, mais aussi parce que la disponibilité touche surtout le secteur du vocabulaire où les emprunts se manifestent d’abord, à savoir, le vocabulaire concret”*¹⁷ (Mackey, Savard, & Ardouin, 1971, Tomo II: 16; Serrano Zapata, 2014: 105).

A continuación, el investigador canadiense empezó la búsqueda, entre los resultados obtenidos, de las palabras inglesas, es decir, los anglicismos. Al final, comprobó que un 13,3% del total de palabras en las listas de disponibilidad estaban en inglés. Los centros de interés donde los informantes aportaron más anglicismos fueron *Les jeux et distractions*, *Le village* y *La cuisine et les utensiles qui s’y trouvent*.

Estando así las cosas, el autor planteó una manera de averiguar qué nivel de influencia había tenido el vocabulario inglés en el vocabulario francés de Acadia. Se propuso comparar el índice de disponibilidad de cada una de las palabras francesas en las listas con el índice de disponibilidad de su sinónimo

¹⁷ Traducción al español del autor de la tesis: “Como la disponibilidad es una manifestación del lenguaje y no solo del habla, es el instrumento perfecto para detectar los préstamos, no solo porque el sujeto bilingüe hace un inventario de su conocimiento de un idioma en particular, sino también porque la disponibilidad afecta principalmente el área del vocabulario donde los préstamos primero se manifiestan, es decir, el vocabulario concreto”.

en inglés encontrado en las mismas. De esta manera, se observó que la mayoría de las palabras inglesas encontradas —la mayor parte de ese 13,3% mencionado— tenían sus correspondientes sinónimos franceses en las listas, pero pocas de ellas superaban sus sinónimos franceses en disponibilidad. Tras los análisis, Mackey concluyó que las apariciones de los anglicismos en sustitución de las palabras francesas estaban sujetas a cada informante concreto, puesto que se daban en pocos informantes. De igual modo, estaban sujetas al centro de interés en particular, porque en unos centros de interés había más anglicismos, pero en otros, no tantos o muchos menos y, además, no había ningún centro de interés donde fueran completas las sustituciones de palabras francesas por las inglesas. Dicho de otra manera, se llegó a la conclusión de que el vocabulario inglés afectaba de modo limitado al vocabulario francés de Acadia.

3.1.7. Siguiendo los primeros trabajos sobre la disponibilidad léxica, en el mismo año en que Mackey publicó *Le vocabulaire disponible du français*, es decir, en 1971, se sacó a la luz una investigación llevada a cabo por John Bailey Victory. Era una tesis de maestría, titulada *A study of lexical availability among monolingual-bilingual speakers of Spanish and English*. En esta investigación, el autor, inspirado en el trabajo que había hecho Dimitrijević, lo siguió muy de cerca —tal y como reconoció el propio Victory en la Introducción de la tesis—, y empezó a estudiar las diferencias cuantitativas y cualitativas entre las palabras disponibles de tres grupos de informantes: 33 monolingües en inglés, 33 monolingües en español y otros 33 bilingües en español e inglés.

Los informantes que componía la muestra del estudio provenían de dos institutos de enseñanza con rasgos socioculturales completamente distintos. Uno era una *escuela preparatoria* de Monterrey, México, y el otro, una *senior high school* de Houston, Estados Unidos; ambos proporcionaban educación secundaria. De los 99 informantes totales de la muestra, los 33 monolingües en español eran del primer instituto, mientras que los 33 monolingües en inglés y los 33 bilingües en español e inglés eran del segundo. Todos eran chicos y chicas de 16, 17 ó 18 años de edad —periodo de la vida en que se asiste,

naturalmente, a la escuela secundaria—. Aunque el autor calificaba la edad de factor altamente influyente en el estudio, insistió en tomar el total de la muestra como un conjunto sin distinguir los estratos de edad, toda vez que su objetivo era analizar la disponibilidad léxica en relación con otras variables como el sexo y características socioeconómicas o socioculturales.

Al respecto del sexo, el autor había pensado en aplicar la prueba en igual número de chicos que de chicas, pero tuvo que trabajar con los estudiantes y el personal escolar disponibles. De los 99 informantes, 46 eran chicos y 53, chicas. Esta pequeña diferencial de 7 informantes entre ambos sexos fue matemáticamente corregida en los cálculos posteriores, equilibrando, con métodos matemáticos, el número de un sexo con el de otro para que las comparaciones según el sexo partieran de igual número de chicos y chicas y fueran, por consiguiente, significativas. Afirmó el autor que, más que demostrar la superioridad en la producción léxica de un sexo sobre otro, tenía como objetivo observar en qué grado podía haber preferencia de cada uno de los sexos por el uso de ciertas palabras y analizar sus divergencias cuantitativas y cualitativas.

Como variable de antecedentes socioeconómicos, Victory recopiló datos sobre el estatus laboral del padre o de la persona que sostenía la familia de cada informante. Además, las dos escuelas encuestadas aportaban, por sí mismas, informaciones sobre el estatus socioeconómico de las familias de los alumnos que estudiaban en ellas. Los alumnos que asistían a la *senior high school* de Houston venían todos de un mismo distrito de la ciudad, porque sus casas se encontraban en la zona de influencia de este centro educativo. Ese distrito, en términos generales, se consideraba una zona mixta comercial-residencial de clase media o media-baja. Mientras tanto, la *escuela preparatoria* de Monterrey no servía a ninguna zona específica de la ciudad. Cabe destacar que el hecho de que estuviera alguien matriculado en esta escuela preparatoria implicaba que ingresaría a la universidad; al contrario, en la escuela de Houston en cuestión, solo alrededor del 20% de los estudiantes iban después a la universidad. Las encuestas también demostraron que la

mayoría de los padres de los informantes en Monterrey ejercían alguna labor profesional.

Las listas de prueba de Vicky eran las mismas que aquellas utilizadas anteriormente por Dimitrijević: listas abiertas, sin restringir ni el número ni el tipo de palabras solicitadas. No obstante, en una prueba previa a la formalmente realizada, Vicky observó que, después de aproximadamente dos minutos en cada centro de interés, la mayoría de los participantes ralentizaban la velocidad, es decir, disminuía la disponibilidad léxica. Por lo tanto, decidió reducir a dos minutos el tiempo de respuesta para un centro de interés, considerando que eran suficientes para recordar las palabras más disponibles.

Con el fin de determinar los detalles de la prueba, el investigador llevó a cabo otra prueba previa de tipo oral utilizando grabadoras, después de la cual comprobó las deficiencias de la prueba oral —muy similares a las ya mencionadas por Dimitrijević— y optó por la prueba por escrito. Afirmó Vicky que los participantes 1) solían tener una sensación de inhibición o aprensión al hablar solos ante un dispositivo de grabación; 2) al pronunciar sucesivamente una respuesta tras otra, tendían a repetir unas palabras a veces sin darse cuenta y otras veces interrumpían su propio pensamiento y comenzaban a comentar que habían mencionado tal palabra; 3) manifestaban una falta de concentración, sobre todo, cuando se comparaban estos resultados con otros en pruebas escritas. A estos tres inconvenientes que obstaculizaban la producción léxica o la transcripción de esta, el autor también añadió otras dos dificultades relativas a la realización de las propias pruebas: para llevar a cabo pruebas puramente orales sobre un gran número de informantes —99 en su caso—, 1) se requería mucho más tiempo y 2) se generaba más desorganización de la rutina escolar que la que las escuelas podían asumir (Vicky, 1971: 14).

Los centros de interés en las pruebas de Vicky fueron los diez siguientes: 1. Animales, 2. La ciudad, 3. Diversión, 4. Familia, 5. Dios, 6.

Trabajos y profesiones, 7. Medicina, 8. Música, 9. Espacio, 10. Guerra¹⁸. De estos, Animales, La ciudad, Diversión y Trabajos y profesiones también fueron usados por Dimitrijević o por Gougenheim y Michéa. El autor explicó que estos ámbitos semánticos fueron seleccionados teniendo en cuenta la relatividad bicultural y binacional. Aseguró que los centros de interés en este estudio suyo, que implicaría algunas distinciones biculturales inevitables, se debían basar en temas que produjeran un alto grado de homogeneidad léxica; es decir, consideró que estos estímulos obtendrían, en sentido general, palabras similares en chicos y chicas de 16 a 18 años, independientemente de su nacionalidad. Estaba convencido de que, si hubiera empleado, por ejemplo, *Comidas y bebidas*, habría una diferencia léxica tan enorme que la comparación sería imposible entre una y otra cultura o país.

Además, el investigador se refirió a los contenidos no descriptivos de sus centros de interés, comentando que en las investigaciones francesas habían dado descripciones extremadamente detalladas y estas dirigieron y redujeron el pensamiento y la producción léxica de los informantes. Victory observó que algunos de los centros de interés en las investigaciones francesas eran frases compuestas de múltiples palabras como, por ejemplo, *Les objets placés sur la table et dont on se sert à tous les repas de la journée*. Creía que los estímulos perifrásticos muy largos podrían limitar el pensamiento de los informantes y estos restringirían su asociación a un área muy estrecha en la que el vocabulario era limitado y especializado. Afirmó que los centros de interés debían estar en favor de una asociación libre de ideas en la que la producción léxica naciera de conceptos relacionados objetiva o subjetivamente con el estímulo. Concretamente, propuso *Dinner table*, en sustitución de *Les objets placés sur la table et dont on se sert à tous les repas de la journée*, porque permitiría una libertad de pensamiento de tal manera que se recabarían

¹⁸ Estos diez centros de interés fueron los utilizados en las encuestas en español para recoger palabras españolas, destinadas a los informantes monolingües en español y a los bilingües en español e inglés. Las encuestas destinadas a los informantes monolingües en inglés y a los bilingües en español e inglés para recoger palabras inglesas iban con centros de interés cuyos nombres eran la traducción de los anteriores: 1. *Animals*, 2. *The city*, 3. *Entertainment*, 4. *Family*, 5. *God*, 6. *Jobs y professions*, 7. *Medicine*, 8. *Music*, 9. *Space*, 10. *War*.

palabras de naturaleza abstracta, a saber, felicidad, conversación, unión, amor, etc., mientras que un centro de interés tan específicamente descriptivo como el establecido para el francés en cuestión limitaría a los informantes a pensar solo objetivamente (Victery, 1971: 19).

En cuanto a la aplicación de las pruebas, el investigador expuso el siguiente procedimiento. La primera prueba se dio en la escuela de Houston entre los monolingües en inglés; la segunda y tercera pruebas se llevaron a cabo entre los bilingües de la misma escuela, primero en inglés y luego en español, con una pausa de descanso entre las dos pruebas¹⁹. Por último, la cuarta prueba se realizó en la escuela de Monterrey entre los monolingües en español. En la primera y cuarta pruebas, todas las instrucciones se daban en las respectivas lenguas de los alumnos monolingües; mientras tanto, en las dos pruebas entre los alumnos bilingües, las instrucciones eran en inglés en la segunda y en español, en la tercera. Victery explicó que estas dos pruebas hechas con alumnos bilingües deberían haber sido organizadas con un intervalo de dos o tres semanas para evitar interferencia de una lengua sobre la otra, pero la programación estaba más allá del control del investigador.

Otras actuaciones relevantes fueron: 1) solicitar al personal de las escuelas que no revelaran con anticipación la naturaleza de la prueba y, sobre todo, no informaran de que se haría una “prueba”; 2) en el momento de la prueba, informar a los alumnos de que se haría un “experimento de asociación de palabras” y que de ninguna manera estaba relacionado con su trabajo escolar ni con sus calificaciones; 3) introducir un centro de interés como ejemplo: *Ships and boats*, seguido de unas palabras libremente asociadas: *water, sail, ocean, blue, palm trees, fish*, etc. y luego pedir a los estudiantes que ofrecieran voluntariamente sus propias palabras asociadas; 4) explicar que la ortografía, en este caso, no era importante y que siempre debían intentar escribir una palabra o abreviarla, esto es, no imponer ninguna restricción

¹⁹ El autor no detalló cuánto tiempo duró la pausa de descanso entre la prueba en inglés y la prueba en español. Por lo tanto, se supone que pudo existir cierto grado de interferencia de una lengua sobre la otra.

ortográfica a la respuesta léxica; 5) avisar a los informantes del tema de un centro de interés solo en el momento de empezar dicho entro, de modo que los estudiantes no supieran de qué trataban los siguientes centros de interés y se concentraran en la temática del momento.

Después de presentar minuciosamente la metodología de su investigación, Victory pasó al tratamiento de los datos obtenidos. Como resultados generales, incluyó en el trabajo 1) sendas listas —según los diez centros de interés— de las palabras que se repetían al menos 8 veces bajo cada uno de ellos; 2) la lista del orden de los centros de interés según la producción de palabras totales y palabras diferentes; 3) la lista de todas las palabras que se repetían al menos 8 veces en todo el corpus; 4) la lista de los diez centros de interés con las respectivas cantidades de palabras totales, de palabras totales en inglés y en español, de palabras diferentes en inglés y en español, de palabras totales por sexo; 5) la lista del orden de los centros de interés según cada una de las categorías del cuarto punto; 6) la lista del orden de los centros de interés según la producción de palabras que se repetían por lo menos 8 veces; y, al final, 7) la lista de las palabras en español según su disponibilidad léxica y la misma para el inglés.

Como resultados de la comparación, el investigador presentó las siguientes listas más importantes:

- 1) la lista con el número de palabras totales comparando los monolingües en inglés y los monolingües en español, por un lado; los monolingües en inglés y los bilingües en inglés, por otro; y, finalmente, los monolingües en español y los bilingües en español;
- 2) la lista del promedio de palabras por informante en cada centro de interés comparando los monolingües en inglés y los monolingües en español, por un lado; los monolingües en inglés y los bilingües en inglés, por otro; y, finalmente, los monolingües en español y los bilingües en español;
- 3) la lista del promedio de palabras por informante en cada centro de interés comparando los chicos y las chicas.

Basándose en los datos que se acaban de mencionar, Victory llegó a las siguientes conclusiones más relevantes. En primer lugar, los monolingües en inglés produjeron el mayor número de palabras, pero los bilingües aportaron el mayor número de palabras diferentes en español. Cuando se hicieron comparaciones entre una lengua y otra, se puso de manifiesto que en inglés los informantes eran capaces de producir más palabras que en español. En cuanto al sexo, se observó que las chicas superaban a los chicos en la producción léxica no solo desde el punto de vista del corpus entero sino también en cada uno de los centros de interés sin excepción. Esta observación difería significativamente de los hallazgos generales de otros estudios, que afirmaron no haber detectado superioridad léxica de un sexo sobre otro.

Además, el investigador notó que el 44.81% de todo el corpus estaba compuesto de palabras con frecuencia de, al menos, 8 repeticiones, lo que demostró claramente la existencia de una concomitancia cognitiva; es decir, había un grado muy significativo de disponibilidad y acuerdo universales sobre estos elementos léxicos de ocurrencia múltiple.

Al comparar las palabras diferentes en español con las diferentes en inglés, el autor observó una notable coincidencia de disponibilidad léxica en las dos lenguas. Así pues, cuando puso en una columna las palabras españolas de disponibilidad superior a 50% y, en otra columna, las palabras inglesas con el mismo grado de disponibilidad, se pudieron encontrar en la segunda columna los homólogos del 75% de las palabras de la primera columna. Esto quiere decir que los informantes de las dos lenguas tendían a recordar los mismos conceptos.

3.1.8. La influencia de las investigaciones sobre el francés de Gougenheim y sus colegas y de Mackey se manifestó en la aparición de varios estudios posteriores que los seguían de cerca. Entre estos se destacó un trabajo de Pierre Emmanuel Njock, que estudiaba la disponibilidad léxica de la lengua francesa y del idioma basaa de Camerún. Esta obra titulada *L'univers*

*familier de l'enfant africain*²⁰ se sacó a la luz en el año 1978. Tenía como objetivos familiarizarse con el entorno de los niños africanos, investigar el vocabulario al que estos estaban expuestos en el hogar y en la escuela, y comparar el vocabulario del idioma nativo —el basaa— con el de la escuela —el francés—.

El propio investigador resumía su trabajo en tres partes principales. La primera parte de la obra era un estudio lingüístico de la situación familiar. El autor empleó las pruebas de disponibilidad léxica con el fin de descubrir el repertorio léxico de los niños, elaborando listas de palabras que conocían estos en su lengua materna, el basaa, así como en el francés, que utilizaban en la escuela. La segunda parte abordaba el hecho de que los niños crecían simultáneamente en dos mundos, el medio cultural africano y el medio educativo. Era la cultura nativa aquella en la que los niños desarrollaban su comprensión y organización del significado; mientras tanto, en el medio educativo se utilizaba otra lengua cuya forma y contenido implicaban muchas realidades totalmente ajenas a los estudiantes. En la tercera parte, el autor comparó, en primer lugar, los datos en francés obtenidos de los niños de Camerún con datos similares de la investigación de Gougenheim y sus colegas; luego cotejó el repertorio léxico de los niños cameruneses en su lengua materna con el obtenido en francés por los mismos niños (Njock, 1978: 1).

Njock trabajó con una muestra que componían 220 alumnos que estudiaban en la escuela primaria. Escogió a los informantes de dos zonas de habla basaa de Camerún, teniendo en cuenta el porcentaje de población escolar de cada distrito para que la muestra fuera lo más representativa posible. Con esta misma finalidad, hizo que los alumnos encuestados se distribuyeran en proporción según el sexo y el tipo de enseñanza —en su caso, enseñanza pública, privada o religiosa (católica o protestante)—. El nivel de educación y la edad eran otros dos factores que consideró el investigador, pero afirmó que se entendían como dos factores distintos, e incluso poco relacionados en el

²⁰ La traducción al español del título de esta obra es: *El entorno familiar de los niños africanos*.

entorno africano. Es que, a diferencia de los informantes en las investigaciones precedentes, que, en términos generales, ingresaban a la escuela a una edad fija y ascendían a un nivel más alto al finalizar el curso anterior, la situación en Camerún no era la misma. Con motivo del alto índice de abandono escolar y de repetición de curso, la edad de los alumnos pertenecientes a un mismo nivel de educación podía oscilar en un rango muy amplio. De todas maneras, los informantes estaban distribuidos proporcionalmente según cada uno de estos dos factores.

Los centros de interés que utilizó Njock en las encuestas en francés eran 1. *Les parties du corps*, 2. *La nourriture et les boissons*, 3. *Les vêtements*, 4. *La maison: matériaux de construction et meubles*, 5. *Les travaux de champs*, 6. *La cuisine*, 7. *La salle de classe*, 8. *Les moyens de locomotion*, 9. *La musique et la danse-genres* y 10. *Les jeux et les divertissements*. En las encuestas en basaa se reproducían los mismos centros. Se puede ver que la mayoría de los ámbitos semánticos eran tomados de los trabajos de los pioneros franceses, mientras que, con un nuevo centro —*Les jeux et les divertissements*— el autor pretendía reflejar las características de la tierra africana. Las encuestas en ambas lenguas fueron completadas por los mismos alumnos con un intervalo de dos semanas. En todas las pruebas, se les concedía a los alumnos quince minutos para cada una de las áreas semánticas (Serrano Zapata, 2014: 109).

Tras elaborar las listas de palabras con su disponibilidad léxica, Njock inició otros procedimientos. Por un lado, investigó el porcentaje de informantes que habían mencionado un mismo concepto en las dos lenguas; y, por otro lado, agrupó las palabras obtenidas según el concepto y comparó el grado de disponibilidad de cada concepto en una lengua con su grado de disponibilidad en la otra lengua. Así, se podía observar cuál de las dos lenguas era la que solían usar los niños africanos para describir su entorno familiar. También comparó la disponibilidad de estos conceptos con la de aquellos recogidos en las investigaciones anteriores en Francia. Como resultado, el autor llegó a la conclusión de que 1) el francés de Francia no podía describir exactamente la

sociedad y cultura propias africanas, 2) la lengua africana, por falta de palabras nuevas, no era capaz de describir el África moderna, pero 3) el bilingüismo en francés y basaa posibilitaba el enriquecimiento del vocabulario para reflejar verbalmente diversos aspectos de la realidad africana.

3.2. Estudios de base de disponibilidad léxica en el mundo hispánico

3.2.1. El Proyecto Panhispánico de Disponibilidad léxica

Como puede verse en el apartado anterior, las investigaciones pioneras que siguieron el primer estudio de disponibilidad léxica en Francia propusieron, por su parte, modificaciones metodológicas con la finalidad de adaptarse a sus respectivos objetivos de investigación, que se centraban o bien en la selección del vocabulario, la enseñanza de lengua y el nivel de anglicismos, o bien en la comparación entre dos culturas y el análisis sociolingüístico en contexto de bilingüismo o de lenguas en contacto.

La introducción de la disponibilidad léxica en el mundo hispánico se debe a Humberto López Morales, quien, al igual que los investigadores de otras lenguas citados en el apartado anterior, emprendió su recogida empírica de materiales léxicos siguiendo las premisas metodológicas de los estudiosos franceses y canadienses. Fue en 1973, año en que el estudioso hispanoamericano inició el primer estudio del léxico disponible en español, circunscrito al área metropolitana de San Juan de Puerto Rico. Este paso dado por López Morales abrió un camino nuevo en el ámbito hispánico porque hizo que los demás estudiosos hispanistas vieran una línea de investigación emergente. Por consiguiente, comenzaron a surgir distintas investigaciones que suponían interesantes aplicaciones de la disponibilidad léxica en la lengua española.

Sin embargo, como había sucedido en los trabajos pioneros relacionados con la tradición franco-canadiense, estas investigaciones sobre el español

también propusieron, por su parte, modificaciones metodológicas y, por lo tanto, presentaron diferencias considerables en la metodología. De este modo, resultaban imposibles comparaciones similares a, por ejemplo, las que pretendía Mackey entre los datos de Francia y los de Canadá. Este inconveniente en la comparación no es difícil de detectar si se imagina que las numerosas investigaciones estudian a informantes de distintas edades, enuncian unos centros de interés dispares u organizan las pruebas de formas divergentes.

De ahí que, más tarde, se ingenió un proyecto panhispánico de disponibilidad léxica, el cual proponía un método único de trabajo para que distintos equipos de investigación se ocuparan de diferentes comunidades de habla o zonas dialectales siguiendo los mismos patrones metodológicos. Es un proyecto que continúa, en la actualidad, en vías de expansión. El proyecto ha sido coordinado por el propio Humberto López Morales, y su objetivo consiste en

“elaborar diccionarios de disponibilidad léxica para las diversas zonas del mundo hispánico, de manera que la homogeneidad de criterios permitirá establecer comparaciones de tipo lingüístico, etnográfico y cultural, dibujar áreas de difusión y, en general, servirá de punto de partida para análisis posteriores”²¹.

En el sitio web del grupo de investigación *Dispolex. Investigación léxica* —<http://www.dispolex.com>—, patrocinado por la Universidad de Salamanca y principal medio de difusión del *Proyecto Panhispánico*, se puede obtener estas informaciones y, además, la bibliografía de las investigaciones realizadas y el directorio de las diversas zonas de estudio, así como el acceso a la base de datos y al programa informático de análisis del léxico disponible. El proyecto reúne bajo una misma metodología un conjunto de investigaciones tanto de España como de Hispanoamérica. Desde entonces, el español se ha

²¹ Extraído del apartado “¿Qué es el Proyecto Panhispánico?” en Bartol Hernández, José Antonio & Hernández Muñoz, Natividad (2003-2020). *Dispolex / Investigación léxica*. Universidad de Salamanca. <http://www.dispolex.com/>.

convertido en la lengua que mejor aprovecha los métodos de la disponibilidad léxica, si bien los primeros trabajos de ella abordaron principalmente el francés y el inglés.

Como se cita en la misma página web, las fases para elaborar cada diccionario de léxico disponible son las siguientes:

- 1) selección de la muestra;
- 2) recogida de materiales mediante encuesta;
- 3) transcripción de los materiales y su homogeneización mediante los protocolos de edición convenidos;
- 4) cálculo de la disponibilidad léxica global y por campos y grupos de informantes, lo que llevará a configurar los diversos apartados del diccionario; y
- 5) publicación del diccionario.

Los principales criterios metodológicos que siguen todas las investigaciones que se inscriben en el proyecto son los siguientes:

- 1) los informantes son estudiantes del último curso del bachillerato, que precede al ingreso en la universidad, a fin de que los resultados no se vean contaminados por los tecnicismos adquiridos en el nivel universitario, los cuales se alejan de la norma de la comunidad;
- 2) se trabaja con los dieciséis centros de interés –temáticas de conversación– que se utilizaron en las primeras investigaciones francesas, aunque no hay impedimento en introducir otros adicionales;
- 3) los informantes responden por escrito a la encuesta y se conceden dos minutos por cada centro de interés para que escriban cuantas respuestas quieran;
- 4) se mantiene homogénea la edad y el nivel cultural de los informantes y se los subcategoriza de acuerdo con las variables sexo, zona geográfica, tipo de centro y nivel sociocultural;
- 5) se adoptan las mismas normas de edición; y

- 6) se usa el mismo programa informático para calcular la disponibilidad léxica.

3.2.2. Estudios fundamentales de disponibilidad léxica

3.2.2.1. Estudios fundamentales en Hispanoamérica

3.2.2.1.1. En su primer estudio de 1973, López Morales intentó analizar sociolingüísticamente el léxico disponible que recopiló del área metropolitana de San Juan de Puerto Rico. Tomó en cuenta la variable socioeconómica — con las variantes estrato bajo y estrato medio—, la variable sexo y la variable nivel de estudios —con las variantes primero, tercero y quinto grados de la escuela elemental—. Trataba de observar la influencia de las variables en la adquisición del léxico y, asimismo, ver la evolución que se manifestaba en la misma. En este primer estudio, se emplearon solamente diez centros de interés²². Más tarde, en 1978, López Morales comparó sus resultados del léxico disponible del año 1973 con las listas de palabras frecuentes que Ismael Rodríguez Bou había recolectado también en Puerto Rico²³. Estas comparaciones permitieron ver que el 47.25%²⁴ de los vocablos que conformaban el léxico disponible de los escolares puertorriqueños no estaban en dichas listas de palabras frecuentes. De modo que se volvió a comprobar que el criterio de los índices de frecuencia no alcanzaba exitosamente todos los campos semánticos del léxico, y no reflejaba el dominio léxico de los hablantes medios. En un trabajo posterior de 1979, López Morales retomó los datos de su primera investigación en San Juan de Puerto Rico, comparando, esta vez, los resultados en función de las variantes de la variable socioeconómica. En consecuencia, se dio cuenta de que existían diferencias en el léxico disponible entre los alumnos de niveles socioeconómicos distintos,

²² Los diez centros de interés en la investigación de 1973 eran: Alimentos, Animales, Muebles, Transporte, Ropa, Partes del cuerpo, La casa, Materiales y herramientas, La cocina y La naturaleza.

²³ Se refieren a las listas de palabras frecuentes incluidas en *Recuento de vocabulario de preescolares*, de Rodríguez Bou, publicado en 1966.

²⁴ Bellón (2011) afirma que, del total de 673 vocablos aportados por la disponibilidad léxica en la investigación de López Morales, 318 no aparecían en el recuento de Rodríguez Bou.

y las diferencias se daban tanto en el sentido cuantitativo como en el cualitativo. Estadísticamente, observó que los informantes de estrato socioeconómico medio mostraban cifras más altas que los de estrato bajo. Este descubrimiento demostró que había una correlación importante entre la disponibilidad léxica y la variable socioeconómica.

En estudios posteriores, López Morales dio importancia a la disponibilidad léxica en lo que se refiere a ámbitos como la enseñanza de la lengua materna²⁵. Fue precisamente este enfoque en la lengua materna el que condujo a la iniciativa del *Proyecto Panhispánico de Disponibilidad léxica*. Bajo este proyecto, que coordinaba él mismo, en el año 1999, Humberto López Morales publicó el *Léxico disponible de Puerto Rico*²⁶. Pasó encuestas a estudiantes procedentes de ocho universidades del país y finalmente 558 encuestas —es decir, 558 informantes después de suprimir las encuestas defectuosas y los informantes que no cumplían con las condiciones establecidas— integraron la muestra del corpus del léxico disponible de Puerto Rico. Todos los informantes cursaban el primer año universitario, lo que parecía no coincidir con el criterio metodológico del *Proyecto Panhispánico*, puesto que este había establecido que se estudiara a informantes del último curso del bachillerato para que los resultados no se vieran contaminados por los tecnicismos adquiridos en el nivel universitario. Pero, en realidad, la muestra de López Morales no constituyó un caso contradictorio, porque la enseñanza universitaria de Puerto Rico estipulaba que los alumnos del primer curso realizaran estudios generales preparatorios antes de ingresar a estudiar en el segundo año en las diferentes facultades. Por tanto, los datos que recogió el investigador no se vieron afectados por los tecnicismos relacionados con las especialidades que iban a cursar y, por consiguiente, no se alejó del panorama léxico de la comunidad puertorriqueña.

²⁵ López Morales abordó la disponibilidad léxica en la enseñanza de la lengua materna en su publicación del año 1986, *La enseñanza de la lengua materna: lingüística para maestros de español*.

²⁶ López Morales ya inició en 1987 la elaboración del léxico disponible de Puerto Rico. Pudo publicarlo finalmente en 1999 después de la encuesta, la edición, codificación y procesamiento electrónico de los materiales, y la demora de la imprenta.

Como estudio fundamental adherido al *Proyecto Panhispánico*, el *Léxico disponible de Puerto Rico* servía de punto de referencia para los criterios metodológicos establecidos. Esto lo podemos comprobar en los centros de interés, la forma de las listas, el control de tiempo y las variables sociales que aplicó López Morales en sus pruebas, todo ello convertido en modelo para las investigaciones posteriores que formaran parte del proyecto. Tal y como hablamos del papel ejemplar de la obra, en el diseño de la prueba asociativa, López Morales se aprovechó de los mismos dieciséis centros de interés que habían empleado los estudiosos franceses, aunque tuvo que introducir dos pequeñas modificaciones en el nombre de dos centros. Es completamente comprensible, porque solo sustituyó “escuela” por “universidad” y “calefacción” por “aire acondicionado” de manera que estuvieran conformes a las circunstancias de los informantes puertorriqueños —estudiantes del primer curso universitario y acostumbrados a un clima tropical—.

La decisión que López Morales tomó sobre la forma de las listas fue utilizar listas abiertas. Estas habían sido empleadas, por primera vez, por Naum R. Dimitrijević, tal y como explicó su metodología en la obra *Lexical availability: A new aspect of the lexical availability of secondary school children*. Al igual que Dimitrijević, López Morales pensaba recoger la mayor cantidad posible de palabras disponibles aportadas por los informantes, de modo que no pidió a los alumnos un número fijo de palabras y abandonó las listas cerradas de veinte palabras de las pruebas de Gougenheim y Mackey. Sin embargo, el control de tiempo correspondiente a cada lista en las pruebas de López Morales era distinto al de Dimitrijević²⁷. El periodo de tiempo cronometrado concedido a cada centro de interés —durante el cual los informantes podían apuntar, en orden, en una lista abierta las palabras que se les ocurrieran— era una cuestión discutible. Según López Morales, había de realizarse una prueba experimental, la hizo y los resultados de la misma sugirieron que el tiempo atribuido a cada centro de interés no fuera superior a

²⁷ Dimitrijević (1969) fijó en cinco minutos el tiempo de respuesta correspondiente a cada centro de interés.

dos minutos. Esto recuerda la prueba previa que había hecho John Bailey Victory antes de las formalmente realizadas para su investigación, procedimiento que explicó en *A study of lexical availability among monolingual-bilingual speakers of Spanish and English* (Victory, 1971). Como ya hemos visto, este investigador aseguró que, después de aproximadamente dos minutos en cada centro de interés, la mayoría de los participantes en la prueba previa iban cada vez más lentos en la producción de palabras, es decir, disminuía la disponibilidad léxica. Se puede ver que López Morales y Victory diseñaron experimentos similares y coincidieron en la idea de limitar a dos minutos el tiempo de respuesta para cada centro de interés.

Las variables sociales, con las que López Morales subcategorizó a los informantes, eran cuatro —tipo de escuela, sexo, procedencia geográfica y nivel sociocultural—. El tipo de escuela era pública, privada o combinada. En cuanto a la procedencia geográfica, se tomaron cuatro zonas predeterminadas del país. El nivel sociocultural se estratificó en medio-alto, medio, medio-bajo y bajo. A los informantes se les atribuía un nivel sociocultural concreto en función de las puntuaciones que obtenían de tres factores: la escolaridad²⁸, la profesión²⁹ y los ingresos³⁰ de los padres.

Mientras determinó las bases metodológicas, López Morales pudo recopilar el léxico disponible de Puerto Rico. Estando así las cosas, el autor expresó, al acabar la investigación, que esperaba que sus resultados, junto con el léxico básico ya establecido en un estudio anterior³¹, pudieran servir de

²⁸ La escolaridad se puntuó según su posición en la siguiente escala: 1. No ha ido a la escuela; 2. Tiene algunos años de escuela elemental; 3. Completó la escuela elemental; 4. Tiene años de escuela intermedia y / o superior; 5. Terminó la escuela superior; 6. Tiene años de universidad; 7. Terminó el bachillerato; y 8. Tiene un título graduado.

²⁹ La profesión fue codificada conforme a la especificación de las siguientes posibilidades: 1. Obreros sin cualificar; 2. Obreros con cualificación; 3. Empleados medios; 4. Pequeños empresarios / autónomos; 5. Medianos empresarios; 6. Profesionales liberales; y 7. Altos directivos / grandes empresarios.

³⁰ El ingreso familiar se valoró de acuerdo con la siguiente escala: 1. Menos de 10.000 dólares; 2. Entre 10.000 y 14.999 dólares; 3. Entre 15.000 y 19.999 dólares; 4. Entre 20.000 y 24.999 dólares; 5. Entre 25.000 y 29.999 dólares; 6. Entre 30.000 y 39.999 dólares; 7. Entre 40.000 y 49.999 dólares; y 8. Más de 50.000 dólares.

³¹ Se refiere al *Léxico básico del español de Puerto Rico* de Amparo Morales (1986).

material fundamental para la planificación léxica de la enseñanza del español en Puerto Rico. Aprovechando el corpus del *Léxico disponible de Puerto Rico*, López Morales llevó a cabo, en el mismo 1999, un estudio que analizaba los anglicismos presentes en los datos recogidos. En este estudio, se concluyó que la presencia de las palabras inglesas en el corpus era baja; además, se daba una clara diferencia entre los centros de interés: había unos que se mostraban reacios a la incorporación de los anglicismos y había otros tantos donde se registraban sus apariciones.

En territorio hispanoamericano, hay muchas otras investigaciones que han enriquecido el estudio de la disponibilidad léxica; sin embargo, algunas de las aportaciones más interesantes se han quedado fuera o han ido más allá del marco del *Proyecto Panhispánico*. Estas investigaciones o bien han profundizado, en otros sentidos, los estudios sobre la disponibilidad léxica³²; o bien han relacionado la disponibilidad léxica con otros campos de estudio³³; o bien han seguido métodos propios, diferentes de los del *Proyecto Panhispánico*, al indagar en el léxico disponible³⁴. En Puerto Rico, de estas características, se han destacado todavía los siguientes estudiosos.

En 1985, Belén Román estudió la disponibilidad léxica en escolares de Dorado, Puerto Rico. Siguió la misma metodología que había empleado López Morales en 1973 sobre el léxico disponible del área metropolitana de San Juan de Puerto Rico, porque se basó en la idea de hacer posteriores comparaciones entre las dos zonas geográficas. Es decir, Román trabajó con los mismos diez centros de interés y utilizó idénticas variables sociales: nivel socioeconómico —estratos medio y bajo—, sexo y nivel de estudio —primero, tercero y quinto

³² Por ejemplo, las investigaciones de búsqueda de una fórmula para calcular el grado de disponibilidad hechas por Roberto Lorán Santos y Humberto López Morales (1983); y las posteriores hechas por Juan López Chávez y Carlos Strassburger Frías (1991 y 2000) para mejorar el cálculo del índice de disponibilidad.

³³ Han sido muchos los estudios que utilizan los datos del léxico disponible para nutrir investigaciones de, por ejemplo, la dialectología, la etnolingüística, la psicolingüística, etc. Los estudios que hizo López Morales en contacto con el anglicismo o la sociolingüística eran investigaciones de esta índole.

³⁴ Un ejemplo de este tipo de estudio es el que hizo López Morales en 1973 sobre el léxico disponible del área metropolitana de San Juan de Puerto Rico, tal y como se ha especificado en los apartados anteriores.

grados de la escuela elemental—. También analizó, por su parte, la influencia de estas variables en la adquisición del léxico. Como resultado, confirmó que las chicas tenían una cifra media superior a los chicos; los informantes de zona urbana consiguieron mayor disponibilidad léxica que los de zona rural; y la disponibilidad léxica incrementaba a medida que aumentaba el nivel de estudio. Además, su observación sobre la relación entre el nivel socioeconómico y la disponibilidad léxica dio resultados dispares en comparación con los de López Morales. En la investigación de Román, los estudiantes de nivel socioeconómico bajo sobrepasaron a los de nivel medio, lo cual tampoco coincidía con las expectativas de la propia investigadora.

El cálculo de la disponibilidad léxica era otro problema que preocupaba a los estudiosos. Al principio, se utilizó la frecuencia de aparición de las unidades léxicas para representar su disponibilidad. Charles Müller (1973) fue el primero en advertir que había que tener en cuenta el lugar de aparición de las unidades léxicas, puesto que las palabras que se actualizaban en los primeros puestos de las listas no tenían el mismo grado de disponibilidad que las que se actualizaban en los últimos lugares. En 1983, los investigadores portorriqueños Roberto Lorán Santos y Humberto López Morales propusieron una fórmula matemática que ponderaba la frecuencia y el orden en que aparecía cada palabra. Fue la primera fórmula matemática que permitía calcular el índice de disponibilidad y fue aplicada en varias investigaciones de la época. Otra investigadora portorriqueña, Gloria Butrón, también mostró su interés en el estudio de la fórmula matemática para la disponibilidad. Hizo, en 1987, una tesis doctoral sobre el cálculo del índice de disponibilidad, analizando el valor de diferentes fórmulas y publicó, en 1991, un trabajo que propuso una nueva aplicación para el cálculo de la disponibilidad.

Entre las investigaciones de Butrón, todavía se destaca otra en 1989 que estudiaba la repercusión de las variables sociales. Analizó ocho centros de interés³⁵ y consideró tres variables sociales: nivel socioeconómico, sexo y

³⁵ Los ocho centros de interés que analizó Butrón fueron *Alimentos, Animales, La casa, Ropa, Naturaleza, Colores, El cuerpo humano y Juegos y entretenimientos*.

zona geográfica. Su punto de partida era demostrar la utilidad de los instrumentos cuantitativos en el estudio del léxico y concluyó que las comparaciones sociolingüísticas mostraron, efectivamente, diferencias cuantitativas relevantes. Lo más importante era que las cifras variaban en gran medida según el nivel socioeconómico y la zona urbana llevaba una gran ventaja sobre la zona rural.

3.2.2.1.2. En la República Dominicana, los estudios de disponibilidad léxica del *Proyecto Panhispánico* los llevó a cabo Orlando Alba, de los cuales la publicación más importante fue *El léxico disponible de la República Dominicana* en 1995 (Alba, 1995b). En esta obra suya, además de recaudar el léxico disponible del país caribeño, Alba tenía también como objetivo el análisis de correlación entre la disponibilidad léxica y las tres variables sociales que tomó en cuenta, a saber, nivel sociocultural, sexo y tipo de escuela. Pensada así la investigación, estratificó a los informantes —un total de 347 estudiantes dominicanos provenientes de 11 universidades en 4 ciudades del país— en función de dichas variables, distinguiendo entre ellos el nivel sociocultural alto, medio y bajo; el sexo femenino y masculino; y el tipo de escuela pública y privada. La asignación de las tres opciones del nivel sociocultural se basa en la puntuación que se hizo observando tres parámetros cuantitativos: la profesión u oficio de los padres³⁶, el grado de educación de los mismos³⁷ y los ingresos familiares³⁸.

Como investigación adherida al *Proyecto Panhispánico*, *El léxico disponible de la República Dominicana* siguió su metodología en la aplicación de las pruebas asociativas del léxico. Adicionalmente, Alba justificó la técnica de fijar en dos minutos la asociación de palabras para cada centro de interés.

³⁶ Para la profesión u oficio de los padres, se asignó una puntuación a cada informante de acuerdo con las siguientes categorías que se tomaron en cuenta: 1. Profesionales, gerente, ejecutivo, empresario; 2. Técnico (contador, secretaria...); 3. Comerciante, empleado; y 4. Ama de casa, obrero no especializado.

³⁷ El grado de educación de los padres fue valorado según los siguientes cuatro niveles: 1. Educación universitaria; 2. Educación secundaria o técnica; 3. Educación primaria; y 4. Ninguna (analfabeto).

³⁸ La puntuación del ingreso familiar se midió conforme a una escala que iba desde 2.000 pesos mensuales o menos —alrededor del salario mínimo— hasta 20.000 pesos o más.

En este aspecto, su consideración era igual que la de Victory (1971) y López Morales (1999b), que ya hemos comentado. Alba afirmó que el objetivo de la prueba era incitar la formulación de palabras realmente disponibles de los estudiantes; de modo que, si no hubiera un límite de tiempo o se les concediera un tiempo excesivo, sería una exploración mental del vocabulario y los informantes buscarían en la memoria palabras que no estaban inmediatamente disponibles. Aparte de ello, con respecto al uso de los dieciséis centros de interés que estrenaron Gougenheim et al. (1964), Alba planteó sus dudas, aunque estaba conforme con su empleo. Por una parte, estaba de acuerdo con Mackey (1971) cuando este señaló que dichos centros de interés no podían cubrir todos los intereses humanos. Por otra parte, observó que algunos centros de interés mostraban una definición semántica bien enmarcada, como *Partes del cuerpo humano* o *Los animales*, mientras que había otros centros con fronteras bastante imprecisas y difusas, como *La ciudad* o *El campo*. Una inadecuación de esto último resultaba ser que una palabra se repetía en distintos centros de interés³⁹.

Entre otros resultados cuantitativos, Alba publicó las cifras del total de palabras y de palabras diferentes (vocablos) por centro de interés. Llamó la atención sobre la errónea correlación que se podía imponer entre el número de palabras diferentes en los centros de interés y la disponibilidad léxica de los informantes en los mismos. Explicó que una ojeada a las cifras de palabras diferentes por centro de interés podía llevar a la idea inadecuada de que, según el número de palabras diferentes, unos centros eran más productivos que otros —o, lo que viene a ser lo mismo, los estudiantes tenían mayor disponibilidad en unos centros que en otros—. Esta sería una conclusión incorrecta si entre los informantes no había una coincidencia mayor de respuestas, puesto que la cantidad de palabras diferentes tenía su origen en la variedad o disparidad de respuestas y no necesariamente en la productividad de cada centro de interés. Para aclarar la cuestión, Alba planteó un caso hipotético: en un grupo de 10

³⁹ Alba citó como ejemplo la aparición de “casa” y “escuela” tanto en *La ciudad* como en *El campo*, y se hallaban “nevera” y “estufa” al mismo tiempo en *Los muebles de la casa* y en *La cocina y sus utensilios*.

informantes, cada uno escribía 5 palabras iguales; en otro grupo igual de 10 informantes, cada uno escribía 1 palabra diferente. Las palabras diferentes del primer grupo eran 5 mientras las del segundo eran 10, pero era obvio que el primero sobrepasaba al segundo en cuanto a la disponibilidad léxica, a pesar de que era inferior respecto a las palabras diferentes.

Como podemos ver en las líneas anteriores, la falta de coincidencia —en otras palabras, la disparidad o variedad— en las respuestas por los informantes influía directamente en las palabras diferentes. Dada su importancia, Alba propuso dos formas sobre el cálculo de coincidencia, que permitía ver el mayor o menor grado de conformidad en las respuestas. Por una parte, señaló la *densidad léxica* de cada centro de interés, dividiendo el total de palabras de un centro por el número de palabras diferentes del mismo. Por otra parte, presentó el *índice de cohesión* propuesto por Max Sergio Echeverría (1987), que resultaba de dividir el promedio de respuestas por informante en cada centro de interés por el número de palabras diferentes del mismo centro. La ventaja del índice de cohesión —que es más utilizado en la actualidad— consistía en que el resultado del cálculo sería 1 en el caso de que todos los informantes escribieran las mismas palabras en un centro de interés, lo cual facilitaba bastante la visualización de los resultados cuantitativos. Como ejemplo, en la investigación en cuestión de Alba, *Partes del cuerpo humano* era el centro de interés de mayor densidad y cohesión —con una densidad léxica de 32 y un índice de cohesión de 0,09⁴⁰—, razón por la cual se entendía que era el centro más compacto (cerrado) o, lo que es lo mismo, el menos difuso (abierto).

De todos modos, lo que realmente interesaba era en qué centros de interés los estudiantes tenían mayor disponibilidad léxica. Alba puso de manifiesto que la cifra más reveladora era el *promedio de palabras por informante* en cada centro de interés. En función de este criterio, por ejemplo,

⁴⁰ En la investigación de Alba, el centro de interés *Partes del cuerpo humano* registró 7992 respuestas; entre estas, había 246 palabras diferentes, y el promedio de respuesta por informante en este centro era 23.

Alimentos y Animales, ambos con un promedio de 24 palabras, eran los centros en los que los informantes tenían mayor disponibilidad léxica⁴¹. Para comprobar el papel clave del promedio de palabras por informante, el investigador estableció comparaciones entre los estudiantes de nivel sociocultural alto, los de nivel medio y los de nivel bajo, por una parte, según el promedio de respuestas y, por otra, según la densidad —la que cuantitativamente equivalía al índice de cohesión—. La hipótesis era que existía una relación directamente proporcional entre la disponibilidad léxica y el nivel sociocultural. El resultado de las comparaciones fue que una relación cuantitativa proporcional se daba cuando se hacían cotejos según el promedio de palabras por informante y no según la densidad; los cotejos desde la perspectiva del primero señalaron variaciones bastante regulares y los de la segunda, nada regulares. Por lo tanto, se observó que el promedio de palabras por informante era la cifra que realmente reflejaba la disponibilidad léxica, esto es, “el dato más realista y confiable para la determinación de la disponibilidad en términos cuantitativos” (Alba, 1995b: 22).

Determinada así la importancia reveladora del promedio de palabras por informante, el investigador empezó su análisis de correlación entre la disponibilidad léxica y los demás factores sociales, a través de dicho criterio. En cuanto al factor sexo, se notó que las chicas superaban a los chicos de manera sistemática en todos los centros de interés, pero cabe mencionar que la ventaja que llevaban ellas sobre ellos no resultaba muy notable. Esto último se mostró en que la diferencia entre la cifra de las chicas y la de los chicos se reducía a una o dos palabras en todos los centros. Sin embargo, las comparaciones que se hicieron con respecto al tipo de escuela resultaron, estadísticamente, bastante significativas. Las cifras según este factor social mostraron una regularidad cuantitativa notoria, tal y como se señaló con

⁴¹ La clasificación de los centros de interés en *El léxico disponible en la República Dominicana* según el promedio de palabras por informante es la siguiente: 1. centros de alta productividad: *Alimentos, Animales y El cuerpo humano*; 2. centros de productividad intermedia: *El vestido, La casa, Los muebles de la casa, La cocina, La escuela, La ciudad, El campo, Medios de transporte, y Juegos y profesiones*; y 3. centros de baja productividad: *Objetos colocados sobre la mesa para la comida, Iluminación y aire acondicionado, y Trabajos del campo y del jardín*.

respecto al nivel sociocultural. Las escuelas privadas parecían tener una superioridad estadística sobre las públicas en todos los centros de interés. Se observó un promedio de diferencia de cuatro palabras por centro, cuando se compararon los promedios de palabras por informante entre los dos tipos de escuela.

Antes de finalizar la obra, Alba reiteró el punto de vista que se había tomado para calcular el índice de disponibilidad, en comparación con la mera frecuencia de aparición. Era la posición que ocupaba una palabra en la lista de su centro de interés la que incidía en el cálculo. Como resultado general y particular de la investigación, el autor adjuntó el léxico disponible de acuerdo con el orden del índice de disponibilidad y el léxico disponible según el sexo, según el tipo de escuela y según el nivel sociocultural.

Además de ocuparse del léxico disponible de la República Dominicana, Alba también se interesó por el anglicismo que aparecía en las palabras que utilizaban los dominicanos. Llevó a cabo dos investigaciones al respecto. La primera la hizo en 1995 (Alba, 1995a). Se trata de un análisis de los anglicismos en un corpus construido con dos tipos de fuente léxica. La primera fuente eran unas conversaciones libres y la segunda eran las palabras disponibles en seis centros de interés actualizadas por 65 informantes⁴². Se comprobó que los sujetos de nivel sociocultural alto empleaban un mayor número de anglicismos. A pesar de ello, resultó evidente que la variable que influía más en la aparición del anglicismo era el centro de interés. Esto quiere decir que había centros que se mostraban reacios a la incorporación de anglicismos y había otros, en cambio, más propensos a registrarlos. Pero, en general, la frecuencia de aparición de los anglicismos era baja, lo que coincide con el estudio de López Morales (1999a), como se ha demostrado en el apartado anterior.

⁴² El investigador se aprovechó de una pequeña muestra de 65 sujetos en los siguientes centros de interés: *Alimentos, Juegos y diversiones, La escuela, Medios de transporte, Profesiones y oficios, El vestido.*

La segunda investigación de Alba sobre el anglicismo tuvo lugar en 1999 cuando retomó los datos del *Léxico disponible de la República Dominicana*. Se enfocó en la densidad de las palabras de origen inglés y en la correlación de esta con dos variables sociales: el sexo y el nivel sociocultural. Los resultados no se alejaron mucho de los ya obtenidos en su propia investigación de la misma índole en 1995 ni de los de López Morales en 1999 sobre el mismo tema. Unos centros eran receptivos a palabras originaras del inglés y otros eran refractarios a dichos vocablos. Esta vez, aprovechándose de los datos del *Léxico disponible de la República Dominicana*, el investigador pudo observar que los primeros centros eran aquellos semánticamente abiertos que trataban de actividades sociales como *Medios de transporte, Juegos y diversiones, El vestido y Los muebles de la casa*, porque en estas áreas el predominio anglosajón era incuestionable. Mientras tanto, los centros que no promovían el anglicismo eran aquellos semánticamente cerrados o de inventario léxico limitado como *Las partes del cuerpo humano y Animales*.

En lo concerniente a las variables sociales, Alba afirmó que el sexo ejercía poco efecto en el uso de las palabras procedentes del inglés. El nivel sociocultural sí influía significativamente en la densidad de los anglicismos, ya que el porcentaje de estos en los informantes de nivel sociocultural alto era más elevado que en los de nivel bajo. De todas maneras, el autor reiteró que la variable más decisiva en el análisis del anglicismo era el centro de interés, tal y como coincide con otros estudios de este tema.

Al igual que se aprovechó de los datos del *Léxico disponible de la República Dominicana* para nutrir el estudio del anglicismo, Alba (1998) dedicó los mismos datos léxicos a otro tipo de investigación: la dialectología. En esta ocasión pretendió comparar los dialectos de cinco regiones de habla hispana: la República Dominicana, Puerto Rico, Concepción (Chile), Madrid y Ciudad de México⁴³. Su método de estudio consistía en comparar los vocablos de alto

⁴³ Los datos comparados en este estudio son los siguientes: el léxico de la República Dominicana es el elaborado por el mismo Orlando Alba; el de Puerto Rico, por Humberto López Morales; el de Concepción (Chile), por Mónica Mena Osorio; el de Madrid, por Pedro Benítez Pérez y el de México corresponde al léxico del nivel escolar “sexto grado”

grado de disponibilidad en las cinco regiones. El investigador no manejó todos los datos que quedaban a su disposición, sino que optó por tres centros de interés —*El cuerpo humano*, *Medios de transporte*, *Alimentos*— y analizó solamente las primeras 50 palabras. Alba revisó las 250 palabras de cada uno de los tres centros⁴⁴, extrajo de ellas las palabras diferentes, en estas buscó las que aparecían en las primeras 50 palabras de todas las regiones, y calculó el porcentaje de estas palabras comunes a los cinco dialectos entre el número total de palabras diferentes anteriormente extraídas. Los resultados revelaron que *El cuerpo humano* contaba con el mayor porcentaje de vocablos comunes mientras *Alimentos* registraba la menor cifra. Por lo tanto, en opinión del investigador dominicano, se manifestaban correlativos el grado de cohesión de un centro de interés y su nivel de compatibilidad dialectal. Dicho de otro modo, el centro de interés, semánticamente, más compacto o cerrado es asimismo el centro más compatible o coincidente en el sentido dialectológico, como era el caso de *El cuerpo humano* en esta investigación; y el centro de interés semánticamente más difuso o abierto es también el centro con más diversidad dialectal, tal y como se presentaba en *Alimentos*.

A pesar de que entre los centros de interés había diferencias con respecto al grado de compatibilidad dialectal, Orlando Alba destacó que, en general, los resultados mostraron una compatibilidad léxica bastante considerable entre los dialectos. En sus observaciones sobre los vocablos comunes, el autor descubrió que un vocablo que no sea común a los cinco dialectos tampoco es propio de una región dialectal. Es decir, eran muy escasas las palabras exclusivas de una sola región. Encima, en la mayoría de los casos, los vocablos que solo se presentaban en algunos dialectos eran en realidad conocidos también en otros, pero ocupaban lugares retrasados y no entraron en las primeras 50 palabras más disponibles. En fin, Alba advirtió un alto grado de compatibilidad dialectal, por lo general, del léxico de español.

de las investigaciones que coordinó Juan López Chávez en Ciudad de México.

⁴⁴ Estas 250 palabras eran la suma de las primeras 50 palabras en cada centro de interés de las 5 regiones estudiadas.

3.2.2.1.3. En el resto de Hispanoamérica, las tempranas investigaciones que contribuyeron a las bases metodológicas del estudio de disponibilidad léxica se enmarcaron fuera del *Proyecto Panhispánico*. En México, los investigadores más destacados y que incorporaron enormes aportaciones fueron Juan López Chávez y Carlos Strassburger Frías. Como ya se ha dicho, ellos colaboraron en proponer otro cálculo del índice de disponibilidad léxica, aunque ya existía la fórmula planteada en 1983 por López Morales y Lorán Santos. Lo que sucedía era que esta fórmula anteriormente empleada, si bien ponderó la frecuencia y el orden de aparición de las palabras, mostró defectos que se fueron comprobando. El propio artífice de la misma, López Morales, advirtió de que su fórmula perdía poder a partir de cierta posición de la palabra y que, aun con las mejoras que introdujo Gloria Butrón en 1987, el desajuste no fue completamente corregido (López Morales 1995: 247). Estando así las cosas, López Chávez y Strassburger Frías publicaron en 1987 una fórmula nuevamente diseñada en su artículo “Otro cálculo del índice de disponibilidad léxica”⁴⁵. Los defectos anteriores fueron eliminados en la formulación de los investigadores mexicanos, en la que contribuían al índice de disponibilidad varios factores: la frecuencia absoluta de la palabra, el número de informantes en las encuestas, el número de posiciones alcanzadas en el centro de interés, la posición de la palabra, entre otros. Posteriormente, López Chávez y Strassburger Frías continuaron en el camino del cálculo del índice de disponibilidad y publicaron, respectivamente en 1991 y en 2000, “Un modelo para el cálculo del índice de disponibilidad léxica individual” y “El diseño de una fórmula para obtener un índice de disponibilidad léxica confiable”. Su fórmula —véase la Figura 1— es la que están utilizando las investigaciones de hoy en día.

$$D(P_j) = \sum_{i=1}^n \left(\frac{f_{ij}}{l} \right) e^{-2.3 \left(\frac{i-1}{n-1} \right)}$$

⁴⁵ Según lo expuesto por Bellón Fernández (2011), Bellón Fernández, Samper Hernández, & Samper Padilla (2003), López Morales (1995) y Mateo García (1998). Los textos originales de este artículo no han sido accesibles para el autor de la tesis que aquí se presenta.

Figura 1. Fórmula del índice de disponibilidad
por López Chávez y Strassburger Frías (1991 y 2000)

En donde:

$D(P_j)$ = disponibilidad de la palabra j .

n = máxima posición alcanzada en el centro de interés.

l = número de posición de que se trata.

j = índice de la palabra en cuestión.

e = número natural (2.718181818459045).

f_{ij} = Frecuencia absoluta de la palabra j en la posición i .

i_1 = número de informantes que participaron en la encuesta.

Además de esta formidable contribución a la base metodológica del estudio de la disponibilidad léxica, López Chávez también se dedicó a otras aplicaciones en esta línea de investigación. En 1991, estudió los extranjerismos en el español de México en dos centros de interés: *Partes del cuerpo* y *Medios de transporte*. Comprobó que el primero se mostraba reacio al anglicismo mientras que el segundo daba cabida a muchos préstamos del inglés que ocupaban lugares importantes. La dialectología fue otra área en la que se metió el estudioso mexicano, puesto que en 1992 estableció comparaciones dialectales entre cuatro comunidades de habla de los dos lados del Océano Atlántico⁴⁶. En este trabajo, comparando los dieciséis centros de interés de las cuatro regiones y analizando sus índices de compatibilidad, López Chávez descubrió que la compatibilidad era baja, si bien las innegables coincidencias, con ciertas restricciones, apuntó a un “español atlántico” (López Chávez, 1992: 26). Además, igual que las investigaciones del *Proyecto Panhispánico* estudiaban el léxico disponible de una comunidad de habla determinada, López Chávez coordinó, en 1993, un gran trabajo sobre el léxico disponible de los escolares mexicanos. Colaboró con otros investigadores en la recogida del léxico disponible, en los dieciséis centros de interés clásicos, sobre alumnos de preescolar y de todos los cursos de primaria⁴⁷. Con estos

⁴⁶ Los léxicos disponibles comparados en este trabajo de López Chávez son los de Madrid (de Pedro Benítez Pérez), Las Palmas de Gran Canaria (de José Antonio Samper Padilla), la República Dominicana (de Orlando Alba) y Puerto Rico (de Humberto López Morales).

⁴⁷ López Chávez elaboró, 1) con Rosa María Mesa Canales, el léxico disponible de

datos léxicos pudo hacer un seguimiento del lexicón mental de los alumnos y estudiar, por consiguiente, la formación, la organización y el desarrollo del mismo (Ayora Esteban, 2006: 52; Bellón Fernández, 2011: 48).

Además de las importantes aportaciones que hizo López Chávez, México todavía contaba con otros tempranos estudios que, en los primeros momentos, pusieron en práctica la emergente teoría de la disponibilidad léxica. Cabe mencionar a tres investigadoras al respecto: Hortensia Guadalupe Justo Hernández, Alva Valentina Cañizal Arévalo y Araceli Ruiz Basto, autoras de sendas memorias de licenciatura sobre la disponibilidad léxica. La primera investigadora (Justo, 1987) estudió la disponibilidad léxica en colores, utilizando listas cerradas y diez centros de interés⁴⁸. Aportó la novedad de recoger adjetivos y pretendió buscar de qué manera las variables sociales — la ubicación de la escuela y la situación socioeconómica del informante— afectaban a los resultados de disponibilidad léxica en los colores. Cañizal (1987) recopiló el léxico disponible en informantes con los estudios de primaria terminados, utilizando seis centros de interés⁴⁹. Con este trabajo, la autora intentó observar la influencia de unas variables —sexo, tipo de centro, tiempo dedicado a la televisión— en la disponibilidad léxica. Ruiz Basto (1987) empleó los mismos seis centros de interés que los de Cañizal, pero trabajó con informantes de primer curso universitario, amplió el tiempo de acción a cinco minutos y consideró solamente dos variables: sexo y turno —matutino o vespertino—. La investigadora pudo concluir que entre las alumnas había mayor coincidencia en las palabras que entre los alumnos; asimismo los

preescolares mexicanos, 2) con María Trinidad Madrid Guillén, el de primer grado de primaria, 3) con Martha Julián Peña, el de segundo grado de primaria, 4) con Rosalía Bolfeta Montes de Oca, el de tercer grado de primaria, 5) con Marcela Flores Cervantes, el de cuarto grado de primaria, 6) con Lilia Castellanos Medina, el de quinto grado de primaria y 7) con Eva Núñez Alonso, el de sexto grado de primaria.

⁴⁸ Con el objetivo de estudiar las relaciones entre la ideología y la realidad en lo que se refiere al color, Justo Hernández empleó los siguientes diez centros de interés cercanos a la vida de los informantes: *Ropa; Zapatos; Cosméticos; Automóviles; Anuncios comerciales; Mochilas, bolsas, maletas; Muebles; Tapices, alfombras; Pintura de casa (exterior-interior) y Vestido.*

⁴⁹ Los seis centros de interés de Cañizal eran *Las partes del cuerpo, Objetos que hay en el interior de la casa, Alimentos y bebidas, Medios de transporte, Profesiones y oficios y Defectos y cualidades físicas y morales.* Con este último, introducido por primera vez, pensó en recabar adjetivos.

informantes de turno matutino tenían mejor resultados que los del vespertino.

En Chile, los estudios de disponibilidad léxica en los primeros momentos también fueron fructíferos, pues siguieron inmediatamente la primera investigación al respecto de López Morales. Empezaron en 1986 con la tesis de maestría de Mónica Mena Osorio, *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Esta investigadora reutilizó los diez centros de interés que había empleado el estudioso portorriqueño en 1973⁵⁰. Recabó el léxico disponible en 300 alumnos de educación básica, teniendo en cuenta el grado de escolaridad y el nivel sociocultural. Observó que 1) el vocabulario crecía progresivamente a medida que incrementaba el nivel escolar y 2) la superioridad de los informantes de nivel sociocultural medio-alto sobre los de nivel bajo era homogénea en todos los centros de interés.

Mientras Mena se ocupó de informantes de educación básica, el estudio sobre los alumnos de educación media lo llevaron a cabo Max Sergio Echeverría, María Olivia Herrera, Patricio Moreno y Francisco Pradenas. El equipo publicó en 1987 la *Disponibilidad léxica en Educación Media*, analizando respuestas aportadas por 400 alumnos de educación media de Concepción, al sudoeste de Chile. Igualmente, retomó los diez centros de interés propuestos por López Morales (1973) y estratificó a los informantes según el sexo, el grado escolar y el nivel sociocultural. Una vez más se comprobó que los alumnos de nivel sociocultural alto obtenían mejores resultados que los de nivel bajo. Llama la atención que el equipo planteara en este trabajo un nuevo índice que describía la fisonomía de los centros de interés. Se trata del “índice de cohesión”, cuya fórmula consistía en dividir el promedio de respuestas por informante en cada centro de interés por el número de palabras diferentes del mismo centro, con lo que medía la

⁵⁰ Humberto López Morales recopiló en 1973 el léxico disponible del área metropolitana de San Juan de Puerto Rico, con lo cual inició los estudios de disponibilidad léxica en el mundo hispánico. Recordemos que los diez centros de interés de esta investigación eran tenidos en cuenta por López Morales, en esta ocasión, fueron: *Alimentos, Animales, Muebles, Transporte, Ropa, Partes del cuerpo, La casa, Materiales y herramientas, La cocina, y La naturaleza*.

homogeneidad en las respuestas dadas por los informantes en un mismo centro de interés. De este modo, era capaz de describir si un centro de interés era compacto (cerrado) o difuso (abierto). Como resultado, Echeverría y sus colaboradores afirmaron que la *Ropa* era el centro de interés más compacto, es decir, la mayor coincidencia de palabras ocurría en este centro.

El estudio chileno de disponibilidad léxica de mayor magnitud se debe a Alva Valencia Espinoza y Max Sergio Echeverría. Con un conjunto de trabajos hechos a lo largo de los años noventa, lograron recabar el léxico disponible chileno de ámbito nacional. En esta magna investigación, analizaron una muestra compuesta por 2052 alumnos del último curso de la educación media chilena. En cuanto a los centros de interés, con el objetivo de “abarcar un amplio espectro de los núcleos de conocimiento y actividad de los adolescentes” (Valencia Espinoza, 1997: 203), por un lado, adoptaron 11 centros afines a los clásicos⁵¹ y, por otro, introdujeron 7 nuevos: *La policía*, *La actividad económica*, *Las artes*, *Ciencia y tecnología*, *Mundo espiritual*, *Procesos mentales* y *Problemas del ambiente*. Las variables extralingüísticas consideradas para la estratificación de la muestra eran las habituales en las investigaciones de disponibilidad léxica, a saber, sexo, nivel socioeconómico —alto, medio o bajo—, régimen de estudio —público o privado—, regionalidad —región metropolitana o el resto de las regiones— y procedencia poblacional —urbana o rural—. Las conclusiones sacadas sobre la productividad léxica en relación con dichas variables coincidían con las expectativas⁵²: 1) no se detectaron diferencias significativas entre ambos sexos, aunque, en general, las chicas manifestaban mejores medias que los chicos; 2) los niveles socioeconómicos sí presentaban diferencias marcadas, observándose una ventaja léxica del nivel alto sobre el medio y la de este, a su vez, sobre el bajo; 3) el régimen de estudio, que, prácticamente, venía afectado por el nivel

⁵¹ Estos 11 centros de interés adoptados de los clásicos eran *Las partes del cuerpo*, *Ropa y calzado*, *La casa*, *Muebles*, *Alimentos*, *La cocina*, *Medios de transporte*, *Trabajos de campo y jardín*, *Plantas y animales*, *Juegos y entretenimientos* y *Profesiones y oficios*.

⁵² Con respecto a la influencia que ejercen las variables en la productividad léxica, los resultados cuantitativos aquí referidos eran parecidos a los obtenidos en Puerto Rico, la República Dominicana y Las Palmas de Gran Canaria.

socioeconómico, permitía ver mejor manejo léxico de los alumnos de centros privados que los de centros públicos; 4) con respecto a la regionalidad, los que estudiaban en la región metropolitana presentaban, obviamente, mayores productividad y homogeneidad léxicas que los de las doce regiones restantes en bloque; 5) el análisis del efecto de la procedencia poblacional apuntó a que los estudiantes rurales tenían un vocabulario bastante más reducido que los urbanos. Además de los análisis de correlación mencionados arriba sobre las variables extralingüísticas, el estudio chileno también se ocupó del índice de cohesión que había propuesto Echeverría. A este respecto, se observó que los centros de interés más compactos eran *Las partes del cuerpo, Ropa y calzado, Muebles y Medios de transporte*; en contraposición de *Mundo espiritual, Ciencia y tecnología y Procesos mentales*, que eran los más difusos. Afirmó la propia Valencia Espinoza que esta investigación suya, por lo voluminosa que era y por comprender centros de enseñanza repartidos proporcionalmente por todo el territorio nacional, podía llegar a describir con acierto la fisonomía léxica en los alumnos del último grado de enseñanza secundaria chilena (Valencia Espinoza, 1997: 225).

Respecto de los estudios chilenos de disponibilidad léxica en fechas tempranas, a los ya citados se añade todavía el de Patricia Ester Vargas Sandoval (1988). Esta autora sugirió que la disponibilidad léxica tuviera como objetivo el ser herramienta de planificación curricular para la enseñanza de la lengua. Vargas hizo un trabajo sobre el léxico del hablante medio de la entonces V Región de Chile. Trabajaron con 500 encuestados, de 20 a 49 años de edad, de nivel sociocultural medio. Aplicaron seis centros de interés, novedosos y abstractos⁵³, con los cuales pretendían, en parte, aproximarse al vocabulario abstracto (Bellón Fernández, 2011: 50), pero comprobaron la dificultad de los encuestados para actualizar estas clases de palabras (Ayora Esteban, 2006: 51).

Costa Rica era otro país hispanoamericano donde había exploraciones

⁵³ Los seis centros de interés originalmente ideados en la investigación de Vargas eran *Arte, Derecho, Economía, Política, Sentimientos y Teología*.

sobre la disponibilidad léxica en fechas tempranas. Marielos Murillo Rojas, en los años 1993, 1994 y 1999, puso en práctica la metodología de este tipo de estudio, aplicándola en muestras de informantes de preescolar costarricenses para revisar su desarrollo léxico. En Murillo (1993), la autora se dedicó al estudio sobre 50 niños preescolares en cinco campos semánticos: *Partes del cuerpo, Familia, Animales, Medios de transporte y Profesiones y oficios*. Al tratarse de niños que no sabían leer todavía, la investigadora costarricense utilizó, en la presentación de los centros de interés, láminas que permitían estímulos pictóricos a los pequeños informantes. De igual modo, Murillo (1994) llevó a cabo otro trabajo sobre 50 niños, pero analizó en este caso solamente dos centros de interés: *Comidas y Bebidas*. En estos dos estudios, la investigadora realizó pruebas en dos períodos distintos del curso escolar y comprobó, en ambos estudios, un incremento léxico considerable de los informantes —manifestado por el índice de disponibilidad de las primeras palabras de las listas y por el número de palabras aportadas por los niños (Bellón Fernández, 2011: 58)—.

Más tarde, con idéntica preocupación por el crecimiento léxico, Murillo (1999) llevó a cabo otra investigación más laboriosa de esta índole — estudiando, esta vez, a 289 informantes—. En este artículo, la autora explicó que los 289 alumnos provenían de doce centros de preescolar y primaria y pertenecían a cuatro niveles de estudio: preescolar, primero, segundo y tercer grado de primaria. Se aplicaron diez centros de interés: *Juegos y juguetes, Alimentos, Animales, Profesiones y oficios, Comunidad, Partes del cuerpo, Ropa y accesorios, Medios de transporte, Medios de comunicación y Familia*. Igual que en las dos investigaciones anteriores, la estudiosa empleó láminas motivadoras y preguntas al respecto para mostrar a los alumnos los contenidos de cada centro de interés. Cabe mencionar que Murillo, en este trabajo, optó por las listas cerradas de 20 palabras. Además, la autora tenía en cuenta tres variables sociales de estratificación: ocupación del padre, procedencia geográfica y sexo. Entre las conclusiones más destacadas, se encuentra que el mayor incremento léxico, en todos los centros de interés, se daba de preescolar a primer grado de primaria; de este último a segundo y de segundo

a tercer grado, el avance de la disponibilidad léxica era lento o incluso contrario a lo esperado: por ejemplo, en el centro de interés *Ropa y accesorios*, no había ningún incremento de primero a segundo de primaria y, asimismo, en la transición entre los mismos dos niveles, había una obvia disminución en *Juegos y juguetes* (Murillo Rojas, 1999: 202).

3.2.2.2. Estudios fundamentales en España

3.2.2.2.1. En España el concepto de la disponibilidad léxica se dio a conocer por primera vez en el año 1982 con la tesis doctoral de María José Azurmendi Ayerbe⁵⁴. La autora se inspiró, sobre todo, en las investigaciones pioneras sobre la lengua francesa realizadas por Mackey. Lo que procuró la investigadora española era aprovecharse de la disponibilidad léxica para hacer una descripción sociolingüística del bilingüismo en la Comarca de San Sebastián. Intentando abarcar todos los centros educativos en cuestión, el volumen de la muestra de la tesis resultó bastante grueso: 1224 individuos encuestados.

Mientras revisaba los centros de interés empleados en los estudios iniciales por Mackey (1971) en Canadá y por Njock (1978) en Camerún, Azurmendi decidió aplicar, en las pruebas asociativas de su propia investigación, 15 centros de interés⁵⁵. No obstante, tal y como afirma la investigadora en su tesis, dado que su “finalidad era la elaboración y comprobación de un modelo para la descripción sociolingüística del

⁵⁴ Esta tesis doctoral se titula *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la Comarca de San Sebastián*.

⁵⁵ Los quince centros de interés tenidos en cuenta en las pruebas asociativas por Azurmendi fueron: 1. *Partes del cuerpo*, 2. *Prenda de vestir y de calzar*, 3. *Partes y objetos que hay en una vivienda (casa o piso)*, 4. *Alimentos. Bebidas*, 5. *Relaciones familiares (madre, tío...)* y *no familiares (alumno, amigo...)*, 6. *Ciudad o pueblo: sus partes; lo que se ve y se hace en ellos*, 7. *Los transportes: medios y vehículos que se utilizan*, 8. *Enseñanza: clases de centros y sus partes: objetos y material utilizados, asignaturas, etc.*, 9. *Trabajo: clases de trabajos, de oficios, de profesiones*, 10. *Juegos. Diversiones. Entretenimientos*, 11. *Deportes: prendas de vestir y objetos que se utilizan*, 12. *Acciones que normalmente realizamos todos los días*, 13. *La naturaleza. Las vacaciones*, 14. *Agricultura y vegetales. Ganadería y animales. Pesca y peces* y 15. *Dinero. Bancos. Comercio*.

bilingüismo” (Azurmendi, 1982: 147), solamente analizó cuatro centros de interés, que resultaron ser: *Partes del cuerpo, Prendas de vestir y de calzar, Juegos. Diversiones. Entretenimientos y Agricultura y vegetales. Ganadería y animales. Pesca y peces*. En cuanto a la forma de las listas aplicadas en los cuestionarios, la autora no afirma explícitamente si empleó listas cerradas o abiertas⁵⁶; solo alude a que “en el espacio en blanco... los alumnos respondieron, separando o con una raya o con un blanco, el conjunto de expresiones y palabras producidas por cada Centro de Interés” (Azurmendi, 1982: 237). De modo que se supone que Azurmendi adoptó las listas abiertas⁵⁷, con la restricción del tipo de palabras solicitadas —en los cuestionarios figuraba la petición de sustantivos y verbos como respuesta a cada centro de interés—. Por su parte, el tiempo de reacción que dio la autora a los alumnos era diez minutos por centro.

A pesar de ello, las pruebas asociativas del vocabulario en la investigación de Azurmendi se realizaron de manera muy diferente de las habituales en el resto de estudios de disponibilidad léxica. En sus pruebas, los alumnos no aportaban respuestas en los quince centros de interés estudiados, sino que cada uno contestaba a tres centros, porque el cuestionario venía dado en una sola hoja que contenía solamente tres campos temáticos. Así que había en total cinco tipos distintos de encuesta. Los profesores encargados de la prueba controlaban el procedimiento y hacían que los cinco tipos de encuesta se distribuyeran de manera proporcional entre los informantes, para que los quince centros semánticos pudieran recogerse entre todos los alumnos; de modo que un alumno trabajaba en tres centros de interés; otro, en otros tres; un tercero, en otros tres, y así sucesivamente hasta que la suma de las encuestas cubría los quince centros de interés totales.

⁵⁶ Mackey, investigador muy citado en la tesis de Azurmendi, utilizó las listas cerradas de veinte palabras. En cambio, las listas abiertas fueron planteadas, por primera vez, por Dimitrijević en el año 1969, como ya se ha detallado anteriormente.

⁵⁷ Sin embargo, en la reseña que hizo Pedro Benítez Pérez (1991: 355) sobre esta tesis, se lee que Azurmendi empleó listas cerradas. Lo mismo se lee también en Carmen Ayora Esteban (2006: 53).

De los datos obtenidos a partir del cuestionario, Azurmendi analizó los siguientes índices lingüísticos: 1) la cantidad de producción (productividad lingüística cuantitativa), 2) la clase o categoría léxica, 3) la distancia interlingüística, 4) la corrección lingüística y 5) las divisiones significativas o semánticas. Dedicó buena parte de la tesis a observar las correlaciones que tenían estos índices entre sí de acuerdo con los cuatro centros de interés analizados. Junto al cuestionario lingüístico, la investigadora española diseñó también un cuestionario sociolingüístico para estudiar las siguientes variables: 1) geografía, 2) edad⁵⁸, 3) sexo, 4) clase de escolarización, 5) lengua materna, 6) lengua más utilizada en casa, 7) lengua más utilizada con los amigos y 8) lengua utilizada en las respuestas de la encuesta —español o euskera—. Con estas variables sociolingüísticas, la autora se propuso analizar las correlaciones que tenían estas con cada uno de los índices lingüísticos arriba mencionados. En la última parte de la tesis, tomando como ejemplo los datos del centro de interés *Prendas de vestir y de calzar*, Azurmendi presentó, de manera cualitativa, cuatro diccionarios de disponibilidad: el conceptual general, el del euskera y del español, el de los dos grupos de edad en euskera y, por último, el de los vocablos y de la corrección lingüística.

3.2.2.2.2. La primera investigación española sobre la disponibilidad léxica adherida al *Proyecto Panhispánico* se llevó a cabo en Madrid en el año 1992, con la publicación de *Disponibilidad léxica en la zona metropolitana de Madrid* por Pedro Benítez Pérez (1992a). El autor estudió a 257 alumnos del último curso de la enseñanza media —concretamente, del curso de orientación universitaria (COU) en el sistema educativo español de ese momento—. Adoptó los dieciséis centros de interés ya propuestos por el *Proyecto Panhispánico*, más un centro temático “Colores”. Tuvo en cuenta tres variables extralingüísticas: sexo, tipo de educación —pública o privada— y nivel sociocultural. Posteriormente, el autor dedicó otros artículos abordando temas propios del estudio de disponibilidad léxica o en relación con él: el uso de las

⁵⁸ Con respecto a la edad, Azurmendi la vinculó a las etapas escolares, distinguiendo así entre el quinto (10-11 años de edad) y el octavo cursos (13-14 años de edad) de la enseñanza primaria obligatoria de aquella época (la EGB, o Enseñanza General Básica).

listas cerradas o abiertas empleadas en la prueba asociativa, los niveles socioculturales que influían en el léxico disponible, el anglicismo en las listas de palabras disponibles y el léxico seleccionado en algunos manuales de español como lengua extranjera. Con respecto a este último, Benítez Pérez reveló la inadecuación del criterio de esos manuales al incluir unos vocablos de poco valor de disponibilidad y sugirió que los vocablos de alto índice de disponibilidad aparecieran con mayor frecuencia en este tipo de manuales (Benítez Pérez, 1994).

Una novedad que aportó España a los estudios relacionados con la disponibilidad léxica fue el trabajo que realizó Florentino Paredes García en 1999 sobre la ortografía en las encuestas. Se aprovechó del léxico disponible que recopiló de los alumnos de enseñanza secundaria de Alcalá y su comarca, para alertar del alto porcentaje de aparición de errores ortográficos en el cuestionario. En las conclusiones del artículo, el autor planteó “la necesidad de una labor didáctica de la ortografía” y propuso que “la disponibilidad léxica se convirtiera en una herramienta para estudiar la situación de la población al respecto” (Paredes, 1999). Los resultados generales del léxico disponible en Alcalá de Henares y su comarca pudieron publicarse en 2001, para los cuales el investigador elaboró una muestra de 484 alumnos de enseñanza secundaria⁵⁹ y trabajó con diecisiete centros de interés.

3.2.2.2.3. En la isla de Gran Canaria, los estudios de disponibilidad léxica se iniciaron en fechas paralelas a los de Madrid⁶⁰ y los resultados de las primeras investigaciones aparecieron finalmente en 1994 y 1997. Los trabajos estaban dirigidos por José Antonio Samper Padilla y Clara Eugenia Hernández Cabrera, quienes analizaron una muestra de 539 alumnos de COU para estudiar la variación sociolingüística. Aplicaron diecisiete centros de interés — los dieciséis clásicos y el de los colores— y distinguieron cinco variables

⁵⁹ Los niveles de enseñanza secundaria de estos alumnos eran el tercero y cuarto de Educación Secundaria Obligatoria (ESO) y el primero y segundo de Bachillerato.

⁶⁰ En el año 1992, José Antonio Samper Padilla y Clara Eugenia Hernández Cabrera ya hicieron la primera exposición sobre los datos léxicos de Las Palmas de Gran Canaria en las Jornadas de Lingüística Aplicada celebradas en la Universidad de Granada.

sociales: 1) sexo, 2) nivel sociocultural, 3) tipo de enseñanza —pública o privada—, 4) ubicación de los centros educativos públicos —urbano o rural— y 5) situación de los centros públicos urbanos —centro o periferia—. A partir de estas estratificaciones, los dos autores publicaron, en 1997, *El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística*. Los investigadores encontraron en este estudio una diferencia de disponibilidad léxica entre los informantes de distintos niveles socioculturales aun teniendo idénticos nivel y condiciones de estudio. Como complemento a este trabajo, varios autores⁶¹ emplearon los datos léxicos recogidos por Samper Padilla y Hernández Cabrera y analizaron, en 1994, dos variables más en el vocabulario grancanario: la geografía y el tipo de educación.

Además de recopilar el léxico disponible de Gran Canaria, Samper Padilla también se interesó por la comparación dialectal, o lo que es lo mismo, el cotejo entre las distintas modalidades diatópicas, lo cual ya habían estudiado Juan López Chávez (1992) y Orlando Alba (1998)⁶². Sin embargo, los resultados de los cotejos realizados por estos dos investigadores eran desiguales con respecto a la compatibilidad dialectal: el primero descubrió una compatibilidad léxica baja, mientras que el segundo la observó bastante considerable. Por lo tanto, Samper Padilla pensaba que “si faltaban criterios comunes a la hora de editar los distintos léxicos, el trabajo de comparación —sobre todo en sus aspectos cuantitativos— podía verse seriamente afectado, hasta el punto de que podría llegar a proporcionar unas diferencias dialectales que no se ajustaban completamente a las que en realidad se producían” (Samper Padilla, 1998: 312). De ahí que el estudioso grancanario expusiera, en un artículo del año 1998⁶³, sus consideraciones y criterios de edición sobre las respuestas dadas en las encuestas de disponibilidad léxica. Estos fueron

⁶¹ Estos investigadores eran María Jesús García Domínguez, Vicente Marrero Pulido, José Alberto Pérez Martín y Gracia Piñero Piñero (1994).

⁶² Como se ha comentado en los apartados anteriores, López Chávez (1992) comparó los léxicos disponibles de Madrid, Las Palmas de Gran Canaria, la República Dominicana y Puerto Rico; Alba (1998), por su parte, lo hizo entre los datos léxicos de la República Dominicana, Puerto Rico, Concepción (Chile), Madrid y la Ciudad de México.

⁶³ Se refiere al artículo con el título de “Criterios de edición del léxico disponible: sugerencias”.

una aportación metodológica trascendental que hizo Samper Padilla a nuestra línea de investigación, dado que los criterios presentados en ese artículo han sido pauta para la mayor parte de los siguientes estudios.

Una vez con los criterios de edición establecidos, Samper Padilla abordó el tema del cotejo dialectal. En 1999, se puso a comparar el léxico disponible de Puerto Rico establecido por Humberto López Morales en el mismo año y el suyo de Gran Canaria, dado que ambos habían utilizado criterios de edición uniformes. Los análisis indicaron que la coincidencia léxica entre las dos sintopías era grande, lo cual ratificó la conclusión de Alba (1998) y la suposición de López Chávez (1992) sobre la unidad de la lengua española en los dos lados del Océano Atlántico. Posteriormente, de igual manera, Samper Padilla, junto con Clara Eugenia Hernández Cabrera y Juan José Bellón Fernández, comparó, en 2003, el léxico disponible de Gran Canaria y el de Córdoba. Este estudio de cotejo demostró de nuevo que la homogeneidad de los léxicos de diferentes comunidades era definitiva, lo que daría luz verde a un léxico general del español. Por otro lado, a pesar de la mayor coincidencia que el léxico disponible grancanario tenía con las demás sintopías, los dialectalismos que aparecieron en él aún despertaron el interés de Samper Padilla y Hernández Cabrera. Por eso, los dos investigadores se dedicaron, en 2003, al análisis del canarismo en las respuestas aportadas por los informantes en el centro de interés *Alimentos y bebidas*.

3.2.2.2.4. Andalucía es la comunidad española donde los estudios de la disponibilidad léxica han avanzado de forma contundente. En 1992, apareció la primera exploración —finalmente publicada en 1994— realizada por Francisco Joaquín García Marcos, María Victoria Mateo García y Antonio Daniel Fuentes González. En este primer estudio, recurriendo a la disponibilidad léxica, los autores analizaron las palabras relacionadas con el ámbito cañero en la costa granadina —en Motril, concretamente—, porque esas palabras sufrían una posible mortandad lingüística. Sobre la base de esto, Mateo García llevó a cabo su memoria de licenciatura, tratando de la estratificación social del léxico en Motril en el centro de interés de la

alimentación. Más tarde, en 1995, la misma investigadora comparó los primeros veinte vocablos de alimentación de Motril con los del mismo centro de interés obtenidos en Madrid por Benítez Pérez (1992a) y en México por Ruiz Basto (1987)⁶⁴. Los resultados de este cotejo revelaban que había, por lo menos en la temática de la alimentación, un léxico común entre las tres regiones.

En 1995, el *Proyecto Panhispánico* empezó a adquirir importancia en Andalucía. En ese año, Humberto López Morales y Francisco Joaquín García Marcos establecieron un proyecto en esta comunidad española que pretendía recolectar el léxico disponible andaluz en general y por provincias. Las investigaciones del proyecto andaluz mantenían las pautas metodológicas comunes. Adoptaron los dieciséis centros de interés tradicionales más el de los colores y el de la mar. Pasaron aproximadamente 400 encuestas entre los alumnos andaluces de COU. Utilizaron listas abiertas con un límite de reacción de dos minutos para cada centro de interés. Con respecto a las variables extralingüísticas, tuvieron en cuenta el sexo, el entorno sociocultural familiar, el tipo de colegio y el nivel sociocultural. Se calculó este último sumando la puntuación asignada al nivel de instrucción de los padres y a la ocupación laboral de los mismos. Todos los datos léxico se sometieron a la fórmula informática diseñada por Chávez López y Strassburger. La finalidad primordial de estas investigaciones andaluzas era elaborar un diccionario general de léxico disponible andaluz y otros provinciales, a los cuales se añadían trabajos monográficos basados en ellos (García Marcos & López Morales, 1995).

A partir de ahí, surgieron en Andalucía estudios monográficos de la disponibilidad léxica que cubrirían todas las provincias de la comunidad: el de María Victoria Mateo García (1998) en Almería, el de Adolfo González Martínez (2002) en Cádiz, el de Josefina Prado Aragonés y María Victoria Galoso Camacho (2005) en Huelva, el de Ignacio Ahumada Lara (2006) en Jaén, el de Antonio Manuel Ávila Muñoz (2006) en Málaga, el de María Ángeles Pastor

⁶⁴ Se refiere a la memoria de licenciatura por Araceli Ruiz Basto en 1987 titulada "Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades Plantel Naucalpan".

Milán y Francisco José Sánchez García (2008) en Granada, el de Juan José Bellón Fernández (2011) en Córdoba y, por último, el de Ester Trigo Ibáñez (2011) en Sevilla. Asimismo, algunos investigadores de la Universidad de Cádiz y la Universidad de Granada también confeccionaron el léxico disponible de Ceuta y Melilla —en concreto, el de Ceuta, María Carmen Ayora Esteban (2006), y el de Melilla, Gérard Fernández Smith, Ana María Rico Martín, María José Molina García y María Ángeles Jiménez Jiménez (2008)—. A partir de mediados de los años 90, parece que los estudios de disponibilidad léxica empezaron a hallarse en un nuevo período, en el cual la homogeneidad en la metodología ha permitido un rápido arranque y desarrollo de distintas investigaciones y una tentativa de aplicar la disponibilidad léxica en otros ámbitos de estudio. Así es que, además de haber recopilado los diccionarios de léxico disponible, en la comunidad de Andalucía, por ejemplo, los temas del dialectalismo, el anglicismo, la influencia de las variables, el cotejo entre distintas sintopías y la enseñanza de lengua materna y extranjera han sido también muy estudiados.

3.2.2.2.5. Del léxico disponible de Castilla y León y de Castilla-La Mancha se han ocupado varios investigadores de la Universidad de Salamanca. Un equipo de esta universidad, dirigido por el profesor Julio Borrego Nieto, ha realizado y mantenido la página web del *Proyecto Panhispánico* y del programa informático *Dispolex*⁶⁵; los responsables de la página son José Antonio Bartol Hernández y Natividad Hernández Muñoz.

La primera investigación de disponibilidad léxica en estas comunidades corresponde a María Victoria Galloso Camacho, quien recopiló, en un principio, el léxico disponible de los estudiantes preuniversitarios de la provincia de Zamora. Siguió las pautas metodológicas comunes a otras investigaciones afines, analizando una muestra compuesta por 100 alumnos de nivel preuniversitario. Concretamente, contempló los dieciséis centros de interés tradicionales, estableció dos minutos de reacción para cada centro, adoptó las

⁶⁵ La dirección de dicha página web es <http://www.dispolex.com>.

listas abiertas y tuvo en cuenta variables sociales tales como sexo, tipo de centro, ubicación del centro, residencia de los padres, estudios de los padres y profesión de los mismos (Galoso, 1997a y 1997b). A base de las palabras obtenidas en Zamora, la investigadora hizo un interesante análisis semántico, estudiando las relaciones asociativas entre las respuestas aportadas por los informantes en un centro de interés. Afirmó la detección de grupos compactos de palabras que siempre aparecían juntas. Era un tipo de estudio similar al que realizó Naum Dimitrijević en 1969, cuando observó parejas de palabras que frecuentemente se presentaban juntas —*perro* y *gato*, por ejemplo— y casos en los que los informantes empezaban una lista enumerando palabras relacionadas con alguna área específica y no pasaban a otra área sin terminar el recorrido completo de la anterior.

En continuación de la línea de investigación, Galoso retomó los datos que tenía de Zamora y publicó, en 2003, *El léxico disponible de Ávila, Salamanca y Zamora*. Hasta fechas recientes, han aparecido, con respecto al resto de Castilla y León, estudios del léxico disponible en Soria (Bartol Hernández, 2004), Burgos (Fernández Juncal, 2008) y Segovia (Santiago Guervós, 2008). Por otro lado, Natividad Hernández Muñoz (2006) recabó el léxico disponible de los estudiantes de Castilla-La Mancha y trabajó en la producción léxica desde la perspectiva psicolingüística. Al igual que en Andalucía, en la nueva etapa del estudio de la disponibilidad léxica, con respecto a los trabajos llevados a cabo dentro de la Universidad de Salamanca, han sido muy tratados los temas sobre el cotejo dialectal, la sociolingüística, la psicolingüística y la enseñanza del español a inmigrantes y a extranjeros.

3.2.2.2.5. En las demás comunidades españolas, los estudios de disponibilidad léxica se iniciaron, más o menos, de la misma manera como se ha comentado arriba, es decir, siguieron las primeras investigaciones fundamentales en el mundo hispánico —concretamente, las de Puerto Rico, de México, de Madrid o de Gran Canaria—. El más temprano fue el estudio de Maitena Etxebarria Arostegui, en 1996, en el área metropolitana de Bilbao, en el País Vasco. Trabajó con 225 estudiantes —120 chicos y 125 chicas— de

COU en situaciones de contacto de lenguas y, a partir de los modelos de enseñanza bilingüe, estudió la disponibilidad léxica en la lengua española de los estudiantes. Posteriormente, con los resultados que se habían sacado de ese estudio y analizando los niveles sociolingüísticos en los informantes, la investigadora también estudió, en 1999, la caracterización sociolingüística del léxico disponible del País Vasco.

De la disponibilidad léxica del Principado de Asturias se ocupó el estudioso Alberto Carcedo González en el año 2001. Consideró las viables sociales habitualmente adoptadas: sexo, centro escolar —público o privado—, tipo de comunidad —urbano o rural—, nivel sociocultural y zona geográfica; pues analizó la incidencia de estas en el léxico disponible. Más tarde, estudió, de acuerdo con cada una de estas variables, la compatibilidad entre los léxicos estratificados. Asimismo, comparó el léxico disponible de Asturias con los de Cádiz, Zamora, Chile, Puerto Rico y la República Dominicana.

En la Comunidad Valenciana, los estudios de disponibilidad léxica se desarrollaron en sus tres provincias, con la labor de equipos de investigación diferentes. José Ramón Gómez Molina y María Begoña Gómez Devís se ocuparon del léxico disponible de la provincia de Valencia, como resultado se publicó, en 2004, la tesis doctoral de Gómez Devís con el título de *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Con respecto a la provincia de Castellón, fueron José Luis Blas Arroyo y Manuela Casanova Ávalos quienes se dedicaron al proyecto de la disponibilidad léxica. Dieron a luz publicaciones sobre el anglicismo (Casanova, 2008) y el valencianismo (Casanova, 2013) en el léxico disponible de Castellón y, finalmente, se dio a conocer, en 2017, la tesis doctoral de esta autora titulada *Léxico disponible de Castellón. Estudio y diccionarios*. En la provincia de Alicante, trabajaron en esta línea de investigación Francisco Gimeno Menéndez y Esther Martínez Olmos. Se publicó, en 2008, la tesis doctoral de esta investigadora, titulada *Disponibilidad léxica en las comunidades de habla alicantinas*.

Los estudios de disponibilidad léxica en Cataluña se iniciaron en la provincia de Lleida, donde se hicieron cargo de ellos M^a Ángeles Calero Fernández y Maribel Serrano Zapata. Las investigaciones realizadas en Lleida tenían un enfoque sociolingüístico y estudiaban las dos lenguas en contacto — español y catalán—. Se recabaron los léxicos disponibles en las dos lenguas cooficiales con datos recogidos en el año 2002 y se publicaron la tesis doctoral de Serrano Zapata (2014), que lleva por título *Disponibilidad léxica en la provincia de Lleida: estudio comparado de dos lenguas en contacto*, y otros trabajos sobre aspectos sociolingüísticos del léxico disponible en Lleida (Serrano Zapata, 2004), los fenómenos de contacto de lenguas (Serrano Zapata, 2006), las transferencias léxicas en comparación con los resultados del corpus PRESEEA-Lérida⁶⁶ (Calero Fernández & Serrano Zapata, 2017) y la incidencia del factor sexo en este tipo de vocabulario (Calero Fernández & Serrano Zapata, 2019). Del léxico disponible del resto de Cataluña y de las islas Baleares, se ha encargado un equipo de investigadores de la Universidad de Barcelona —formado, concretamente, por Esther Blasco Mateo, Antonio Torres Torres, María Ángeles Pérez Edo y Mar Forment Hernández—. Los trabajos de estos investigadores están todavía en proceso.

Galicia es otra comunidad española donde se llevaron a cabo investigaciones de la disponibilidad léxica en contextos de bilingüismo. Belén López Meirama (2008) realizó trabajos sobre el léxico disponible en el español de Galicia. A partir de estos datos, la autora también estudió temas como la ortografía, la interferencia y el galleguismo.

En Aragón, la recopilación del léxico disponible fue coordinada, en 2004, por María Luisa Arnal Purroy y, posteriormente, esta investigadora ha

⁶⁶ El PRESEEA (*Proyecto de Estudio Sociolingüístico del Español de España y América*) es un proyecto impulsado por Humberto López Morales con el objeto de crear un macrocorpus de lengua española hablada que represente la variedad dialectal y sociolectal del territorio hispanohablante sobre el que desarrollar investigaciones que sean comparables entre sí. Este proyecto promueve una metodología común para todos los equipos que participan en él. La Universidad de Lleida está integrada en este proyecto desde sus inicios y cuenta con el corpus PRESEEA-Lérida, dirigido por M^a Ángeles Calero. Véase <https://preseea.linguas.net/>.

estudiado, sobre todo, el dialectismo en el vocabulario disponible de dicha comunidad (Arnal, 2008b y 2008c), pero también este tipo de léxico entre los jóvenes aragoneses (2008a). Con respecto a Navarra, Marta Mangado Cruz y María Areta Lara (2008) estudiaron, en un principio, el procesamiento informático para la elaboración de diccionarios de disponibilidad léxica. Un año más tarde, Areta (2009) dio a luz su tesis doctoral con el título de *El léxico disponible de los estudiantes preuniversitarios navarros. Estudio sociolingüístico*. En cuanto a Cantabria, fue María del Carmen Fernández Juncal, de la Universidad de Salamanca, quien acabó, en 2013, el léxico disponible de la comunidad con un enfoque sociolingüístico.

3.3. Aplicaciones de la disponibilidad léxica

Los corpus de léxico disponible, junto con las diferentes variables asignadas a los informantes que componen las muestras, sirven de recursos de análisis y, así, facilitan una amplia gama de estudios en contacto con varias ramas de la lingüística. Como se puede comprobar en los trabajos científicos comentados en los apartados anteriores, la disponibilidad léxica ha sido una herramienta que, más allá de dar con el vocabulario de uso habitual, nutre la exploración de ámbitos de investigación como la dialectología, la sociolingüística, la psicolingüística, la etnolingüística, la didáctica de lenguas, etc. Humberto López Morales (1995) fue el primero en llamar la atención sobre los contactos de la disponibilidad léxica con dichas disciplinas. A continuación, vamos a observar las aplicaciones de la disponibilidad léxica en dichas áreas de estudio y destacamos el contacto de la disponibilidad léxica con la enseñanza de la lengua, que es el propósito de la presente tesis.

3.3.1. Disponibilidad léxica y dialectología. Dado que los estudios de disponibilidad léxica tienen como objetivo elaborar diccionarios de léxico disponible de una comunidad de hablantes determinada, son aprovechables para hacer comparaciones diatópicas entre los datos léxicos de distintas regiones geográficas. Naturalmente, las comparaciones de este tipo permiten detectar los vocablos comunes de las sintopías, los vocablos diferentes entre

las mismas y los vocablos exclusivos de alguna de ellas. Por consiguiente, la detección de los dialectalismos hace posible ver la homogeneidad o variedad léxicas de la lengua estudiada. La aplicación de la disponibilidad léxica en la dialectología ya fue una preocupación de las primeras investigaciones.

William Francis Mackey et al. (1971), al recopilar el léxico disponible en francés de Acadie (Canadá) decidieron mantener los dialectismos canadienses con la finalidad de estudiarlos y compararlos con los resultados anteriormente sacados del francés de Francia por Georges Gougenheim et al. (1964). Naum Dimitrijević (1981), en su trabajo sobre el léxico disponible de los alumnos monolingües y bilingües de Novi Sad de la antigua Yugoslavia, observó que los informantes bilingües habían escrito palabras dialectales en su segunda lengua —el serbocroata—, las cuales no aparecían en las listas de los hablantes monolingües cuya lengua materna era el mismo serbocroata.

En el mundo hispánico, Juan López Chávez (1992) fue el primero en meterse en el ámbito de la dialectología, puesto que estableció comparaciones entre los léxicos disponibles de cuatro zonas geográficas diferentes: el de Madrid (recogido por Pedro Benítez Pérez), el de Las Palmas de Gran Canaria (recogido por José Antonio Samper Padilla), el de la República Dominicana (recogido por Orlando Alba) y el de Puerto Rico (recogido por Humberto López Morales). Aun partiendo de la suposición de que existía un vocabulario español comúnmente compartido entre las diferentes zonas dialectales, López Chávez descubrió, como ya hemos comentado, una baja compatibilidad léxica. Orlando Alba (1998) retomó el tema sobre el cotejo dialectal y comparó los vocablos de alto grado de disponibilidad de tres centros de interés —*El cuerpo humano*, *Medios de transporte* y *Alimentos*— entre cinco comunidades de habla distintas: además de Madrid, República Dominicana y Puerto Rico, comparadas ya por López Chávez, añadió Concepción (Chile) (estudiada por Mena Osorio), y México⁶⁷ (analizada por López Chávez). A su vez, Alba puso

⁶⁷ Los vocablos de México sometidos al estudio comparativo de Orlando Alba corresponden al léxico del nivel escolar “sexto grado” de las investigaciones que coordinó Juan López Chávez, en 1993, en la ciudad de México.

de relieve un alto grado de compatibilidad léxica entre los dialectos de regiones geográficamente distantes entre sí, lo cual sería corroborado, un año más tarde, en Samper Padilla (1999). De las aplicaciones a la dialectología hispanoamericana, hubo otra aportación por Alba Valencia (2005), quien estudió los dialectalismos en el léxico disponible chileno.

En España, José Antonio Samper Padilla, Clara Eugenia Hernández Cabrera y Marta Samper Hernández son tres estudiosos que han abordado, en varios trabajos científicos, los dialectalismos en el léxico disponible de Gran Canaria. Además, como se ha mencionado más arriba con respecto al cotejo dialectal, Samper Padilla (1999) realizó una investigación de comparación entre el léxico disponible de Puerto Rico y el de Gran Canaria, en la cual pudo comprobar el alto grado de compatibilidad léxica propuesto por Orlando Alba. Más tarde, en 2003, Samper Padilla y Hernández Cabrera, junto con Juan José Bellón Fernández, llevaron a cabo otro estudio que comparaba el léxico de Gran Canaria y el de Córdoba. Adolfo González Martínez y Pepa Orellana Ramírez (2000), asimilando los primeros trabajos comparativos de los investigadores canarios, cotejaron los léxicos disponibles de Cádiz (recogido por González Martínez) y Zamora (recogido por María Victoria Galoso Camacho). Julio Borrego Nieto y María del Carmen Fernández Juncal (2002) trataron de las cuestiones metodológicas en las investigaciones de aplicación del léxico disponible a los estudios dialectales. Alberto Carcedo González, en 2002, presentó una comunicación⁶⁸ que comparaba el léxico disponible de Asturias con los de Zamora, Cádiz, Puerto Rico, República Dominicana y Chile. María Luisa Arnal Purroy (2008b y 2008c), por su parte, dedicó dos trabajos al léxico dialectal de Aragón. Además de estos trabajos, existen todavía otros que estudian la dialectología a partir de la disponibilidad léxica, por ejemplo, sobre el léxico dialectal de Ávila (Llorente Pinto, 2005), el de Castilla-La Mancha (Hernández Muñoz, 2009), el de Sevilla (Trigo Ibáñez, 2011), etc.

⁶⁸ Se trata de “La variable léxica disponible en la comparación interdialectal: compatibilidad de la norma asturiana con otras sintopías hispanohablantes”, presentada en el *XIII Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina (ALFAL)*, celebrado en San José de Costa Rica y cuyas actas se publicaron en 2004.

3.3.2. Disponibilidad léxica y sociolingüística. La sociolingüística es otra rama de ciencia que se ha aprovechado de los resultados de las investigaciones de disponibilidad léxica. Esto se debe a que las iniciativas metodológicas de los estudios sobre el léxico disponible plantearon recoger datos extralingüísticos —sociológicos, en su mayoría— de los informantes. Entre estos, los que se conocen como variables sociales han dado luz verde a estudios posteriores de estratificación social con respecto a los componentes en la muestra. Las variables sociales más consideradas en las investigaciones aplicadas a la sociolingüística son el sexo (varón o mujer), el nivel sociocultural (alto, medio o bajo⁶⁹), el tipo de centro (público o privado) y la zona geográfica (urbana o rural). Para las investigaciones que se realizan en comunidades bilingües o de lenguas en contacto, también es común que, aparte de las variables citadas arriba, se tienen en cuenta la lengua materna, la lengua de uso habitual y el modelo de enseñanza (monolingüe o bilingüe). Así establecidas las cosas, los investigadores de disponibilidad léxica pueden emprender trabajos sobre la incidencia de una variable en el léxico disponible de los informantes, analizando, cuantitativa o cualitativamente, las cifras o las respuestas de distintos grupos sociales.

El estudio de las variables extralingüísticas o el estudio sociolingüístico de la disponibilidad léxica data de las primeras investigaciones de esta línea. Gougenheim et al. (1964) ya analizaron la relevancia que tenían los factores sociales sexo y procedencia geográfica en las respuestas aportadas por los informantes. Dimitrijević (1969) revisó cuantitativamente las variables sociales sexo y tipo de centro escolar. Llegó a la conclusión de que la primera era un factor menos relevante que la segunda y, en la segunda, pudo observar manifestaciones estadísticas —mediante la producción léxica— de que un *senior secondary school* era superior a un *junior secondary school*. Esto se puede traducir en la correlación del nivel de estudio con la disponibilidad léxica.

⁶⁹ En algunas investigaciones también se consideran los niveles medio-alto o medio-bajo según lo vea conveniente el investigador en su diseño de estratificación. Los diferentes niveles suelen ser evaluados en función de la suma de las puntuaciones que se asignan a cada uno de los indicadores tenidos en cuenta.

En otra investigación comparativa entre estudiantes monolingües y bilingües, Dimitrijević (1981) analizó cuantitativamente las variables lengua materna, edad, sexo e inteligencia y, en todas, encontró correlaciones con la disponibilidad léxica. De manera similar, John Bailey Victory (1971) calificó la edad de factor sumamente influyente en el léxico disponible y estudió las correlaciones que el sexo y el nivel socioeconómico mantienen con él.

Las investigaciones posteriores en el mundo hispánico y, sobre todo, las enmarcadas en el *Proyecto Panhispánico* han incluido factores extralingüísticos más o menos comunes en la descripción tanto del panorama léxico de las comunidades de habla como de las características léxicas propias de los sociolectos. De esta manera, se estudian las variaciones léxicas en función de las variables. Entre los condicionantes habitualmente analizados, cabe destacar el nivel sociocultural, que, en comparación con otros, ha sido un factor social de mucho peso en la investigación sociolingüística. López Morales (1979) ya advirtió de la importancia que tenía el nivel sociocultural en relación con las diferencias léxicas. De ahí en adelante, al ser ampliamente estudiada esta variable social, se ha revelado la tendencia, por lo general, de que los estratos de un nivel sociocultural más alto también cuentan con una producción léxica mayor. Aquí citamos algunos de los estudios sociolingüísticos de la disponibilidad léxica como el de Gran Canaria (García Domínguez et al., 1994; Samper Padilla & Hernández Cabrera, 1997), Almería (Mateo García, 1998), Aragón (Arnal Purroy, 2008), Cantabria (Fernández Juncal, 2013) y Chile (Valencia Espinoza & Echeverría, 1999).

Un enfoque de los estudios sociolingüísticos de la disponibilidad léxica en el mundo hispánico se ha dado en las regiones bilingües o de lenguas en contacto. De esta índole se destacan los trabajos de Lleida (Serrano Zapata, 2003, 2014; Calero Fernández & Serrano Zapata, 2017), Valencia (Gómez Devís, 2004), Castellón (Casanova Ávalos, 2006, 2017), Alicante (Martínez Olmos, 2008), País Vasco (Azurmendi, 1982; Etxebarria Arostegui, 1996, 2012), Navarra (Areta, 2009), Galicia (López Meirama, 2008, 2011), Ceuta (Ayora Esteban, 2006), Melilla (Fernández Smith et al., 2008), etc.

3.3.3. Disponibilidad léxica y psicolingüística. El lexicón mental es uno de los objetos de estudio de la psicolingüística y es el elemento que conecta esta con las investigaciones de la disponibilidad léxica. Se trata de una especie de diccionario mental que permite a los individuos hacer actualizaciones léxicas en situaciones comunicativas orales o escritas. Las pruebas asociativas del vocabulario, que consisten en incentivar las asociaciones léxicas de los informantes en torno a un estímulo temático —conocido también como los centros de interés—, aplicadas en las investigaciones de disponibilidad léxica favorece el acercamiento de los científicos a una representación de la estructura del lexicón mental.

A este respecto, el *índice de cohesión*⁷⁰, creado —como ya hemos dicho— por Echeverría (1987), ha resultado sumamente importante, porque mide el grado de coincidencia entre las actualizaciones léxicas de los informantes para un centro de interés. De esta manera, se puede saber lo compacto (cerrado) o difuso (abierto) que es un centro de interés. A los psicolingüistas, les interesan los centros más compactos, puesto que estos — en los que los vocablos y las posiciones de los mismos presentan unas relaciones más estrechas— son, prácticamente, las áreas semánticas que más se acercan y mejor representan la estructura del lexicón mental correspondiente a dichas áreas. La *densidad léxica*⁷¹, propuesta por Orlando Alba (1995b), es otra forma de cálculo sobre el grado de conformidad en las respuestas dentro de un centro de interés. Aunque el índice de cohesión y la densidad léxica funcionan de manera igual al describir el grado de coincidencia, el primero ha sido más empleado, porque el valor máximo del índice es 1 y esto presenta una mejor visualización —en comparación con la densidad— de los resultados cuantitativos. Son muchos los estudios de la disponibilidad

⁷⁰ El índice de cohesión resulta de dividir el promedio de respuestas por informante en cada centro de interés por el número de palabras diferentes del mismo centro. El resultado del cálculo será 1 en el caso de que todos los informantes escribieran las mismas palabras en el mismo centro de interés.

⁷¹ La densidad léxica resulta de dividir el total de palabras de un centro de interés por el número de palabras diferentes del mismo.

léxica que han investigado cuál es el centro de interés más compacto. Citamos aquí, por ejemplo, el de Echeverría et al. (1987), que señaló que *La ropa* era el centro de interés más compacto. En otro trabajo de Echeverría y Alba Valencia Espinoza (1999a), se comprobó que era *Las partes del cuerpo*. Por su parte, Alba (1995) confirmó que el centro más cerrado era *El cuerpo humano*.

Por otro lado, las respuestas léxicas producidas en las pruebas de disponibilidad léxica pueden servir de fuente de estudio para conocer cómo se reproducen y organizan los vocablos del lexicón mental. Para esto, los investigadores han recurrido a la *teoría de la red*, propuesta por Aitchison en 1987, con la finalidad de analizar las tres relaciones típicas entre el estímulo y la reacción: 1) los elementos del mismo campo léxico de la palabra original, 2) las parejas si la palabra estímulo es un par o tiene opuestos evidentes, 3) las palabras de la misma clase gramatical (apud Gómez Devís, 2004; López Morales, 1995).

Dentro del centro de interés, los investigadores también estudian las relaciones entre las mismas respuestas léxicas producidas. En 2008, un equipo de investigadores coordinado por Max Sergio Echeverría diseñó el programa computacional “Dispógrafo”, que presenta en forma de grafo las redes semánticas de las palabras más disponibles. Ha sido un gran apoyo al análisis psicolingüístico del léxico disponible. En este tema han investigado autores como Juan López Chávez (1994), María Victoria Galloso Camacho (1997b), María Begoña Gómez Devís (2004), etc. Estas últimas dos investigadoras han descrito —cada una por su parte— las relaciones asociativas entre las palabras producidas en un listado de términos. Según Gómez Devís (2004: 49-50), las formas de asociación que se establecen dentro de una red semántica, aunque son variadas según diferentes autores, pueden resumirse en las cuatro siguientes: 1) asociaciones por el significante —*lanza* y *balanza*, *camarero* y *camionero*—, 2) asociaciones por los significados —*escuela* y *aula*, *enseñanza* y *educación*—, 3) asociaciones semánticas —*burro* y *asno*, *frío* y *calor*—, 4) asociaciones sintagmáticas —*carta* y *certificada*—.

Cabe mencionar que la tesis doctoral de Natividad Hernández Muñoz (2006), titulada *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*, es el intento psicolingüístico más serio y novedoso hasta la fecha en el marco de la disponibilidad léxica. La autora abordó, mediante experimentos psicolingüísticos, el fundamento cognitivo en la producción del léxico disponible y el componente semántico que generaba el listado de palabras.

3.3.4. Disponibilidad léxica y enseñanza de la lengua. Una de las principales utilidades de la disponibilidad léxica es su aplicación en la enseñanza de la lengua. Como es bien sabido, el léxico tiene un papel de relevancia en el proceso del aprendizaje de la lengua. No obstante, en los diccionarios normativos abundan terminologías, tecnicismos y cultismos poco habituales o regionalismos propios de una diatopía y desconocidos por las demás, hecho que niega su uso como fuente de aprendizaje del léxico, porque, con ellos, los aprendices no pueden tener acceso al léxico real de uso corriente. Por tanto, la planificación de la enseñanza del léxico debe buscar una vía al léxico fundamental de las comunidades de habla, el cual, como se ha demostrado en el capítulo anterior, está constituido por el léxico básico y el disponible. Los estudios de disponibilidad léxica sirven para la enseñanza de la lengua en este sentido, ya que presentan a los estudiantes el caudal del léxico disponible en su curso de adquisición de la competencia referida al vocabulario. Ofrecen unos resultados muy provechosos para la elaboración de materiales didácticos y diccionarios temáticos. Por otro lado, y al revés, la metodología de la disponibilidad léxica —en concreto, las pruebas asociativas del léxico— también brindan una forma de evaluación sobre la competencia léxica de los aprendices en sus diferentes fases de adquisición del vocabulario.

Evidentemente, se ha visto que la disponibilidad léxica se ha aplicado a la didáctica de la lengua materna, la segunda lengua y la lengua extranjera. Con respecto al primer caso, López Morales (1978) ya manifestó que la frecuencia y disponibilidad léxicas podían servir para elaborar la programación

curricular del vocabulario. López Chávez y Rodríguez Fonseca (1992), refiriéndose a la utilidad de la disponibilidad léxica y la teoría del lexicón mental, plantearon un modelo para la programación de la enseñanza del vocabulario a partir de unos vocablos rectores y, por consiguiente, otros secundarios. Un ejemplo reciente de esta índole se da en la tesis doctoral de María José García Casero (2013), la cual indagó el léxico disponible de los estudiantes de cuarto curso de Educación Secundaria Obligatoria en Santander. En uno de los capítulos de la tesis, la autora aprovechó los resultados de la muestra cántabra para elaborar una propuesta que pondría en práctica la enseñanza del vocabulario en clase y que ampliaría la riqueza léxica de los alumnos.

Además, también existen trabajos que han realizado análisis sobre las distintas fases de adquisición del léxico en lengua materna. Paredes (2001) recabó, en Alcalá, el léxico disponible en informantes de tercero y cuarto de la educación secundaria obligatoria (ESO) y de primero y segundo del bachillerato para analizar, entre otras, la variable nivel de estudio. Antonio García Megía (2003) trabajó con estudiantes entre los nueve y los doce años de edad en la ciudad de Almería. Marta Samper Hernández (2003b) estudió la evolución de la disponibilidad léxica en estudiantes grancanarios de enseñanza primaria y secundaria, con informantes de tercero y quinto de primaria y primero y tercero de la ESO. De manera similar, Bárbara Ortolano Ríos (2005) analizó la disponibilidad de los estudiantes de segundo y cuarto de la ESO y de segundo del bachillerato en Ayamonte (Huelva).

En relación con la ciencia informática, también ha tenido éxito la aplicación didáctica de la disponibilidad léxica. El investigador chileno Max Sergio Echeverría diseñó, en 2001, un programa informático denominado “Vocabulario Disponible”⁷² que podía hacer un diagnóstico sobre el vocabulario disponible de los estudiantes de educación media en dieciocho centros de interés⁷³. Basado en los resultados de una serie de investigaciones

⁷² En la página web <http://www.c5.cl/ieinvestiga/actas/tise99/html/papers/metacognicion/> se encuentra la descripción detallada del diseño y funciones del programa.

⁷³ Estos son *Cuerpo humano, Muebles, Medios de transporte, Juegos y entretenimientos, Economía, Mundo espiritual, Ropa y calzado, Alimentos, Campo y jardín, Profesiones y*

llevadas a cabo a lo largo de los años 1994, 1997, 1998 y 1999 —mediante las cuales se logró recabar el léxico disponible chileno de ámbito nacional—, dicho programa permite comparar la autoevaluación del alumno con el nivel medio nacional. Se presentan, al final del diagnóstico léxico, los vocablos de alto grado de disponibilidad que el alumno ha olvidado y se diseñan actividades de aprendizaje de nuevo vocabulario sobre la base de tareas interactivas (texto, imagen, audio y video).

Otra área en la que la aplicación didáctica de la disponibilidad léxica ha tenido repercusión es la enseñanza de segundas lenguas. Existen bastantes estudios de esta índole en las zonas de lenguas en contacto. Citamos aquí como ejemplo el de María del Carmen Ayora Esteban (2004), que investiga las deficiencias en la adquisición del léxico como problema específico de las lenguas en contacto; el de Maribel Serrano Zapata (2006), que estudia las consecuencias del contacto de lenguas sobre la base de las interferencias detectadas en las encuestas de disponibilidad léxica en Lleida; y el de María Begoña Gómez Devís (2010), que trata los aportes de la disponibilidad léxica a la investigación sobre fenómenos de contacto lingüístico.

Aparte de ello, la puesta en práctica de la disponibilidad léxica en la enseñanza de la lengua extranjera es un tema de estudio que ha adquirido mucha fuerza en épocas recientes, aunque la utilidad de esta línea de investigación como herramienta de enseñanza y aprendizaje para los extranjeros ya se concibió, como hemos visto, en las primeras investigaciones del tema. Georges Gougenheim y sus colegas (1964) averiguaron cuál era el léxico fundamental de la lengua francesa, para lo que aplicaron por primera vez pruebas de disponibilidad léxica. La iniciativa de esas investigaciones era facilitar el aprendizaje del francés a los inmigrantes llegados a Francia y, asimismo, a habitantes de países africanos que habían formado la Unión Francesa, en donde el francés era lengua oficial o la lengua extranjera más influyente. De igual manera, la aplicación pedagógica de la disponibilidad

oficios, Las artes, Procesos mentales, La casa, Cocina, Plantas y animales, Política, Ciencia y tecnología y Problemas del ambiente.

léxica sobre el español como lengua extranjera ha obtenido grandes desarrollos. Recordamos que las primeras investigaciones de este tipo las llevó a cabo Carcedo (1998a, 1998b, 1999, 2000), quien analizó el léxico disponible de los dieciséis centros tradicionales en 350 estudiantes fineses de bachillerato y de universidad, dando prioridad a los 150 informantes del último curso de bachillerato. Como resultado, el investigador realizó trabajos relativos al análisis de errores, las incidencias de las variables y la evolución del léxico.

Matías Hidalgo Gallardo (2017d) y Lahoussine Aabidi (2019) repasaron de manera pormenorizada los estudios de disponibilidad léxica en español como lengua extranjera (ELE) en sendos trabajos. Aquí destacamos que ambos autores coinciden en dividir los estudios de disponibilidad léxica en ELE en los siguientes dos tipos, en función del perfil de los informantes: estudios que tienen como objeto de investigación 1) los estudiantes de ELE en su país de origen —en países extranjeros donde la lengua materna de la gente no es el español— o 2) los estudiantes de ELE en situación de inmersión lingüística —en España o en Hispanoamérica—.

El primer tipo es el que preocupa a la presente tesis, ya que nuestra investigación es una de este tipo. Según Hidalgo y Aabidi, ha habido investigaciones de disponibilidad léxica en ELE en países extranjeros como Finlandia, Alemania, Polonia, Eslovenia, Islandia, Rumania, Portugal, Italia, Estados Unidos, Brasil, China, Turquía, Corea del Sur, Egipto y Marruecos⁷⁴. Con respecto a China, llamamos la atención sobre dos estudios importantes llevados a cabo, respectivamente, por Lin (2012) y por Hidalgo (2019b). En cuanto a la metodología, Lin (2012) contempló los dieciséis centros de interés clásicos, pasó encuestas a alumnos universitarios procedentes de cinco ciudades del país y tuvo en cuenta las variables del sexo, la edad, el nivel de español y la zona geográfica de los estudiantes. En cambio, Hidalgo (2019b) estudió dieciocho centros de interés⁷⁵, pasó encuestas en seis universidades

⁷⁴ Recomendamos la consulta de los dos artículos referidos para conocer las investigaciones de disponibilidad léxica en ELE en los citados países extranjeros.

⁷⁵ Los centros de interés contemplados por Hidalgo Gallardo son 01. *El cuerpo humano*, 02. *La ropa y complementos*, 03. *La casa*, 04. *Comidas y bebidas*, 05. *Escuela y*

situadas en sendas ciudades diferentes del país y consideró las variables del curso académico, el nivel de español, el prestigio de la universidad, el prestigio de la facultad de español, el conocimiento de otras lenguas extranjeras y la estancia en un país hispanohablante. Siendo los primeros trabajos de disponibilidad léxica de ELE en China, ambos estudios realizaron análisis tanto cuantitativo como cualitativo de los datos léxicos recogidos y obtuvieron resultados relevantes acerca de la competencia léxica de los sujetos que constituyeron la muestra. Cabe mencionar que, tanto en el trabajo de Lin como en el de Hidalgo, los informantes provienen de universidades que se cuentan entre las mejores de China en lo referente a su calidad de enseñanza del español y, así, no reflejan los distintos perfiles de todos los alumnos de ELE en China. Por otra parte, las ciudades en las que se ubican las universidades estudiadas están muy distantes entre sí⁷⁶ y son apenas capaces de representar todas las zonas del país si tenemos en cuenta su inmenso territorio. Por eso, en esta tesis, en la selección de las muestras, nos centramos en una zona específica del país —Tianjin (China)—. Al mismo tiempo, tenemos en cuenta a alumnos de todos los perfiles diferentes de la zona: estudiantes de universidades tanto públicas como privadas y tanto destacadas como medianas (véase el capítulo de la metodología).

El segundo tipo de estudios según Hidalgo (2017d) y Aabidi (2019), que tiene como objeto de investigación a los estudiantes de ELE en situación de inmersión lingüística, se ha llevado a cabo en muchas zonas del España (Salamanca, Huelva, Sevilla, Cádiz, Andalucía en conjunto, Madrid, Alcalá de Henares e Ibiza) y de Hispanoamérica (México y Cuba)⁷⁷. Bajo estos estudios realizados en zonas con el español como lengua materna, además de las

universidad, 06. La ciudad, 07. El medio ambiente, 08. Medios de transporte, 09. Los animales, 10. Ocio y tiempo libre, 11. Profesiones y trabajos, 12. Tiempo meteorológico y clima, 13. Acciones y actividades habituales, 14. Aspecto físico y carácter, 15. La familia, 16. Viajes y vacaciones, 17. El dinero y 18. Tecnología e Internet.

⁷⁶ Según Lin (2012), los informantes son estudiantes de cinco ciudades: Cantón, Pekín, Shanghái, Xi'an, Chongqing y Luoyang. En Hidalgo (2019), los estudiantes encuestados provienen también de cinco ciudades: Catón, Pekín, Xi'an, Chongqing y Jilin.

⁷⁷ Pueden consultarse los artículos de Hidalgo (2017d) y de Aabidi (2019) para conocer más detalles sobre los estudios de ELE en situación de inmersión lingüística en dichas zonas de España e Hispanoamérica.

investigaciones que recaudan el léxico disponible de los informantes, también se destacan dos sub-tipos de trabajos: 1) los que investigan muestras de inmigrantes en España y 2) los que aprovechan los resultados cuantitativos y cualitativos del léxico disponible para variados objetivos: comparación entre hablantes extranjeros y nativos, evolución léxica de los informantes extranjeros, análisis de las características culturales del léxico de aprendices extranjeros y cotejo del léxico disponible de estudiantes extranjeros con el vocabulario de manuales didácticos de ELE.

4. METODOLOGÍA

4.1. Entorno de la investigación

4.1.1. Tianjin

En China, las provincias forman parte de las divisiones administrativas inmediatamente inferiores al Estado. La provincia es la división administrativa de primer nivel más conocida. De esta manera, una provincia en China es más o menos equivalente a una comunidad en España o a un estado en los Estados Unidos. Además de las provincias, todavía hay otros tres tipos de división que, junto con las provincias, constituyen el total del primer nivel administrativo. Los otros tres tipos son: regiones autónomas, municipios bajo jurisdicción central y regiones administrativas especiales. La actual organización territorial de China incluye 23 provincias, 5 regiones autónomas, 4 municipios bajo jurisdicción central y 2 regiones administrativas especiales. Es decir, China cuenta con 34 divisiones de primer nivel, o lo que viene a ser lo mismo, de nivel provincial.

Tianjin —cuya ubicación en el territorio de China se puede ver en el mapa 4.1— se cuenta entre los cuatro municipios bajo jurisdicción central. Por su cualidad de ser unidad territorial de nivel provincial, tiene importancia política y económica similar a una provincia, a pesar de ser denominada como municipio. Asimismo, se diferencia, en muchos sentidos, de los simples municipios —los que son subdivisiones administrativas inferiores a las provincias—. Por ejemplo, es mucho más grande y tiene bastante más población que un simple municipio, aunque menos que una provincia de sentido general.



Mapa 4.1. Ubicación del municipio de Tianjin en el territorio de China

El municipio de Tianjin es una ciudad portuaria situada en el norte del país. Está al lado de la capital, Pekín (Beijing). Se caracteriza por su importancia en la industria, el comercio y las finanzas. En cuanto a la educación, que es lo que nos interesa, Tianjin dispone de recursos educativos también iguales a los de una provincia. La ciudad tiene 30 universidades y 27 centros de educación superior de formaciones profesionales.

4.1.2. El español en China

La enseñanza del español como lengua extranjera en China ha gozado de un periodo de auge a partir de la primera década del presente siglo. Según comunicó la Asociación China de Enseñanza e Investigación del Español y Portugués en el Simposio Nacional de Enseñanza del Español en Centros de Educación Superior del año 2016⁷⁸, hasta este momento, llegó a 67 el número

⁷⁸ Se celebró del día 13 al 15 de octubre de 2016 en la Universidad de Estudios Internacionales de Shanghai.

de universidades chinas en las que se impartía la lengua española como especialidad universitaria —concretamente, la Filología Hispánica— (véase la Tabla 4.1). Si nos fijamos en tres años más tarde, es decir, a finales de 2019, ese número había llegado a ser 137.

Año	Número de universidades
1952	1
1999	12
2016	67
2019	137

Tabla 4.1. Número de universidades chinas con especialidad en Filología Hispánica según el año

Un simple vistazo a estas cifras nos revela un crecimiento considerable con respecto al desarrollo de la enseñanza de ELE en China en tan solo tres años, puesto que se ha duplicado. Así estando las cosas, no hay que olvidar que, hasta finales del siglo pasado —en concreto, hasta 1999—, eran solamente 12 las universidades con un departamento de Filología Hispánica. El establecimiento del primer departamento de lengua española tuvo lugar en 1952, lo que supuso el inicio de la enseñanza de este idioma en China. En resumen, tal y como se muestra en el Gráfico 4.1, el siglo XXI ha sido el testigo de una evolución explosiva de la enseñanza de ELE en China en el nivel universitario.

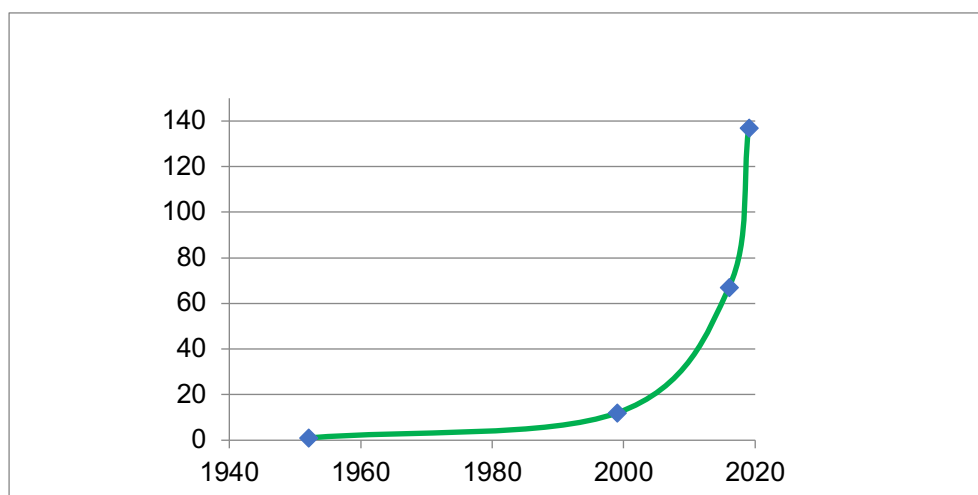


Gráfico 4.1. Número de universidades chinas con especialidad en Filología Hispánica según el año

La presente tesis aborda solamente la enseñanza de ELE en las universidades, porque estas disponen de la mayoría absoluta de los recursos educativos correspondientes a la pedagogía del español como lengua extranjera. En otras palabras, los estudiantes bien preparados que entran cada año en el mercado laboral han recibido su formación completa de la lengua española en la universidad y no en ningún otro tipo de institución. La enseñanza del español en la escuela secundaria, a pesar de su arranque hace tiempo en algunos institutos, no vio grandes desarrollos hasta 2018 con la inclusión de la lengua española en el nuevo programa de educación secundaria publicado por el Ministerio de Educación chino. En los centros de Primaria, el español tiene poca presencia hasta el punto de que suelen ser omitidos al hablar panorámicamente de la pedagogía de esta lengua en China. Por último, el español también se enseña en algunos centros o escuelas de formación de lengua extranjera privados, donde se ofrecen cursos de español de corta duración y a tiempo parcial. Pero, a nuestro juicio, se trata de cursos para estudiantes que tienen un interés específico por el aprendizaje de dicha lengua extranjera o que sienten necesidad de usarla en algún momento de su trabajo. De modo que estos aprendices adquieren cierta competencia comunicativa del español y no se forman para ser profesionales en la lengua española. Además, nos centramos particularmente en el nivel del grado de la educación superior, puesto que este ofrece a los alumnos una formación de la

lengua española desde un umbral cero hasta permitirles estar bien preparados como para acceder al mercado laboral. No abordamos los estudios de posgrado del español, incluidos el máster y el doctorado, por la evidente diferencia que tienen estos en su objetivo para formar a investigadores o académicos, que está más allá de la mera adquisición de una lengua extranjera.

4.2. La muestra y la obtención de las encuestas

4.2.1. Los centros de enseñanza seleccionados

El municipio de Tianjin dispone de tres universidades en las que existe un departamento de Filología Hispánica. Todas ellas son objeto de estudio de la presente tesis, con el fin de investigar la disponibilidad léxica, tanto en español como lengua extranjera (ELE), como en chino como lengua materna (CLM) de los estudiantes de ELE –como se ha comentado más arriba–, en el nivel del grado en los centros de educación superior enmarcados dentro del municipio en cuestión. La recogida de datos se ha hecho a través de encuestas que se pasaron a estudiantes del grado de Filología Hispánica en las tres universidades referidas. Al contrario de las investigaciones del *Proyecto Panhispánico* y de algunas otras de léxico disponible en ELE, en las cuales se analiza a los estudiantes del último año de la educación secundaria, nos fijamos solamente en la enseñanza universitaria en el nivel del grado. Por lo tanto, los informantes son estudiantes chinos de ELE en estas universidades que no tienen antecedentes académicos sobre la lengua española, es decir, que no han aprendido el español en la secundaria y solo empiezan a tener conocimientos sobre el español cuando inician su estudio en la universidad. Por ejemplo, en una de las tres universidades, en el curso académico 2017-2018, 4 de entre 220 alumnos estudiaron el español en la secundaria, o sea, antes de ingresar en la universidad. Estos 4 informantes fueron excluidos de nuestras encuestas.

Las tres universidades en las que pasamos las cuestas presentan tres

tipos diferentes de rasgos, porque, siendo así, representan todos los tipos de universidad con enseñanza de ELE que existen en China. La primera es una universidad pública integral. Ofrece una gama completa de disciplinas. Los estudios e investigaciones que se realizan en ella pertenecen a todos los ámbitos de las ciencias y de las letras. La segunda se fundó como una universidad pública específica en lenguas extranjeras, pero, a lo largo de su evolución y de algunas reformas aplicadas en la institución, se ha convertido en una universidad pública profesional en estudios extranjeros y, asimismo, con algunas ramas de ciencias humanas y sociales. La tercera es una de las muchas universidades emergentes de tipo “privado”. Estas, vueltas a resurgir a partir del siglo XXI, no son universidades públicas, pero tampoco son completamente privadas, porque, por un lado, administrativa y, sobre todo, fiscalmente, obtienen inversiones privadas de empresas y corporaciones para desarrollar y mantener las infraestructuras y su sistema orgánico y, por otro lado, en algunas parcelas académicas y administrativas, se aprovechan y dependen de los recursos de una universidad pública a la que se subordinan. Generalmente, una universidad de este tipo se centra en unas ramas limitadas, sea de letras o de ciencias, por su reducido volumen institucional. La tercera universidad analizada en la presente tesis es una de estas privadas con inclinación por los estudios extranjeros, o sea por las lenguas extranjeras.

4.2.2. Recogida de las encuestas y composición de la muestra para el análisis de la disponibilidad léxica en ELE

La muestra para analizar la disponibilidad léxica en ELE de los estudiantes chinos de ELE fue, al principio, diseñada para componerse de 300 informantes. Según estipula el Ministerio de Educación chino en lo que se refiere a la carrera universitaria de lengua extranjera, los estudiantes realizan los estudios del grado en un ciclo de cuatro años. De manera que, en un principio, pensamos en investigar a 75 estudiantes en cada curso académico del grado, con lo que los informantes analizados cubrirían el total de los niveles del grado. De igual modo, nos propusimos pasar y recoger encuestas de cantidad proporcionada procedentes de las tres universidades en cuestión, con el objetivo de que los

encuestados representaran, equilibradamente, cada tipo de centro —los tres tipos bien distinguidos comentados en el apartado anterior—, manteniendo la distribución diseñada sobre los niveles de estudio. No obstante, a la hora de ponernos en contacto con los profesores de las universidades para ir discutiendo y preparando cómo pasar las encuestas, nos dijeron, unánimemente y sin excepción, que no era posible romper la programación de las clases ni intervenir en las mismas con motivo de la encuesta. Había que añadir que la rutina universitaria que seguía cada estudiante de lengua española no era exactamente la misma porque, además de las asignaturas de especialidad en Filología Hispánica, cada uno tenía, por su parte, asignaturas de educación general y otras optativas. Dichas circunstancias hacían que el tiempo libre fuera de las clases de un informante, con mucha dificultad, coincidiera con el de otro. La única manera de reunir a los estudiantes para hacerles presencialmente la prueba asociativa del léxico era buscar horas marginales, en las que no tuvieran programado ningún asunto importante, como a la hora de descanso al mediodía después de un almuerzo rápido, o cuando se hace muy tarde al finalizar la programación de todas las actividades universitarias. Estábamos bien advertidos de que estas horas marginales del día eran fatales para llevar a cabo la prueba asociativa del vocabulario, puesto que a estas horas los informantes seguramente rendían mucho menos. Por otra parte, debido a que los estudiantes no estaban programados para presentarse en el aula de la prueba y que tenían que sacrificar su tiempo de descanso y animarse a participar en unas encuestas voluntarias a una hora marginal con bastante fatiga, al final no pudimos conseguir el equilibrio pensado en la cantidad de participantes, a pesar de que los profesores de las tres universidades se habían ofrecido a animar a sus alumnos a que participaran y se habían esforzado en ello. A la hora de recoger las encuestas, no había otro remedio, así es que se nos ocurrió corregir el peso de las cifras posteriormente con algunas medidas estadísticas.

En consecuencia, explicamos a continuación los componentes de la muestra real de los estudiantes de ELE, participantes válidos en la prueba asociativa del vocabulario sobre la disponibilidad léxica. Los profesores o

administradores de las universidades sugirieron que ocultáramos en los resultados de la investigación el nombre del centro universitario por varias razones, sobre todo, porque no querían que la publicación de los mismos se presentara como una especie de prueba-evaluación sobre la competencia léxica de sus estudiantes, lo que se podría utilizar como referencia de comparación entre estudiantes de un centro y otro. Por lo tanto, respetando la voluntad de las autoridades académicas, a la primera universidad pública e integral en diversas disciplinas asignamos el sustituto de “pública generalista”; a la segunda pública y profesional en los estudios extranjeros la llamamos “pública especializada”; mientras tanto, a la última, de tipo privado y centrada en carreras de lengua extranjera, nos referimos con la etiqueta de “privada especializada”. El total de informantes que componen la muestra de nuestra investigación es 288.

Curso	Pública	Pública	Privada	Total
	generalista	especializada	especializada	
1º	23	44	38	105
2º	10	46	28	84
3º	9	25	17	51
4º	17	27	4	48
Total	59	142	87	288

Tabla 4.2. Número de informantes de la muestra para el estudio de la disponibilidad léxica en ELE según el tipo de universidad y el curso universitario

En concreto, la especialidad de Filología Hispánica en la universidad pública generalista alberga un solo grupo en cada curso académico. El número de estudiantes en cada grupo oscila entre 19 y 24. Como se puede ver en la Tabla 4.2, pudimos recoger 23 encuestas válidas en el primer curso de esta universidad, lo que suponía una participación completa de todos los estudiantes del grupo del primer curso. Lamentablemente, en los demás cursos, los estudiantes no se mostraron tan participativos. Así que solamente pudimos contar con 10 informantes en el segundo curso —la mitad del total de alumnos en este curso—, así como con 9 y 17 en el tercero y el cuarto cursos,

respectivamente. Como resultado, entre los 288 informantes totales, 59 proceden de la universidad pública generalista. La encuesta para recoger el léxico de los sujetos de esta universidad se pasó de una vez, puesto que la profesora había convocado, un mediodía, al alumnado de todos los niveles en un salón de conferencias para completar los cuestionarios.

A diferencia de ello, la participación de los estudiantes de la universidad pública especializada parece mayor. Se debe a que en esta universidad hay dos grupos en cada curso académico y, por el mayor número de estudiantes, decidimos hacer las encuestas en cuatro ocasiones —cuatro mediodías no consecutivos—, cada vez con alumnos de un solo curso en un aula previamente reservada. La cantidad de alumnos en cada grupo es similar a la de la universidad anterior, es decir, de 19 a 25 estudiantes. Se puede calcular que la participación en los primeros dos cursos de la pública especializada era casi completa, con 44 y 46 encuestas tenidas en cuenta respectivamente. Al contrario, en el tercer curso y el cuarto, la universidad pública especializada solamente proporciona, más o menos, la mitad del alumnado —25 y 27 encuestas válidas, respectivamente—. En total, son 142 informantes los provenientes de esta universidad.

Por último, la situación de la universidad privada especializada era más complicada. Esta institución tiene un volumen mayor de alumnado en comparación con las dos anteriores. Da cabida a tres grupos de estudiantes en cada curso académico y cada grupo cuenta con una media de 25 aprendices. Resultaba evidente que cuanto mayor fuera el volumen del alumnado, mayor dificultad habría en buscar coincidencia en el tiempo libre de los estudiantes, a lo que se añadía otro hecho que no podía ser ignorado: el campus de la universidad privada especializada está situado en la periferia de una zona nueva del municipio, el transporte público que conecta la zona céntrica de la ciudad con el campus tiene un horario con bastantes restricciones y el tiempo de viaje en autobús es de alrededor de una hora y media. Al buscar la conformidad en el tiempo de los alumnos, también teníamos que considerar el horario restrictivo del transporte para ir a aplicar

presencialmente la prueba. Así pues, concertamos pasar las encuestas a los alumnos un mediodía y tres tardes al finalizar todas las actividades docentes. En cada una de las horas concertadas hicimos la prueba a informantes de un solo curso académico. Dada la cantidad de estudiantes, la tasa de participación de la universidad privada especializada fue relativamente baja, sobre todo, en los niveles altos —el tercero y cuarto curso—. Explicaron sus profesoras que los de cursos superiores tenían menos clases y buscaban prácticas en el centro u otras zonas de la ciudad, por lo que no estaban mucho tiempo en el campus, y, precisamente, las horas coordinadas sin actividades universitarias eran las mismas que ellos aprovechaban para ir y venir entre la ciudad y el campus. A fin de cuentas, tuvimos 87 informantes en total en la universidad privada especializada. La cifra iba decayendo de acuerdo con el nivel de estudio: 38, en el primer curso; 28, en el segundo; 17, en el tercero, hasta que solamente se registraban 7 en el cuarto curso.

Las cifras de informantes aquí presentadas reflejan las encuestas válidas y finalmente adoptadas para la investigación. La participación real de los estudiantes supone un número un poco mayor del mostrado. Las encuestas en paquete de folios que pasamos y recogimos en cada una de las pruebas realizadas sumaron 311. Sin embargo, hicimos una revisión sobre los datos y respuestas en los cuestionarios y descartamos 23 encuestas no válidas por diversas razones, las cuales vamos a detallar más adelante en el apartado sobre el procesamiento informático. Estando así las cosas, la muestra final para analizar la disponibilidad léxica en ELE de los estudiantes chinos de ELE se compone de los 288 informantes que ya hemos comentado.

Para finalizar este apartado, añadimos, con respecto a las fechas de las encuestas, que los informantes las completaron en el mes de noviembre del año 2017, o, lo que es lo mismo, a mediados del primer semestre del curso académico 2017-2018. En otras investigaciones, ha habido consideraciones que consisten en que la recogida de datos debe hacerse a finales del segundo semestre de un curso académico, algo en lo que estoy totalmente de acuerdo, puesto que en ese periodo del curso académico los alumnos, probablemente,

tendrán un nivel de conocimiento del español correspondiente a la denominación de su fase de aprendizaje⁷⁹. Sin embargo, a la hora de hacer eso en las universidades chinas, uno se enfrentaría con bastantes obstáculos: generalmente, en el segundo semestre, el investigador encontraría el cuarto curso universitario completamente vacío, sin estudiantes, porque estos no tendrían ninguna clase obligatoria y se habrían ido a hacer prácticas y buscar trabajo. Así que nuestra consideración de pasar las encuestas en el primer semestre y no en el segundo se basó en la idea de intentar evitar la ausencia de informantes de nivel superior. Por otra parte, con respecto a pasar las encuestas a mediados del primer semestre y no a finales del mismo, aclaramos lo siguiente: advertimos, a la hora de coordinar el tiempo libre de los encuestados potenciales, que, a partir del mes de diciembre, los estudiantes se verían involucrados en la preparación y participación en unas pruebas de nivelación o exámenes ^{selectivos}⁸⁰ que consideran sumamente importantes y por los que se desvelan. Para no coincidir con las fechas de estas pruebas o exámenes, ni con las de las propias evaluaciones finales del departamento de Filología Hispánica, ni con algunos días festivos con motivo del fin de año e inicio del año nuevo chino, consideramos mucho mejor ir a pasar las encuestas en noviembre. Por consiguiente, ponemos énfasis sobre el hecho de que, dados los periodos del curso en los que recogimos las encuestas, los informantes, supuestamente, no tenían el nivel de estudio correspondiente al curso universitario en que estaban —o, teóricamente, tenían un nivel de estudio inferior a este—. A pesar de esto, la investigación sigue siendo capaz de dar resultados efectivos, porque la estratificación en los niveles de estudio se mantiene, al tratarse de grupos de informantes pertenecientes a uno u otro curso cuya fotografía lingüística se ha tomado en el mismo punto del proceso de aprendizaje en cada nivel.

⁷⁹ Está claro que sería recomendable buscar un alumno de un nivel cualquiera, por ejemplo, del segundo curso universitario, al final de dicho curso más que en ninguna otra fase de su aprendizaje.

⁸⁰ Citamos como ejemplo las pruebas nacionales del *College English Test*, del *National Computer Rank Examination*, las oposiciones del Estado y de las provincias y los exámenes de inglés TOEFL y IELTS.

4.2.3. Recogida de las encuestas y composición de la muestra para el análisis de la disponibilidad léxica en CLM

La aportación más importante de la presente tesis consiste en analizar la disponibilidad léxica en chino ⁸¹ como lengua materna (CLM) de los estudiantes de ELE en Tianjin y hacer comparaciones cuantitativas y cualitativas entre los datos de disponibilidad léxica en ELE y en CLM. Para ello, además de las encuestas sobre el vocabulario en español referidas en el apartado anterior, también pasamos otras encuestas a los mismos estudiantes, con el fin de descubrir la disponibilidad léxica en lengua materna de los informantes y su diferencia en comparación con la que tienen en español como lengua extranjera.

La muestra para analizar la disponibilidad léxica en CLM de los estudiantes de Tianjin fue diseñada para componerse de los mismos informantes que la muestra de ELE. Antes de tomar esta decisión, se nos había ocurrido otra opción, la que tomaron algunas investigaciones que estudiaron a informantes en situaciones de bilingüismo y aplicaron encuestas también en dos lenguas (Serrano Zapata, 2014: 152). Se trata de dividir a los informantes en dos mitades a la hora de aplicar la encuesta y pedir que cada mitad de los encuestados complete la encuesta en una sola de las dos lenguas —español o chino, en nuestro caso—. Descartamos esta opción por dos razones:

- 1) la división de los estudiantes causaría que el número de informantes en cada lengua se redujera a la mitad y, por consiguiente, obtuviéramos ambas muestras en un volumen limitado; y
- 2) el contraste entre dos grupos de informantes en el momento de la aplicación de la encuesta, generado por la división, contribuiría, a su vez, a una variable nueva —grupo de informantes diferentes—, por lo tanto, afectaría a los resultados de las comparaciones posteriores.

⁸¹ El idioma chino estudiado en esta tesis se refiere al chino estándar moderno, idioma oficial de la República Popular China. Es la lengua de comunicación común en todo el país conocida como *Putonghua* ('lengua general').

Como resultado de estas consideraciones, optamos por la medida de usar a los mismos informantes en las muestras de ELE y de CLM, de modo que aplicamos la encuesta en ELE y la repetimos una vez más en CLM.

De ahí venía otra cuestión: ¿cuánto tiempo debía ser el intervalo entre la aplicación de la encuesta en ELE y la en CLM sobre un mismo informante? A nuestro juicio, la segunda aplicación nunca debía ser inmediata a la primera, porque se imaginaba que los estudiantes, al terminar la encuesta en una lengua y empezar pronto una nueva en otra lengua, tenían muy frescas en la memoria las respuestas que acababan de aportar en la anterior y las podían reproducir, traduciéndolas, en la lengua de la encuesta posterior. Como mínimo, se puede sospechar que, en tan corto tiempo, los informantes mantendrían fácilmente el mecanismo de las asociaciones anteriormente hechas y seguirían actualizándolas en la segunda encuesta, lo que llevaría a unas segundas asociaciones afines a las primeras, aunque no fueran respuestas exactamente iguales a las primeras. Partiendo de esta suposición, el intervalo no debía ser corto.

Tuvimos en consideración los métodos tomados en algunas investigaciones de disponibilidad léxica hechas sobre los mismos informantes en dos lenguas de contacto y que aplicaron dos encuestas en cada lengua. La investigación de Victory (1971) sugería un intervalo de dos o tres semanas para evitar interferencia de una lengua sobre otra (véase el apartado 3.1.7) y la de Njock (1978) definió un tiempo transcurrido de dos semanas (véase el apartado 3.1.8). Más recientemente, Serrano Zapata señaló que optó “de forma intuitiva, por este intervalo (de dos semanas), al no conocer referencias bibliográficas sobre cuál debía ser el margen necesario entre las dos pruebas. No obstante, los pocos estudios existentes realizados también con los mismos informantes en distintas lenguas eligen también casi todos esta opción” (Serrano Zapata, 2014: 253). Dado esto, determinamos, en esta tesis, que el intervalo de tiempo entre las dos aplicaciones de encuesta en lenguas distintas fuera de dos semanas. Estando así las cosas, las encuestas de disponibilidad

léxica en CLM de los estudiantes de Tianjin cayó a finales del mes de noviembre del año 2017. Es decir, dos semanas después de la aplicación de las encuestas en ELE, los alumnos anteriormente encuestados volvían a otra convocatoria para completar las encuestas en su lengua materna.

Siendo los mismos informantes, la composición de la muestra para analizar la disponibilidad léxica en CLM fue pensada para ser idéntica a la tomada mediante las encuestas en ELE. Pero, en el momento de avisar y animar a los estudiantes a que participaran en estas encuestas de segunda remesa, lo que hicimos junto con los profesores no fue dirigirnos solamente a los ya encuestados anteriormente en la primera remesa, sino que hicimos un llamamiento general tal y como habíamos hecho para la encuesta en español. La razón de esto es que estábamos advertidos de un posible descenso de la tasa de participación y, por eso, si solamente hubiéramos pedido que volvieran los ya encuestados, la cifra final de los participantes en esta segunda remesa hubiera sido mucho menor que la anterior. Llegamos a esta conclusión al observar que se mantenía la misma percepción de los alumnos sobre nuestra encuesta y nos dimos cuenta de que estas encuestas presenciales de tipo “prueba cronometrada” eran totalmente raras, difíciles y costosas en su opinión, a lo que se sumaba el hecho de que tenían que abandonar su reducido tiempo de descanso aun con tanta fatiga por la rutina diaria. En resumen, la pasividad notable en los estudiantes no era difícil de detectar, a pesar del ánimo y esfuerzo puesto por los profesores. Como resultado, la cifra de informantes reales que se presentaron a la prueba asociativa, pese al llamamiento general, no cumplió con las expectativas. A continuación, explicamos cómo se ha compuesto la muestra de informantes encuestados en CLM.

Curso	Pública	Pública	Privada	Total
	generalista	especializada	especializada	
1º	10	27	31	68
2º	6	40	6	52
3º	5	10	5	20
4º	10	18	4	32
Total	31	95	46	172

Tabla 4.3. Número de informantes de la muestra para el estudio de la disponibilidad léxica en CLM según la universidad y el curso universitario

Al igual que en las encuestas en ELE, en CLM la universidad pública especializada también fue la que proporcionó más informantes, concretamente, 95. Se percibe que los alumnos de niveles bajos eran más participativos que los de niveles altos. Dado que había dos grupos en cada curso con un número de estudiantes comprendido entre 19 y 25 en cada grupo, la proporción de encuestas válidas en el primer curso suponía un poco más de la mitad del total en este curso, o sea, 27 informantes. Los estudiantes participantes del segundo curso fueron casi la totalidad, con una cifra registrada de 40. En el tercero y cuarto curso, el número de encuestados fue de 10 y 18, respectivamente.

Por último, al igual que en el caso de la muestra en ELE, en las encuestas en CLM la universidad privada especializada presentaba situaciones aún menos controladas y, por consiguiente, aportaciones a la muestra bastante reducidas. Dado el mayor volumen del alumnado en la universidad privada especializada —tres grupos en cada curso y una media de 25 estudiantes por grupo—, la tasa de participación fue realmente decepcionante. De las 46 encuestas válidas que pudimos recoger de esta universidad, 31 procedían del primer curso. Los demás cursos apenas contribuyeron a la composición de la muestra, con 6, 5 y 4 informantes, respectivamente, en cada curso.

4.3. Diseño de la encuesta y procesamiento de los datos

4.3.1. El cuestionario y la prueba

4.3.1.1. Los métodos que empleamos en esta investigación para obtener los datos de disponibilidad léxica en ELE y en CLM han sido análogos a los establecidos por el *Proyecto Panhispánico* (véase el apartado 3.2.1), que afectaban tanto a la estructura de nuestra encuesta como a su contenido. La recolección del léxico disponible de cada informante que formaba parte de la muestra se efectuó a través de un cuestionario que consistía en una prueba asociativa de vocabulario planteada en torno a unos centros de interés, o sea, campos temáticos. Durante la realización de la prueba, pedimos a los estudiantes encuestados que escribieran las palabras que les vinieran a la mente en relación con cada centro de interés de que se trataba. Los centros de interés que utilizamos en esta investigación fueron los mismos 16 empleados en muchos de los estudios de disponibilidad léxica del *Proyecto Panhispánico* y, asimismo, los más habituales en los estudios del tema sobre el español como lengua extranjera. Son los siguientes:

01. Partes del cuerpo humano	01. 人体部位
02. Ropa y complementos	02. 服装配饰
03. Partes de la casa (sin muebles)	03. 房屋的各个部分（不含家具）
04. Los muebles de la casa	04. 房间家具
05. Alimentos y bebidas	05. 食物和饮料
06. Objetos colocados en la mesa para la comida	06. 餐桌上的用餐物品
07. La cocina y sus utensilios	07. 厨房及厨房器具
08. La escuela: sus muebles y materiales	08. 学校：教学设施和学习用品
09. Iluminación y calefacción	09. 照明和采暖设备
10. La ciudad	10. 城市
11. El campo	11. 乡村
12. Medios de transporte	12. 交通工具
13. Trabajos del campo y del jardín	13. 田地园林劳作
14. Los animales	14. 动物
15. Juegos y distracciones	15. 游戏、运动、娱乐活动
16. Profesiones y oficios	16. 各种职业与行业

Como se puede ver, los centros de interés en español de nuestra investigación se mantenían idénticos a los clásicos del primer estudio de disponibilidad léxica (véase el apartado 2.2.3) y, por consiguiente, a los utilizados en las investigaciones del *Proyecto Panhispánico* —puesto que estos adoptaron los primeros con motivo de la uniformidad metodológica—, excepto por algunas leves diferencias en la denominación. Por ejemplo, en vez de *El cuerpo humano* o *Partes del cuerpo* en algunas otras investigaciones, empleamos *Partes del cuerpo humano* para referirnos lo más explícitamente posible. Nuestro centro *Alimentos y bebidas* se diferenciaba de algunos denominados “Comidas y bebidas”, pero esta distinción era prácticamente nula con respecto a la función orientativa de los centros de interés a la hora de la aplicación de la prueba. Sin embargo, en el centro 02, sustituimos el habitualmente usado *La ropa* por *Ropa y complementos*, porque tuvimos en cuenta lo que comentó Serrano Zapata (2014: 255) al respecto: “algunos autores ya advierten de que los alumnos enumeran en este campo tanto prendas de vestir como accesorios y complementos” y, revisando bastantes

listas de palabras de este centro en muchos otros estudios, encontramos que era cierto. De ahí que adoptáramos la denominación utilizada por Serrano Zapata para este centro de interés: *Ropa y complementos*.

Por otra parte, los homólogos de los 16 centros de interés en chino se tradujeron directamente de sus denominaciones en español, excepto el centro 15 *Juegos y distracciones*. La palabra “juego” en español hace referencia a las actividades recreativas de diversa índole incluyendo los entretenimientos y los deportes, lo cual se puede comprobar fácilmente en las listas de palabras de muchas investigaciones si nos damos cuenta de la coexistencia de, por ejemplo, “internet” y “fútbol”. No obstante, en la lengua china, los entretenimientos y los deportes se nombran con dos palabras particulares, es decir, no hay en chino una sola palabra que englobe estas dos referencias de la palabra “juego” en español. Por lo tanto, modificamos la denominación de este centro de interés en chino, la cual se interpretaría como “Entretenimientos, deportes y distracciones”, con la finalidad de conservar la función orientativa de la palabra española “juego” en dicho centro. Las demás traducciones al chino de los centros de interés aplicados en esta tesis fueron literales y, al mismo tiempo, explícitas, sin haber añadido ni restado interpretaciones sobre las denominaciones en español.

A la hora de aplicar la prueba mediante cuestionarios en papel, con respecto a cada centro de interés, exigimos no solo a los estudiantes encuestados que anotaran las palabras que se les ocurrieran relacionadas con el centro de interés correspondiente, sino que también las escribieran una tras otra y en orden para que, en fases posteriores de la investigación, el investigador pudiera ver cuál de las palabras escritas era aquella que les había venido primeramente a la mente, cuál se les había ocurrido después y, así, sucesivamente. De este modo, se determinaría que una palabra era más disponible que otra, porque el informante sabía actualizar más rápidamente una palabra que otra en una situación comunicativa bajo cierta temática. Por cada centro de interés, los estudiantes tenían un límite temporal de dos minutos para apuntar sus palabras-respuestas. El lapso de tiempo para

responder a cada centro de interés que aquí aplicamos era el establecido en el *Proyecto Panhispánico* y el unánimemente utilizado en muchos estudios aun fuera del proyecto, toda vez que se había comprobado en estudios anteriores que, después de aproximadamente dos minutos en cada centro de interés, la mayoría de los participantes ralentizaba las respuestas. Desde nuestro punto de vista, los dos minutos eran suficientes para que los alumnos recordaran sus palabras más disponibles, ya que, después de ese tiempo, aparecían palabras mucho menos disponibles que contribuirían poco a la investigación. Además, los dos minutos eran estrictamente cronometrados. No solicitamos a los informantes un número determinado de palabras, así es que, podían escribir cuantas palabras quisieran —lo que coincidía con las listas de palabras abiertas del *Proyecto Panhispánico*, en contraposición a las listas cerradas de algunos otros estudios—.

4.3.1.2. El cuestionario, además de disponer de las páginas con espacios para recolectar las palabras disponibles de los informantes, también contenía dos partes más: la de los datos personales del informante que completaba el cuestionario —de los que se extraerían las variables extralingüísticas de nuestra investigación— y la de las instrucciones sobre la prueba. Cabe destacar que el cuestionario que se utilizó para recoger el léxico disponible en ELE y el empleado para el léxico en CLM eran el mismo. Más concretamente, utilizamos un cuestionario bilingüe en español y chino⁸². La idea de hacer una encuesta en las dos lenguas tenía como motivo facilitar la comprensión lectora de los participantes, sobre todo, de los alumnos de nivel de estudio bajo. Se imaginaba que una encuesta totalmente en español contribuiría a la pasividad de los estudiantes, porque estos, además de someterse a una prueba de léxico, tendrían que intentar entender el contenido del cuestionario leyendo en lengua extranjera. Por lo tanto, todos los contenidos de nuestra encuesta se presentaron en las dos lenguas en cuestión, incluidos los títulos y opciones de los datos personales, las instrucciones sobre la prueba y las denominaciones de los centros de interés.

⁸² El cuestionario entero se puede consultar en el apéndice 1.

En la cabeza de la página de los datos personales, se pedía a los informantes que apuntaran la fecha de la encuesta, de modo que, cada fecha correspondía a un turno de prueba determinado. Como tomamos nota de las fechas en las que pasamos las encuestas a cada colectivo de alumnos, la identificación de cada fecha y, por consiguiente, de cada turno de prueba servía para saber a qué universidad pertenecía cada informante —según esto último se le asignaría a cada informante una opción de la variable “tipo de centro (público o privado)”—. Fue una especie de codificación sobre el nombre de los centros educativos que protegía el anonimato de los mismos, puesto que nunca se pedía en el cuestionario que escribieran el nombre de la universidad. De esta manera, si, en el futuro cuando la investigación se acabe, se desechan los cuestionarios en papel, al verlos nadie podrá relacionar los contenidos y respuestas en ellos con ningún centro educativo.

A continuación, se pedía en el cuestionario que los informantes indicaran si eran varón o mujer. Luego cada uno debía informar de su edad y de la provincia de la que procedía. Después, se preguntaba por el curso universitario en que estaba estudiando cada informante y este podía marcar entre el primero, el segundo, el tercero y el cuarto. También se indagaba sobre si tenían alguna experiencia de haber estudiado en un país hispanohablante. En el caso que sí, debían hacer saber cuánto tiempo había durado el periodo de estudio.

Por último, se les solicitaba señalar qué nivel tenían del inglés. A este respecto, habíamos pensado en intentar conseguir las notas que obtuvieron los estudiantes en la asignatura de lengua inglesa, pero resultó que esta asignatura se desarrollaba en las tres universidades de manera muy desigual y las pautas según las que estas tres evaluaban a sus propios alumnos también eran distintas. Por eso, se nos ocurrió definir el nivel de inglés con unas pautas comúnmente aplicables a los universitarios chinos. Estas consistían en 1) dos fases de estudio del inglés con niveles de enseñanza bien diferenciados: el bachillerato y la universidad y 2) dos niveles de una prueba nacional del inglés

como lengua extranjera: CET nivel 4 y CET nivel 6⁸³. Tomamos como referencia esta prueba porque está altamente generalizada. Como es una prueba nacional de peso, comúnmente reconocida en todo el país y, sobre todo, de fácil acceso, veíamos, según nuestra experiencia empírica, que todos los alumnos tenían una gran inclinación a participar en ella siempre que pudieran. Los que no tomaran parte en ella, muy probablemente, no tendrían un nivel correspondiente. Por tanto, la prueba CET podía servir como una pauta general aplicable a todos nuestros informantes para determinar su nivel de inglés. En el caso de que un informante no hubiera aprobado dicha prueba, le preguntábamos si su nivel de inglés correspondía, más o menos, a la enseñanza del bachillerato o a la de la universidad, puesto que, por lo menos, lo habría estudiado o estaría estudiándolo en una de las dos fases. Asimismo, también se pedía en el cuestionario que especificaran, en su caso, algunos otros cursos o certificados de inglés que hubieran obtenido, por si tenían algún nivel de excelencia. Si el informante no sabía nada de inglés, podía simplemente marcar la opción de no saber el idioma. En fases posteriores de la investigación, vamos a asignar una valoración a cada una de las opciones que han dado los encuestados sobre el nivel de inglés.

4.3.1.3. Después de la página para recabar los datos personales de cada informante, el cuestionario iba con otra página de instrucciones sobre la prueba asociativa del vocabulario. En primer lugar, se informaba a los estudiantes del mecanismo de la prueba. Se señalaba que 1) la encuesta contenía 16 temas correspondientes a 16 columnas —enumerados desde A 01 hasta P 16⁸⁴—, 2) que el investigador les iría indicando progresivamente cada tema, y 3) que, en

⁸³ CET es la sigla en inglés de *College English Test*. Se trata de una prueba de inglés como lengua extranjera en China. Está destinada a los universitarios que estudien dicho idioma, pero, no como especialidad universitaria. Antes fue obligatoria porque examinaba el dominio del alumno en la lengua inglesa para garantizar que alcanzaran los requisitos del aprendizaje de algunas asignaturas parcialmente en inglés —aún pocas para los de otras especialidades que no fuera la lengua inglesa—. Ahora es una prueba optativa en el ámbito nacional y sigue siendo bien estimada y de peso. Tiene solamente dos niveles, que son CET nivel 4 y CET nivel 6.

⁸⁴ Las columnas correspondientes a los 16 temas, o sea, 16 centros de interés, se marcaron en el cuestionario con letra y número —por ejemplo, P 16 hace referencia al centro de interés número 16—, con el fin de diferenciarse de las líneas, marcadas en cada columna solo con un número, en las cuales se escribían las palabras.

cada columna desde la línea 1 hasta la 30, debían escribir durante dos minutos cronometrados todas las palabras que se les ocurrieran relacionadas con el tema que les hubiera dado el investigador. Se puede ver que el mecanismo era el mismo que el establecido en los estudios del *Proyecto Panhispánico*, pero hicimos una modificación en lo que se refiere a la presentación de cada centro de interés. En muchas investigaciones, la designación de los centros de interés se mostraba en la cabeza de las columnas en las que los informantes escribían las listas de palabras, de modo que estos, en el momento de completar la lista de un centro, podían ver cuál era el siguiente o, incluso, los siguientes. El enfoque en los campos temáticos que quedaban por rellenar podía desviar el pensamiento de los alumnos y ello podía tener repercusiones en sus respuestas. Es más, en el peor de los casos, los alumnos podían pasar por alto el tiempo cronometrado y empezar a dar respuestas en el siguiente centro antes de que se acabaran los dos minutos, o incluso volver a alguno anterior.

En consecuencia, empleamos en esta tesis el método que había planteado Hernández Muñoz (2006: 262-263), quien señaló que “en las columnas no rezaba el epígrafe del centro de interés correspondiente, con el objetivo de no interferir en las respuestas de ninguna de las áreas temáticas. El entrevistador, o el profesor que ejercía como tal, era el encargado de dictarlas en voz alta antes del comienzo de los dos minutos correspondientes a cada uno de los campos”. Dado esto y para que los informantes se concentraran en cada campo léxico durante dos minutos efectivos, escondimos en nuestro cuestionario las denominaciones de los centros de interés, pusimos en su lugar una letra y un número, tal y como se especificó en las instrucciones de la prueba y acabamos de comentar, y fuimos avisando, a lo largo de la prueba, del tema correspondiente a cada columna. A diferencia de lo planteado por Hernández Muñoz, quien dictó los centros de interés, nosotros mostramos a todos los participantes un cartel con la denominación — en español y en chino— del centro correspondiente antes de proceder a cronometrar los dos minutos correspondientes. En nuestro caso, el hecho de hacer dictado en lengua extranjera podría obstaculizar la comprensión de los informantes y afectar sus respuestas. De todas maneras, la realización de

nuestras pruebas tuvo en cuenta todos esos detalles para que las asociaciones léxicas hechas en cada columna se realizaran de manera realmente controlada y poco desviada.

Estando así las cosas, lo que se enseñaba a continuación en las instrucciones de la prueba era que no hacía falta completar todas las líneas de una columna, pero si sus respuestas rebasaban las 30 líneas previamente trazadas, podían seguir en el dorso del folio siempre que no se hubieran acabado los dos minutos cronometrados. Por el contrario, se alertaba de que, si daban por terminada una columna sin poder actualizar más palabras antes de que venciera el límite temporal, debían esperar tranquilos hasta que el investigador pusiera fin a ese centro de interés y diera indicación de comienzo para la siguiente columna. Además, con el objetivo de recolectar las más respuestas posibles, se instruía a los participantes en no coartarse por motivos de dudas ortográficas. Se les animaba a poner una respuesta incluso si no estaban seguros de la palabra o de su grafía, hasta el punto de que podían apuntar palabras malsonantes. Estas palabras las podríamos corregir e incluir o descartar y excluir posteriormente según los criterios de edición. Junto con eso, se advertía en las instrucciones de que no debían preguntar, ni al investigador o al profesor ni a sus compañeros, cómo se escribía determinada palabra o si se podía escribir una palabra u otra, porque las preguntas podían orientar las respuestas de los demás. Asimismo, con la finalidad de garantizar que la prueba reflejara fielmente el léxico disponible de los informantes, se explicaba que, aunque sus respuestas fueran pocas, no debían –y tampoco era necesario– copiar las respuestas de los compañeros, subrayando que la prueba no tenía nada que ver con la calificación de ninguna asignatura. Al final, se ponía énfasis en el anonimato de los informantes para asegurarles que no era posible identificar a ninguno de ellos mediante la encuesta, con lo cual se intentaba conseguir que los estudiantes efectuaran asociaciones realmente libres y sin limitaciones.

En resumen, el cuestionario fue repartido, de manera presencial y en el momento de la prueba, a los alumnos participantes a través de un paquete de

hojas formado por las tres partes arriba referidas, que eran, en orden, la de los datos personales, la de las instrucciones y la de las columnas y líneas donde poner las respuestas. Tenía en total seis hojas, las dos primeras partes ocupaban sendas hojas y la última empleaba cuatro, de manera que un folio daba cabida a cuatro columnas, o lo que es lo mismo, a cuatro centros de interés. Concertamos las fechas y horas y reservamos las aulas tal y como hemos comentado en el apartado 4.2.1. Fue el mismo autor de esta tesis el que siempre se presentó a la hora y en el aula de cada una de las pruebas para controlar y garantizar el desarrollo de ellas, porque conocía mejor el mecanismo y los procesos que los profesores de los alumnos —aunque en algunas ocasiones el profesor o profesora de los estudiantes también estaba presente acompañando por si estos no tomaban en serio el acto—. A ello se añade otra ventaja de que el investigador estuviera siempre presente: las pruebas paralelas en distintos grupos de informantes se llevaban a cabo siguiendo unos procedimientos y detalles homogéneos y, por consiguiente, generaban condiciones parecidas de asociaciones y repuestas en los informantes. Antes de la prueba, el investigador aclaraba brevemente el motivo de la misma y agradecía la colaboración de los presentes. En el caso de las pruebas en ELE, el investigador siempre recordaba que habría una segunda prueba en CLM. A continuación, el investigador pedía a los participantes leer las instrucciones en la encuesta mientras explicaba en voz alta el contenido de las mismas. Una vez iniciada la prueba cronometrada, no se interrumpía hasta que esta llegara a su fin. Dentro de ella, además de los dos minutos concedidos a cada centro de interés, se dejaba unos segundos como pausa de descanso entre el final de un centro y el comienzo del siguiente. Así pues, una prueba completa en la presente investigación duró unos 50 minutos.

4.3.2. *Procesamiento informático*

4.3.2.1. Inclusión o exclusión de las encuestas cumplimentadas

En este apartado vamos a explicar cómo funcionó el procesamiento, mediante un programa informático, de los datos que obtuvimos de las

encuestas. Antes de ello, habíamos de decidir cuáles de los materiales recogidos eran los que se debían someter al procesamiento informático. Como hemos mencionado en el apartado 4.2.1, el número final de las encuestas en ELE que pudimos incluir en la investigación para su posterior análisis era un poco menor que la cantidad de los participantes en las pruebas —288 encuestas finalmente incluidas frente a 311 recogidas—. Esto se debe a que descartamos y decidimos no incluir algunas encuestas que considerábamos defectuosas y, por eso, no válidas. Hicimos una revisión de los materiales recabados en las encuestas y tomamos decisiones sobre si excluir o no cada encuesta de la muestra y, por consiguiente, del procesamiento informático. Finalmente, 23 encuestas en ELE fueron exceptuadas porque presentaban uno o más de uno de los siguientes rasgos.

- 1) No haber rellenado los datos personales, por lo tanto, las respuestas resultaban ser de índole indefinida ya que no se podía asignar las opciones de las variables al informante que había dado sus respuestas. Esto sucedió en seis encuestas, que, por tal razón, quedaron descartadas.
- 2) No aportar palabras en las diferentes columnas, Esto sucedió en tres encuestas en las que el informante solamente reprodujo el epígrafe de los centros de interés en la primera línea de cada columna, sin aportar ninguna otra palabra.
- 3) Faltar algún folio. Eso sucedió en una encuesta, en la que había desaparecido la última hoja del paquete con los últimos cuatro centros de interés. Se comprobó que dicha hoja se había desprendido y las respuestas, al haberse perdido un cuarto de las mismas, resultaron deficientes.
- 4) Responder todo el cuestionario en inglés. Para nuestra sorpresa, descubrimos un caso particular en el que el informante había apuntado sus respuestas, al cien por cien, en inglés. Aunque, en general,

preferíamos mantener las respuestas en inglés y no negábamos una posibilidad de hacer futuras investigaciones sobre el anglicismo en las listas de palabras, decidimos desechar esta encuesta única porque veíamos que contestar en inglés en su totalidad iba en contra de nuestra iniciativa de recabar el léxico en español como lengua extranjera y tampoco aportaba informaciones sobre la interferencia del inglés en el español adquirido por el informante.

- 5) Escribir de manera ilegible. Curiosamente, había otro caso igual de particular que el anterior. En esta encuesta, el informante no había dado sus respuestas ni en español ni en inglés, sino que había escrito las palabras —aun en alfabeto latino— todas absolutamente indescifrables. A nuestro juicio, dicho estudiante o bien sabía manejar una “lengua” muy alejada del propósito de nuestra investigación o bien no tomó en serio la encuesta, aparentando contestar esta a la hora de la prueba.
- 6) Contestar menos de la mitad de los centros de interés. Las demás encuestas, hasta el total de 23, fueron excluidas porque sus informantes solo habían contestado algunas pocas columnas. Sin embargo, llamamos la atención sobre el hecho de que nunca descartamos una encuesta porque su informante no hubiera completado todas las columnas. Está claro que un alumno podía tener disponibilidad deficiente en alguno o algunos de los centros de interés y dejó vacío una o varias columnas, lo que era totalmente aceptado. Las encuestas que no admitimos para el procesamiento informático fueron aquellas que tenían vacía, más o menos, la mitad de las columnas y en cuyas pocas columnas rellenas, las respuestas eran realmente escasas —una o, en el mejor de los casos, tres—. No sabíamos si dicha situación se debía a una disponibilidad verdaderamente baja o a que no tomaron parte en la encuesta con responsabilidad y en serio. Por eso, puesto que estas encuestas denotaban informaciones bastante ambiguas con respecto a la

disponibilidad léxica del informante, decidimos no emplearlas.

Dentro de las encuestas con columnas no completadas en su totalidad —las que se admitían en la investigación excepto si se trata del caso anterior—, había dos casos especiales. Estos dos alumnos llegaron tarde al acto de la prueba, de modo que dejaron sin contestar, respectivamente, los primeros tres y cuatro centros de interés. Como no les daba tiempo para volver a hacer la prueba en esos centros al finalizar el acto, tuvimos que desechar sus encuestas ya que aportaron informaciones parciales e inconsecuentes —en ambos casos, no contestaron nada en las primeras columnas, pero muy bien en el resto—.

Cabe destacar que, en cuanto a las encuestas en CLM, ninguna fue exceptuada, es decir, pudimos incluir en la investigación y someter al procesamiento informático los materiales en todas las 172 encuestas recogidas en CLM. En nuestra opinión, la validez total de estos datos se debe a que los participantes en esta segunda remesa de pruebas estaban todos de veras interesados y fueron responsables con la investigación, toda vez que aquellos con menos ganas de participar no se animaron y no volvieron a presentarse. Asimismo, como era la segunda vez que completaban el cuestionario, conocían mejor el mecanismo y los requisitos de la prueba y aportaron encuestas satisfactorias.

4.3.2.2. Uso del programa *Dispolex*

Una vez determinadas las encuestas válidas y los datos utilizables, recurrimos al programa informático *Dispolex*, accesible en la página web <http://www.dispolex.com> (véase el apartado 3.2.1), para el procesamiento de los datos. Como se ha comentado en el capítulo anterior, es el sitio web coordinador de las diferentes investigaciones del *Proyecto Panhispánico*. Cuenta con un banco de datos de disponibilidad léxica, donde se dan de alta los distintos proyectos de investigadores de todas partes. Aunque se ha creado para coordinar los estudios del *Proyecto Panhispánico*, está también abierto a

otros que no estén en el Proyecto. Cualquier proyecto dado de alta en dicho banco de datos puede aprovecharse de las funciones del programa informático.

La manera de insertar los datos en el programa tiene que ver con si la investigación en cuestión pertenece o no al *Proyecto Panhispánico*. Como este establece unas pautas comunes, una investigación del Proyecto dada de alta tiene, automáticamente, predeterminados los centros de interés, las variables y las opciones de estas, y el investigador puede insertar sus datos o bien manualmente o bien desde un documento ya existente en formato txt. En nuestro caso, como investigación fuera del Proyecto, debíamos, primeramente, definir los ítems. Antes de nada, tuvimos que introducir los centros de interés estudiados en esta investigación. Después, incorporamos las variables tenidas en cuenta en la investigación y sus opciones, para luego corresponder cada una de estas con el informante tratado. A continuación, el programa pidió que añadiéramos diferentes grupos de estudio. En nuestro caso, creamos tres grupos de estudio y, en cada uno, agrupamos los informantes procedentes de cada una de las tres universidades. A cada grupo le pudimos asignar una opción de la ubicación —urbano o rural— y una opción del tipo de centro —público o privado—. Así pues, debajo de cada grupo, pudimos ir añadiendo a los informantes, escogiendo una opción que le correspondía de cada variable para determinar el perfil de los mismos. Una vez aquí, ya pudimos empezar a insertar las respuestas aportadas por cada informante debajo del fichero del mismo. En el fichero de cada informante, aparecían los botones de los centros de interés. Cuando se finalizaba la inserción en un centro, se podía cambiar de centro para pasar a otro. Hemos de advertir de que la inserción de las palabras en el programa, si se trata de una investigación fuera del *Proyecto Panhispánico*, es totalmente manual. Es decir, no se permite la inserción automática mediante ningún otro documento existente, sea fichero de texto o de un excel.

El mecanismo de la inserción de las respuestas era el siguiente. Debajo de cada informante y de cada centro de interés, el programa exponía dos columnas donde poner las palabras. En la columna izquierda se debía copiar

las respuestas originales del informante, o sea, las respuestas dadas por él sin ninguna modificación, mientras tanto, en la derecha, se ponían las palabras sometidas a los criterios de edición. Nos dimos cuenta, en fases posteriores, de que la ordenación de las palabras y los cálculos matemáticos se realizaban a partir de las palabras sometidas a los criterios de edición, es decir, las de la columna derecha. Así que las respuestas de la columna izquierda servían como una copia de los datos originales y, por consiguiente, se podían extraer para otros usos —como para analizar la ortografía—, aunque también manualmente. Por último, en lo que se refiere a la lengua insertada, descubrimos que el programa no solo se había diseñado para editar las palabras en español, sino que también era capaz de mostrar y procesar los caracteres chinos de las encuestas en CLM.

A continuación, vamos a presentar los cálculos matemáticos que nos pudo ofrecer el programa *Dispolex*. En primer lugar, permitía sacar listas de palabras de cada uno de los centros de interés según variables, tanto las correspondientes al grupo de estudio —la ubicación o el tipo de centro— como las relativas al informante. Para cada una de las palabras en una lista, el programa mostraba cuatro cifras: 1) el índice de disponibilidad, 2) la frecuencia relativa, 3) el porcentaje de aparición y 4) la frecuencia acumulada. En cuanto al índice de disponibilidad, *Dispolex* empleaba la fórmula matemática desarrollada por López Chávez y Strassburger Frías (véase el apartado 3.2.2.1.3). La frecuencia relativa resultaba de dividir el número de aparición de una palabra entre el número total de palabras. El porcentaje de aparición, a su vez, era el resultado de dividir el número de veces que aparecía una palabra entre el número total de informantes⁸⁵. Estas tres cifras eran propias de una palabra, es decir, eran inalterables una vez determinadas las variables. No obstante, la frecuencia acumulada se refería a la suma de las frecuencias relativas en la lista tratada y podía cambiar si variaba la ordenación de las palabras. Servía para dar a conocer la posición de una palabra en una lista

⁸⁵ El número de aparición de una palabra es lo mismo que el número de informantes que han dado una palabra. Por eso, en los cálculos de porcentaje de aparición, también se puede decir que resulta de dividir el número de informantes que han dado una palabra entre el número total de informantes.

determinada. Estando así las cosas, las palabras de una lista podían ordenarse según cada uno de los tres parámetros propios de ellas: según el índice de disponibilidad, la frecuencia relativa o el porcentaje de aparición, respectivamente. Además, el programa también favorecía otra forma de ordenación de las palabras, que era según el orden alfabético. Lamentablemente, esta última forma de ordenación solo nos funcionó en los resultados en ELE y las listas de palabras en CLM no pudieron presentarse en orden alfabético. Comprobamos que el programa, a pesar de que sabía mostrar los caracteres chinos tal y como estos habían sido insertados para después procesarlos, no era capaz de enumerarlos en orden de acuerdo con el alfabeto latino del sistema de transcripción fonética de esta lengua⁸⁶.

A lo largo de nuestro uso de *Dispolex* para generar las listas de palabras, hemos encontrado una deficiencia del programa, de la que consideramos necesario advertir a los futuros investigadores que quieran emplearlo. Esta es que el programa no permite cruzar las variables relativas al informante. En otras palabras, si se fijan las opciones de más de una variable, el programa falla y no da ningún resultado. Por ejemplo, en esta investigación, si se quiere sacar las palabras de las mujeres (opción de la variable “sexo”) del primer curso (opción de la variable “nivel de estudio”) en un centro de interés determinado, no podrán obtenerse resultados. De igual modo, si quiere conseguirse resultados con tres o más variables relativas al informante definidas, el programa falla. Pero, si solo se sacan las palabras de las mujeres o solo las del primer curso, no habrá ningún problema. Al detectar dicho defecto, nos pusimos en contacto con el equipo de *Dispolex*, pero hasta la fecha no se han encontrado soluciones al respecto. Esta cuestión afectó a los análisis que íbamos a hacer con respecto a una variable que habíamos diseñado para la presente investigación, lo cual vamos a explicar en el siguiente apartado que trata sobre las variables extralingüísticas.

⁸⁶ Se trata del sistema de representación de la lengua china mediante el uso del alfabeto latino. Dicho sistema es comúnmente conocido en China como “pinyin”, o sea, literalmente en español, “deletreo de los sonidos del chino”. En la mayoría de los casos, es según el alfabeto del “pinyin” como los diccionarios de la lengua china ordenan sus palabras.

Además de las listas de palabras, *Dispolex* todavía es capaz de calcular otros índices cuantitativos relativos a los centros de interés. Estos son 1) palabras totales, 2) palabras diferentes, 3) palabras por informante y 4) índice de cohesión. El primero se refiere al número total de palabras que dan los informantes en un centro de interés; el segundo indica el número de palabras diferentes en un centro de interés, o, lo que es lo mismo, el número de vocablos en un centro; el tercero consiste en el promedio de palabras en un centro; mientras tanto, el cuarto refleja la coincidencia entre las palabras que dan los informantes en un centro, y resulta de dividir el promedio de palabras por informante en cada centro de interés entre el número de vocablos del mismo centro (véase el apartado 3.2.2.1.2 y 3.2.2.1.3). Al igual que en las operaciones para sacar listas de palabras, los índices cuantitativos de los centros de interés se pueden obtener según las variables. Es decir, se pueden sacar diferentes resultados sobre dichos índices estableciendo distintas condiciones a través de las variables. Sin embargo, la deficiencia del programa anteriormente comentada continúa en estos cálculos: no es posible cruzar las variables relativas al informante. Por ejemplo, en la presente investigación, uno no puede conseguir el promedio de palabras en ningún centro de interés de los varones (opción de la variable “sexo”) del primer curso (opción de la variable “nivel de estudio”). Los propios administradores del programa desconocían esta limitación hasta que contactamos con ellos. Lo esperable es que, si no es posible subsanarla, la web advierta de su existencia, para que los investigadores se planteen alternativas o adecuen sus hipótesis y objetivos a lo que pueden obtener utilizando este instrumento de análisis.

4.4. Las variables extralingüísticas

Son seis las variables extralingüísticas que hemos tenido en cuenta en las encuestas para estratificar a los informantes y analizar la incidencia de cada una. Concretamente, son el nivel de estudio, el sexo, la procedencia, el nivel de inglés, la estancia en países hispanohablantes y el tipo de centro. Vamos a explicar en los siguientes apartados cómo hemos determinado las opciones — o sea, las variantes— de cada variable y dar a conocer la distribución de cada

variante en el conjunto de la muestra tanto de encuestas en ELE como de encuestas en CLM.

Antes de nada, es preciso llamar la atención sobre el hecho de que tuvimos que abandonar la variable “estancia en países hispanohablantes” — aunque recogida en las encuestas—, dada la deficiencia que detectamos en el programa *Dispolex* (véase el apartado anterior). La razón la vamos a ir aclarando a continuación. Es muy común que los departamentos de Filología Hispánica en China —incluidos los tres estudiados en la presente tesis— tengan programas de movilidad y formación para estudiantes. Los alumnos pueden elegir libremente si hacer o no una estancia temporal en una universidad extranjera con la que su universidad tenga algún convenio de intercambio. Normalmente, las universidades de destino son las de algún país hispanohablante⁸⁷, ya que es la lengua española la que interesa a los departamentos de Filología Hispánica. Resulta sumamente tentador averiguar la diferencia —en nuestro caso, la diferencia en la competencia léxica manifestada por la disponibilidad léxica— entre los estudiantes que eligen la movilidad y los que no. Será un análisis de incidencia que nos permitirá revelar el peso de la inmersión lingüística en el aprendizaje del español como lengua extranjera. Por lo tanto, establecimos en nuestras encuestas la variable “estancia en países hispanohablantes” (véase el apartado 4.3.1.2) para investigar más tarde la incidencia al respecto.

Sin embargo, el mecanismo de movilidad en dichos departamentos, por lo general, es enviar a sus alumnos a partir del tercer curso del grado. Esto se debe a que los alumnos de primero y segundo curso, posiblemente, no tienen el nivel de lengua suficiente para poder dominar los estudios en las universidades de destino. Es comprensible porque, como hemos comentado en el apartado 4.1.2, en el primer curso los alumnos empiezan a estudiar el abecedario de la lengua española. En tales circunstancias, se puede ver que

⁸⁷ Algunos departamentos también ofrecen a sus alumnos la opción de hacer la estancia en universidades de destino ubicadas en algunas partes de los Estados Unidos donde el español tiene mucha influencia y es comúnmente hablado.

la movilidad de los alumnos viene plenamente afectada por su nivel de estudio —los cuatro cursos universitarios—. Dicho de otra manera, la variable de estancia no ejerce ninguna influencia en los alumnos del primero y segundo curso y solo tiene sentido de estratificación para los de tercero y cuarto curso. Por eso, se deduce que el análisis de la diferencia léxica dada por la estancia en países hispanohablantes será lógico si se hace solamente en el conjunto de los alumnos de tercero y cuarto. Como el programa *Dispolex* no nos permitía cruzar las variables relativas al informante, no pudimos sacar resultados contrastivos entre los informantes de tercero o cuarto que habían hecho la estancia y los de dichos cursos que no la habían hecho⁸⁸. De todas maneras, lo que pudimos realizar mediante *Dispolex* en lo que se refiere a la variable de estancia fue contrastar los que sí habían hecho la estancia y los que no, sobre la base del conjunto del alumnado. Esto último no es significativo, toda vez que, de ser así, en el grupo de “sí” se calculó sobre los últimos dos cursos y en el grupo de “no” se calculó sobre los cuatro cursos totales.

Por esta consideración, de las seis variables que pudimos extraer de las encuestas, analizamos en esta tesis solamente las cinco siguientes: el nivel de estudio, el sexo, la procedencia, el nivel de inglés y el tipo de centro. A diferencia de la estancia en países hispanohablantes, estas cinco variables son intrínsecas de los informantes, es decir, no vienen afectadas por otro factor de la investigación y, a la hora de sacar resultados, no tuvieron que cruzarse con otra variable. De este modo, con lo que es factible hacer con el programa *Dispolex*, hemos podido estudiar la incidencia que cada una de las cinco variables ejerce en la disponibilidad léxica de los informantes. La variable de la estancia en países hispanohablantes la reservamos para analizarla en el futuro, o bien cuando el *Dispolex* nos lo permita o bien recurriendo a otros programas informáticos.

Adicionalmente, con respecto a la distribución de los informantes

⁸⁸ En el apartado anterior esto ya se ha comentado y, en este caso concreto, no pudimos cruzar las variantes de la variable “nivel de estudio” con las variantes de la variable “estancia en países hispanohablantes”.

correspondientes a las variantes de cada variable, hemos de recordar que la presente investigación no se ha basado en una simetría cuantitativa entre los mismos. Dicho de otra manera, dentro de cada variable, los informantes de la muestra no se distribuyen cuantitativamente por igual según la variante. Por un lado, como se ha comentado anteriormente, a la hora de convocar a los estudiantes a las encuestas, no podíamos prever ni predeterminar a los informantes que coincidían con cada una de las variantes (véase el apartado 4.2.1). De modo que tanto la muestra en ELE como la en CLM fueron posteriormente estratificadas. Por otro lado, no pensamos que era imprescindible que los informantes se distribuyeran simétricamente según la variante en todas las variables, toda vez que la población estadística⁸⁹ en nuestra investigación no se distribuía de esta manera⁹⁰. Estando así las cosas, lo que se ha pretendido en esta tesis en lo que se refiere a la distribución de las variantes ha sido una proporcionalidad parecida entre las variantes de la muestra en ELE y las de la muestra en CLM. Por eso, a continuación, vamos a presentar, sobre todo, la proporcionalidad de las variantes de cada variable en ambas muestras.

4.4.1. Variable “nivel de estudio”

El nivel de estudio de los informantes es la variable más importante de nuestra investigación, porque se imagina que las diferentes fases de estudio constituyen el factor determinante en la disponibilidad léxica de los estudiantes si se enseñan y aprenden unos vocablos en una fase y otros, en cambio, en otra. A diferencia de las investigaciones del *Proyecto Panhispánico*, que estudian a informantes de un nivel de estudio determinado —el nivel preuniversitario, o sea, el último curso del bachillerato—, nosotros tenemos otros objetivos relativos a los aprendientes de español como lengua extranjera.

⁸⁹ La población estadística consiste en el conjunto de individuos del que se obtiene la muestra. Más concretamente, se refiere en esta investigación a todos los estudiantes universitarios de ELE en Tianjin.

⁹⁰ Un ejemplo para que se entienda fácilmente esto es el caso del sexo: las cantidades de chicos y de chicas que estudian español como lengua extranjera en las universidades chinas nunca se presentan de manera simétrica, sino bastante desigual. En otras variables pasa lo mismo.

Uno de nuestros objetivos, como ya se ha indicado en la Introducción a esta tesis, consiste en averiguar las diferencias tanto cuantitativas como cualitativas entre el léxico disponible de los cuatro cursos universitarios en que estudian los alumnos de ELE en Tianjin. A partir de ello, por un lado, se puede observar si existe una evolución diacrónica en la disponibilidad léxica y, en su caso, cómo se manifiesta cuantitativamente; por otro lado, los análisis cualitativos centrados en cada uno de los cursos permiten aportar informaciones sobre los métodos y materiales didácticos correspondientes a cada curso, tales como cuáles son las palabras que se han adquirido por los alumnos en cada fase y qué palabras han de reforzarse en cuanto a su adquisición en una fase de estudio concreta. De todas maneras, el nivel de estudio, por lo general, es el indicador que revela la competencia léxica de los individuos que forman parte de cualquier conjunto de estudiantes de lengua extranjera, siempre que el conjunto se pueda estratificar de acuerdo con él. Por todo esto, incluimos en la presente tesis la variable “nivel de estudio” con el objetivo de estratificar a los estudiantes de ELE en Tianjin tal y como el sistema educativo los secciona según las fases de aprendizaje.

Dentro de esta variable distinguimos cuatro variantes, las cuales se corresponden con los cuatro cursos universitarios en que estaban estudiando los informantes en el momento de la encuesta. La Tabla 4.4 muestra el número de informantes que tuvimos en los tipos de encuestas (la en ELE y la en CLM) en función de cada curso universitario.

Nivel de estudio	ELE	CLM
Primero	105	68
Segundo	84	52
Tercero	51	20
Cuarto	48	32
Total	288	172

Tabla 4.4. Número de informantes de la muestra en ELE y en CLM según el nivel de estudio

Como se ha mencionado en los apartados 4.2.1 y 4.2.2, debido a los

imponderables, los informantes de cada curso fueron convocados a las encuestas de igual modo, pero respondieron de manera distinta. Por lo tanto, tal y como se presenta en la Tabla 4.4, no pudimos obtener unas estratificaciones de porción igual entre los cuatro cursos en ninguno de los dos tipos de muestra —dadas la yuxtaposición de los números 105, 84, 51 y 48 en la muestra de ELE y la de las cifras 68, 52, 20 y 32 en la muestra de CLM—. De igual modo, la Tabla 4.4 también informa de que no tuvimos estratificaciones de porción igual entre las muestras de ELE y de CLM en ningún curso —teniendo en cuenta la contraposición de 105 y 68 en el primer curso; la de 84 y 52, en el segundo; la de 51 y 20, en el tercero, y la de 48 y 32, en el cuarto—.

Sin embargo, si analizamos, respectivamente, el porcentaje que ocupa la porción de cada curso en el total de la muestra de ELE y de CLM —lo que exponen los Gráficos 4.2 y 4.3—, podemos observar una proporcionalidad más o menos parecida entre las dos muestras.

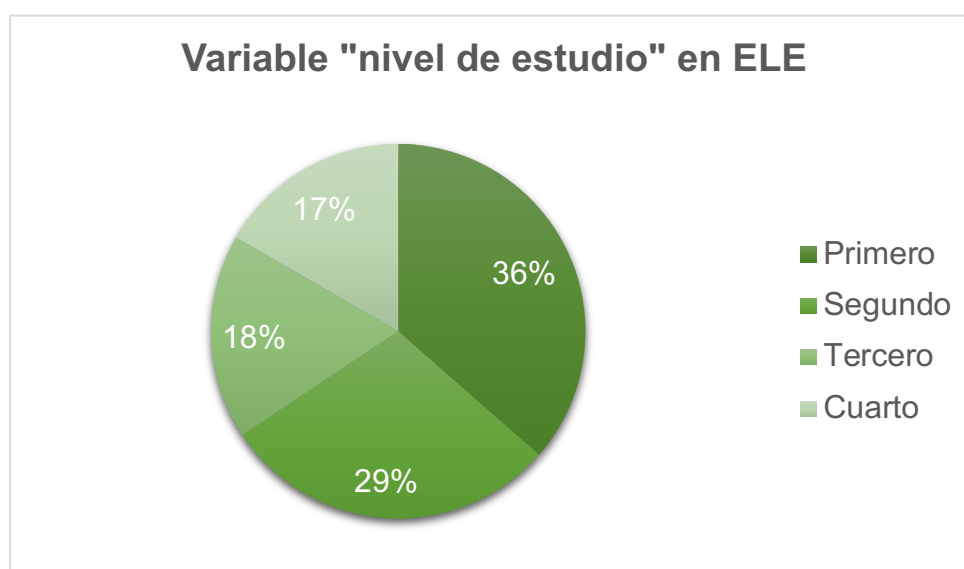


Gráfico 4.2. Distribución de los informantes en ELE según el nivel de estudio

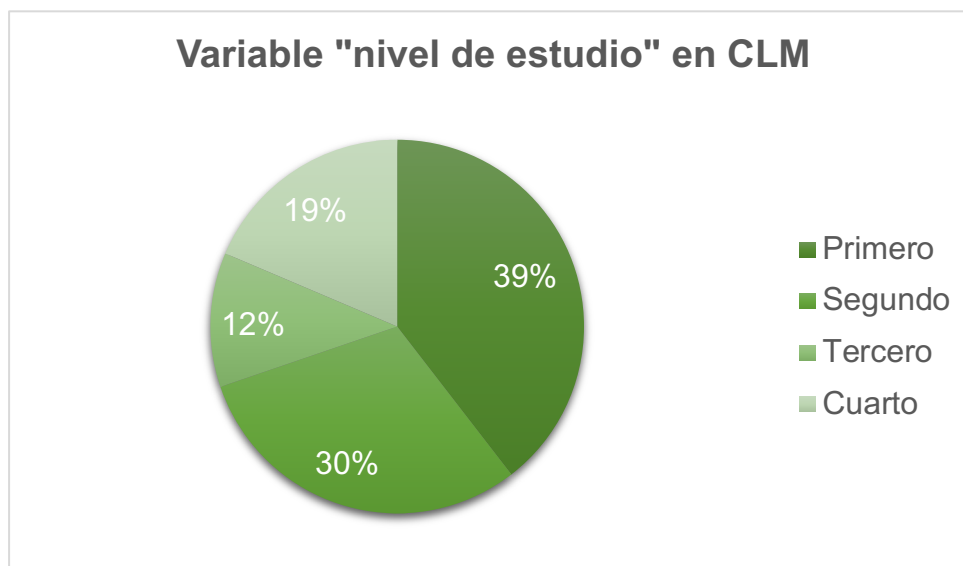


Gráfico 4.3. Distribución de los informantes en CLM según el nivel de estudio

Los porcentajes en ambas muestras presentan una mayoría contundente de los primeros dos cursos. De modo similar, en las dos muestras, los informantes de primer curso suponen un poco más de un tercio de la totalidad —36% en la muestra de ELE y 39% en la de CLM—; mientras que los de segundo cubren apenas un tercio —29% en la muestra de ELE y 30% en la de CLM—. Así mismo, la suma de los porcentajes de tercer curso y de cuarto en ambas muestras representa alrededor del tercio restante —35% en la muestra de ELE y 31% en la de CLM—, aunque, dentro de este bloque, se nota un menor porcentaje de informantes de tercero en la muestra de CLM —un 12%—. En una situación como esta, las distribuciones más o menos parecidas, observadas a partir de la proporcionalidad que presentan la cantidad de informantes provenientes de los cuatro cursos universitarios, hacen viable un estudio de comparación entre las dos muestras con el enfoque en el nivel de estudio de los estudiantes.

4.4.2. Variable “sexo”

El estudio sobre la variable sexo es ya una tradición en las investigaciones de disponibilidad léxica. En los estudios existentes se ha demostrado que el sexo es un factor que influye en el conocimiento y el dominio del léxico.

Tenemos en cuenta esta variable con el objetivo de averiguar si existe una superioridad en la disponibilidad léxica de un género sobre el otro, a pesar de que en muchas de las investigaciones ya se ha constatado la ventaja de las chicas sobre los chicos. Vamos a ver si esta tendencia se repite y si dicha suposición se comprueba en los alumnos de ELE en Tianjin. La Tabla 4.5 presenta el número de informantes que tuvimos en las muestras de ELE y de CLM en función de los dos géneros.

Sexo	ELE	CLM
Varón	59	24
Mujer	229	148
Total	288	172

Tabla 4.5. Número de informantes de la muestra en ELE y en CLM según el sexo

Como en la variable anterior, sigue habiendo una cantidad desigual que se da entre varones y mujeres tanto en la muestra de ELE —59 chicos frente a 229 chicas— como en la de CLM —24 chicos frente a 148 chicas—. El perfil de nuestros informantes se caracteriza por una mayor presencia de las mujeres que de los hombres. No obstante, ello no es raro en los estudios de disponibilidad léxica y, sobre todo, en los aplicados a informantes de ELE. Esto último se debe a que, en cualquier población, siempre hay más chicas que chicos que optan por estudiar una lengua extranjera. Se consideran estudios propios de mujeres.

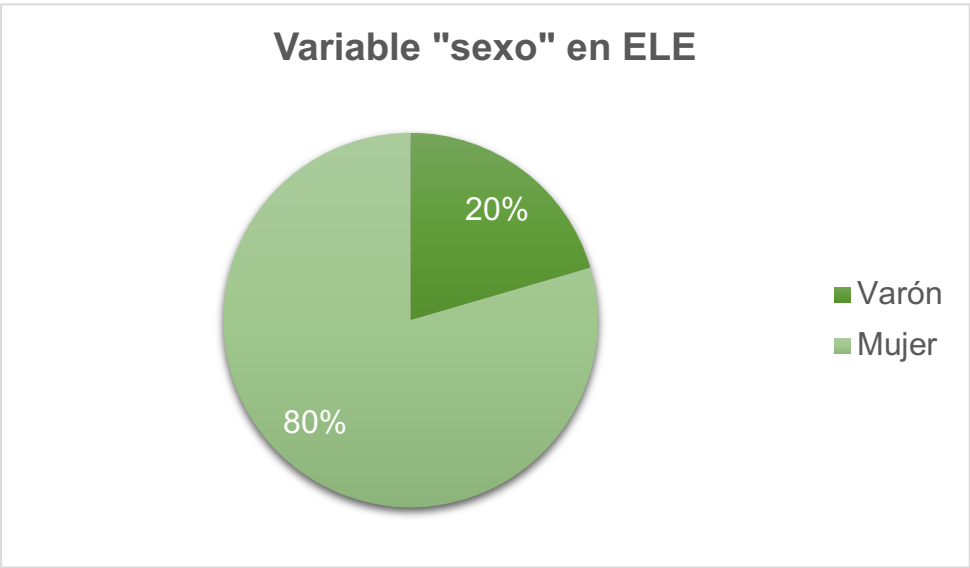


Gráfico 4.4. Distribución de los informantes de la muestra en ELE según el sexo

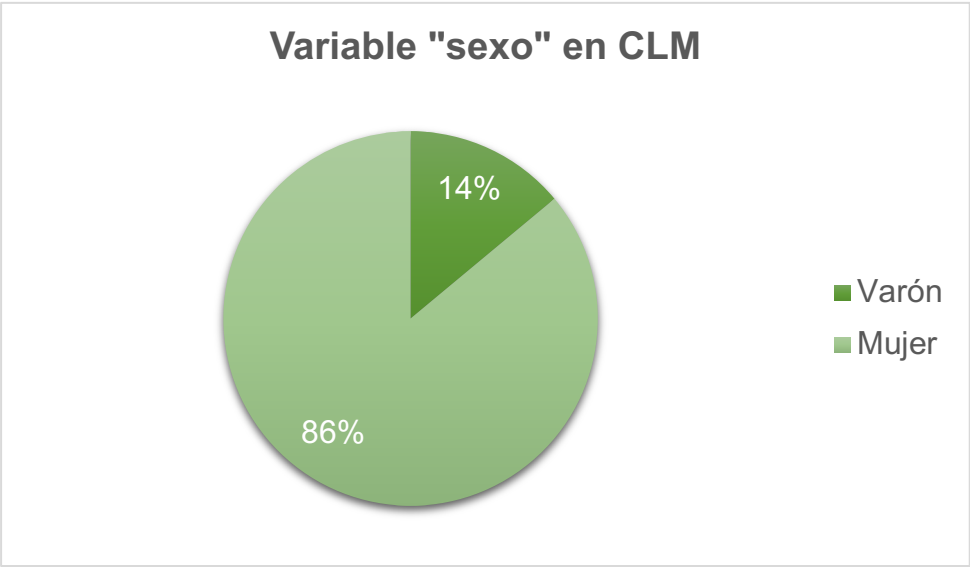


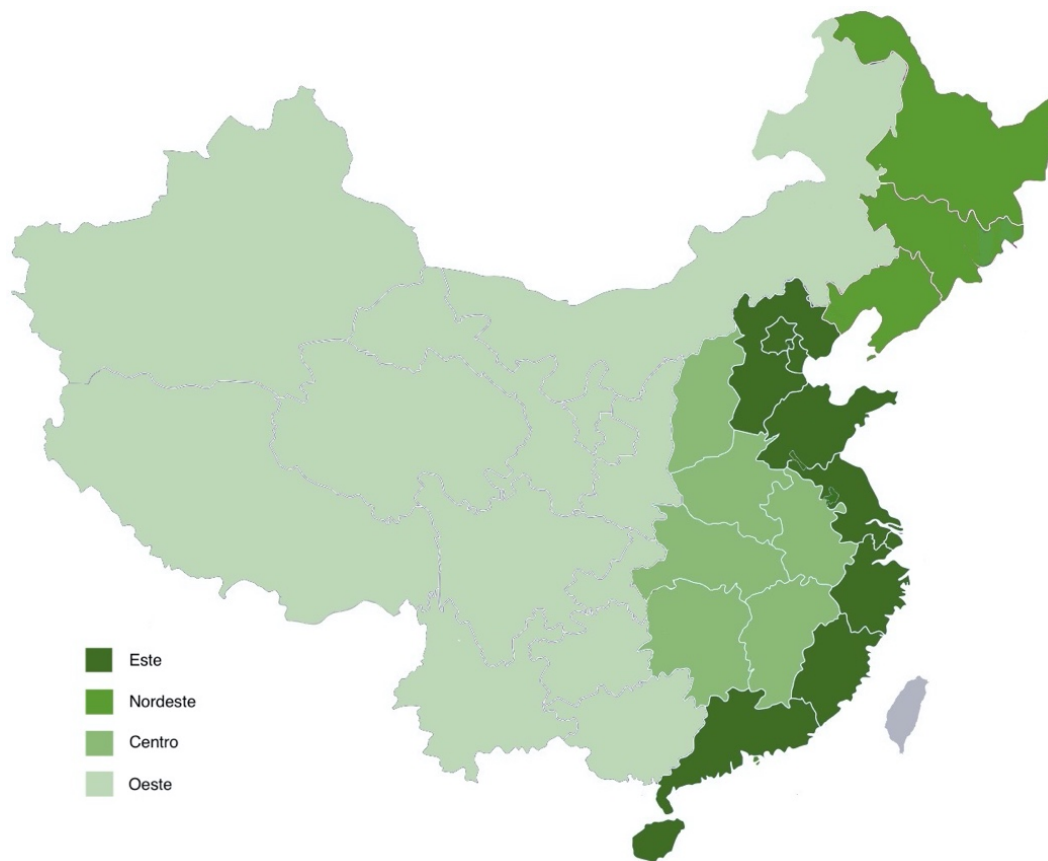
Gráfico 4.5. Distribución de los informantes de la muestra en CLM según el sexo

Según mi experiencia como profesor de ELE en Tianjin, el número de alumnos varones en un grupo de 25 estudiantes oscila entre 3 y 5, o, lo que viene a ser lo mismo, los alumnos varones ocupan entre el 12% y el 20% de la totalidad del alumnado. Los Gráficos 4.4 y 4.5 muestran los porcentajes que ocupan los varones y las mujeres en las dos muestras. Se puede ver en estos gráficos que los chicos suponen un 20% en la muestra de ELE y un 14% en la muestra de CLM. Las porciones de los alumnos varones en ambas muestras

coinciden con nuestra experiencia empírica, lo cual justifica la representatividad de nuestras muestras en lo que se refiere a la distribución de los dos géneros. Por otra parte, la contraposición entre el porcentaje de varones y el de mujeres en las dos muestras —20% frente a 80% en la muestra de ELE y 14% frente a 86% en la de CLM— describe una proporcionalidad parecida. Por lo tanto, tenemos el fundamento sobre el que se apoya un estudio de comparación entre las dos muestras con respecto a la variable del sexo; a pesar de la diferencia que existe entre el número de varones en ELE y en CLM —59 frente a 24— y entre el número de mujeres de la misma índole —229 frente a 148— (véase la Tabla 4.5).

4.4.3. Variable “procedencia”

La procedencia de los estudiantes es una variable interesante, puesto que nos gustaría averiguar si el lugar de origen de los alumnos ejerce alguna influencia en su disponibilidad léxica. Más concretamente, si no podemos poner en análisis todos los lugares posibles, pretendemos, por lo menos, conocer si los estudiantes que proceden de unas zonas que comparten rasgos comunes se diferencian en la disponibilidad léxica de otros procedentes de otras zonas caracterizadas de otro modo. Dado esto, con respecto a la variable *procedencia*, los informantes en nuestra investigación se dividen en estratos de acuerdo con la ubicación de su provincia de origen. Como no es viable ni necesario contrastar todas las divisiones administrativas de nivel provincial en China —las cuales suman 34 tal y como se ha comentado en el apartado 4.1.1—, la medida que tomamos en este estudio es contrastar las cuatro zonas del país bien distintas según la geografía económica. El Mapa 4.2 exhibe las cuatro zonas en cuestión con diferentes tonos del color verde, las cuales son la zona del este, la del nordeste, la del centro y la del oeste.



Mapa 4.2. Zonas de China según la geografía económica

Esta división de cuatro zonas según la geografía económica es conocida a partir del año 2002. Fue planteada primeramente por el Consejo de Estado⁹¹ chino con la finalidad de reflejar los distintos estados de desarrollo socioeconómico de los territorios. Una vez que entró en funcionamiento, es una de las divisiones socioeconómicas según las que el Instituto Nacional de Estadística de China lleva a cabo sus trabajos. En el sitio web de este instituto, se explica cómo las zonas socioeconómicas se han delimitado a partir de las provincias y qué provincias o regiones de nivel provincial pertenecen a cada una de las cuatro zonas⁹². Tradicionalmente, se cree que la zona del este es,

⁹¹ El Consejo de Estado está encabezado por el primer ministro. Es el máximo órgano del poder ejecutivo del país y ejerce, entre otros, los poderes de presentación de proyectos de ley, dirección ejecutiva, administración económica, administración de las relaciones exteriores y administración de la sociedad.

⁹² La zona del Este incluye 7 provincias (Hebei, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong y Hainan) y 3 municipios bajo jurisdicción central (Beijing, Tianjin y Shanghai). La del Nordeste cuenta con 3 provincias (Liaoning, Jilin y Heilongjiang). La del Centro abarca 6 provincias (Shanxi, Anhui, Jiangxi, Henan, Hubei y Hunan). La zona del Oeste comprende 6 provincias (Sichuan, Guizhou, Yunnan, Shaanxi, Gansu y Qinghai), 1

en todos los sentidos, la más desarrollada económicamente y, al contrario, la zona del oeste es la menos desarrollada. Por lo tanto, la del nordeste y la del centro son zonas de nivel de desarrollo intermedio y se cree que ellas dos son más o menos iguales en cuanto al grado de desarrollo. Con el fin de comprobar esta creencia tradicional, recurrimos a las últimas estadísticas a que teníamos acceso y tuvimos las cifras de la Tabla 4.6 que muestran el nivel de desarrollo de las diferentes regiones por medio del PIB per cápita⁹³.

PIB per cápita por provincias 2018 (yuanes por habitante)							
Este		Nordeste		Centro		Oeste	
Beijing	140211	Liaoning	58008	Shanxi	45328	Mongolia Interior	68302
Tianjin	120711	Jilin	55611	Anhui	47712	Guangxi	41489
Hebei	47772	Heilongjiang	43274	Jiangxi	47434	Chongqing	65933
Shanghai	134982			Henan	50152	Sichuan	48883
Jiangsu	115168			Hubei	66616	Guizhou	41244
Zhejiang	98643			Hunan	52949	Yunan	37136
Fujian	91197					Tíbet	43398
Shandong	76267					Shaanxi	63477
Guangdong	86412					Gansu	31336
Hainan	51955					Qinghai	47689
						Ningxia	54094
						Xinjiang	49475
Promedio de PIB per cápita por zonas 2018 (yuanes por habitante)							
Este		Nordeste		Centro		Oeste	
96332		52298		51699		49371	

Tabla 4.6. PIB per cápita por provincias y promedio de PIB per cápita de las cuatro zonas socioeconómicas chinas (año 2018)⁹⁴

municipio bajo jurisdicción central (Chongqing) y 5 regiones autónomas (Mongolia Interior, Guangxi, Tíbet, Ningxia y Xinjiang). Estas informaciones sobre la delimitación detallada de las cuatro zonas socioeconómicas se pueden consultar en la página web http://www.stats.gov.cn/tjzs/cjwtd/201308/t20130829_74318.html.

⁹³ Las estadísticas consultadas son del año 2018 y fueron las más recientes entre las publicadas hasta el año 2020 por el Instituto Nacional de Estadística de China en su sitio web <http://data.stats.gov.cn>.

⁹⁴ Fuente: Elaboración propia a partir de datos del Instituto Nacional de Estadística de China en la página web <http://data.stats.gov.cn>.

El PIB per cápita mide la riqueza material disponible de un habitante medio. Como indicador de productividad, es usado para revelar el desarrollo económico, las condiciones socioeconómicas e, incluso, el bienestar social de una región. Tal y como se presentan los datos en la Tabla 4.6, las estadísticas del 2018 evidencian, en cuanto a la renta económica, la superioridad de la zona del este sobre las otras tres zonas, ostentando una media de 96332 yuanes⁹⁵ por habitante, casi el doble del promedio de la zona del oeste —49371 yuanes—. Se deja claro que la zona del oeste carece de ventaja económica frente a las demás zonas. Asimismo, podemos percatarnos de que la zona del nordeste y la del centro tienen un nivel de desarrollo mediano. Estas dos disponen de fuerzas económicas equivalentes, a pesar de que la primera supera levemente a la segunda en la cifra. Comprobamos, de esta manera, el dicho tradicional⁹⁶ sobre los distintos grados de desarrollo económico de las cuatro zonas estudiadas, hasta el punto de ordenarlas en secuencia de la más desarrollada económicamente a la menos desarrollada como lo siguiente: este, nordeste, centro y oeste. Como solicitamos a los informantes en las encuestas su provincia de origen, pudimos corresponder a cada uno con una de las cuatro zonas socioeconómicas en cuestión. De ahí podremos averiguar si los informantes de distintas zonas socioeconómicas muestran diferencias significativas en la disponibilidad léxica y, en su caso, si la disponibilidad léxica de los estudiantes tiene correlación con el nivel de desarrollo de su zona de procedencia.

A continuación, la Tabla 4.7 enseña el número de informantes que tuvimos en las muestras de ELE y de CLM según las cuatro zonas de procedencia.

⁹⁵ El yuan es la unidad básica de la moneda corriente de la República Popular China. Un yuan equivale, a día de 1 de enero de 2021, a 0.13 €.

⁹⁶ Nos referimos al dicho de que la zona del este es, en todos los sentidos, la más desarrollada económicamente y, al contrario, la zona del oeste es la menos desarrollada.

Procedencia	ELE	CLM
Este	132	66
Nordeste	8	7
Centro	74	48
Oeste	74	51
Total	288	172

Tabla 4.7. Número de informantes de la muestra en ELE y en CLM según la procedencia

Otra vez estamos ante unos estratos de porción dispar entre informantes de las cuatro zonas: en la muestra de ELE se enfrentan, por zona de procedencia, los números 132, 8, 74 y 74 y, en la muestra de CLM, 66, 7, 48 y 51. Por otro lado, entre las dos muestras, tienen su origen en el Este 132 alumnos de la muestra de ELE y 66 de la de CLM; de igual modo, se oponen 8 y 7, 74 y 48 y 74 y 51 en las otras tres procedencias, respectivamente. Al igual que en los casos anteriores, conseguimos hacer balance de esto con la proporcionalidad que implican las procedencias, tal y como indican los Gráficos 4.6 y 4.7.

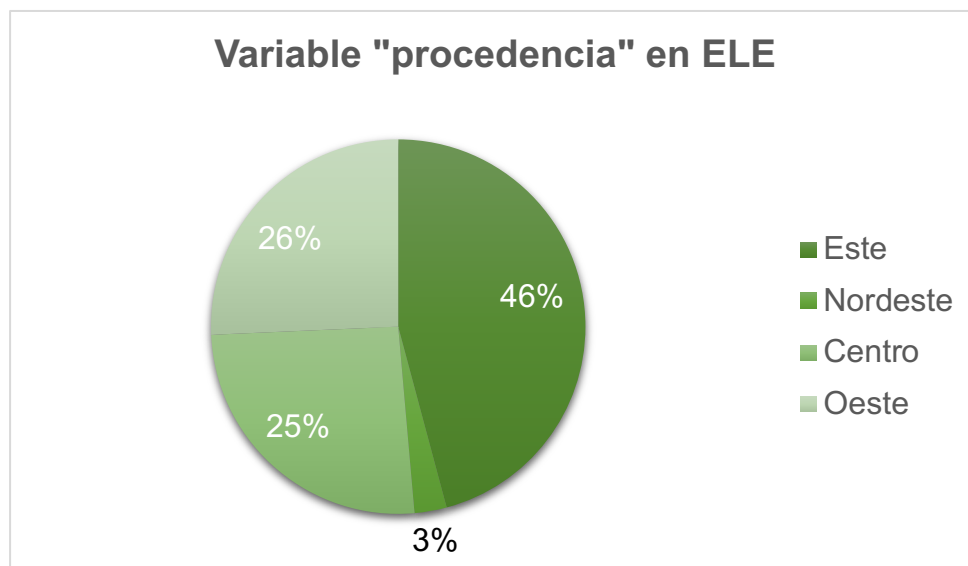


Gráfico 4.6. Distribución de los informantes en ELE según la procedencia

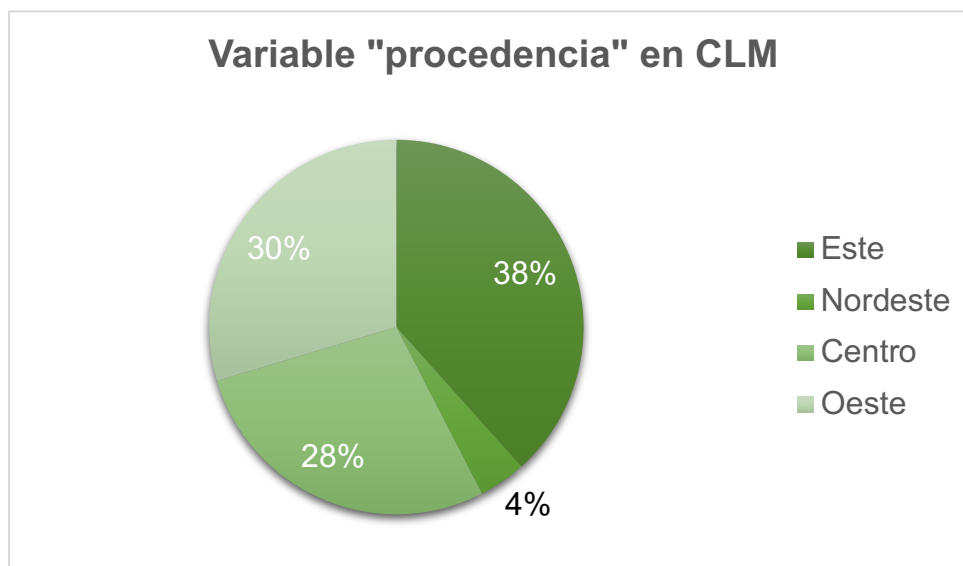


Gráfico 4.7. Distribución de los informantes en CLM según la procedencia

Se puede observar que, en ambas muestras, hay más informantes del Este que de cualquier otra procedencia —46% y 38% en ELE y en CLM, respectivamente—. Los alumnos del Centro y del Oeste ocupan más o menos el mismo peso en cuanto al porcentaje —25% frente a 26% en ELE y 28% frente a 30% en CLM— y, mientras tanto, la suma de estas dos procedencias constituye la gran mayoría en las muestras—51% y 58% respectivamente—. De modo similar, los estudiantes del Nordeste se manifiestan como una minoría: solo un 3% en la muestra de ELE y 4% en la otra. La jerarquía, en lo que se refiere a la composición porcentual de las cuatro procedencias, es la misma: hay más sujetos del Este que del Oeste, más del Oeste que del Centro y más del Centro que del Nordeste. Esta proporcionalidad más o menos análoga sirve de fundamento para llevar a cabo las comparaciones entre las dos muestras con respecto a la procedencia de los informantes.

Por otra parte, la disparidad en la cantidad de alumnos correspondientes a cada procedencia, en ninguna ocasión, constituye un obstáculo que ponga en duda nuestro estudio de cotejo, toda vez que en la población estadística —o sea, en el conjunto de alumnos de ELE en Tianjin— las cuatro procedencias jamás se distribuyen con proporción igual en cada una. No vamos a analizar la composición porcentual de cada procedencia en dicha población, porque los

profesores de las tres universidades estudiadas nos informaron de que iba cambiando todos los años en función de la planificación de admisión de estudiantes del año correspondiente. Esta planificación establece, según las situaciones de cada año académico, las normas de admisión de estudiantes en cada provincia y cuántos estudiantes admite de cada una. Por lo tanto, en ambas muestras, dejamos al muestreo al azar la composición de informantes procedentes de las cuatro zonas socioeconómicas.

4.4.4. Variable “nivel de inglés”

El nivel de inglés es otra variable que tenemos en cuenta en la investigación. La idea de incluir una variable relativa a la lengua inglesa fue tomada en consideración cuando diseñábamos la investigación y las encuestas, por razones que explicamos a continuación. En primer lugar, entre las investigaciones existentes de disponibilidad léxica, son muchos los intentos por estudiar los anglicismos aparecidos en las listas del léxico disponible en español, los cuales implican la influencia del inglés en la adquisición de la lengua española. Como ejemplos, tenemos estudios de este tipo de López Chávez (1991), Benítez Pérez (1993), Orlando Alba (1995 y 1999), López Morales (1999a) y Casanova Ávalos (2008) (véase el apartado 3.2.2).

En segundo lugar, el inglés es un idioma que los estudiantes chinos aprenden como lengua extranjera desde la educación primaria, cuya enseñanza es obligatoria en el sistema educativo chino. Esto quiere decir que el inglés es la primera lengua extranjera de nuestros informantes, hasta que el peso de la enseñanza y aprendizaje del español va aumentando y sobrepasando al inglés en la universidad entre los estudiantes de Filología Hispánica. Por eso, los alumnos, casi sin excepciones, comienzan el aprendizaje del español sobre la base de su conocimiento del inglés. Están muy dependientes, sobre todo, en fases iniciales del aprendizaje, de la morfosintaxis del inglés en la expresión escrita, toda vez que, en comparación con su lengua materna, el idioma inglés es mucho menos lejano del español y el dominio del inglés puede facilitar la adquisición del español. Por ejemplo, los

alumnos pueden encontrar en la lengua inglesa estructuras sintácticas afines o raíces de palabras iguales que las españolas. El hecho de que los estudiantes dominen bien o mal el inglés significa mucho. En consecuencia, en China, por lo general, se estima que el inglés puede influir positivamente en la adquisición del español. Son estas razones por las que indagamos en esta investigación el nivel de inglés de los informantes, con el objetivo de observar si existen correlaciones entre esta variable y la disponibilidad léxica y, también, para estudiar, en futuras investigaciones, correlaciones entre esta variable y los anglicismos en nuestras listas de palabras.

El modo de inquirir, en los cuestionarios, qué nivel de inglés tenía cada informante ya lo hemos aclarado en el apartado 4.3.1.2. En resumen, les pedimos, o bien marcar una de las siguientes cinco opciones: 1) no saber inglés, 2) haber estudiado inglés en el bachillerato, 3) estar estudiando inglés en la universidad, 4) tener aprobado el CET⁹⁷ nivel 4 y 5) tener aprobado el CET nivel 6, o bien especificar qué otros certificados de inglés tenían en ese momento. A la opción o especificación que había proporcionado cada informante le asignamos una valoración que manifestaba su nivel de lengua inglesa y que permitía la estratificación de la muestra. Las valoraciones se hicieron de acuerdo con cuatro niveles de inglés que se presentaban en forma de una escala jerárquica. Estos eran, del más bajo al más alto, el nivel básico, el intermedio, el avanzado y el superior. Una vez que tengamos los estratos de informantes en virtud de dichos cuatro niveles progresivos, pondremos en marcha el análisis sobre si en la disponibilidad léxica de los estudiantes consta alguna progresividad en función de su nivel de inglés. Y, en el caso de que no se manifieste dicha progresividad, podemos por lo menos observar si hay alguna diferencia, con respecto a la disponibilidad, del estrato de un nivel con el de otro. A continuación, explicamos cómo valoramos las informaciones que recogimos de las encuestas en relación con el nivel de inglés.

En primer lugar, no hallamos en los cuestionarios ninguna respuesta que

⁹⁷ Consulte el apartado 4.3.1.2 para conocer de qué trata la prueba CET.

afirmara que el informante no sabía inglés. Es decir, todos los alumnos que formaban parte de nuestras muestras tenían conocimiento de esta lengua. En segundo lugar, después de revisar todos los cuestionarios recabados, nos dimos cuenta de que todos nuestros informantes disponían de un nivel de este idioma superior al del bachillerato: o bien estaban en el nivel de inglés de los cursos universitarios, o bien tenían aprobado algún test de dicha lengua que suponía más allá del nivel de inglés del bachillerato. Por lo tanto, el nivel básico de inglés del que hablamos en esta investigación se refiere al nivel de este idioma correspondiente a la enseñanza básica de la universidad.

Estando así las cosas, estratificamos, de manera general, a los informantes de las dos muestras de la manera siguiente: a los que indicaban que estaban estudiando el inglés en la universidad y no tenían ningún certificado de esta lengua, les asignamos el nivel básico; a los que tenían aprobado el CET nivel 4, consideramos que tenían el nivel intermedio de inglés; y, asimismo, a los que tenían aprobado el CET nivel 6, les atribuimos el nivel avanzado. El nivel superior de inglés lo equiparamos con el nivel que tenían aquellos que habían especificado los siguientes casos particulares: 1) los que afirmaron que consiguieron un puntaje de 7 o más en el examen IELTS⁹⁸; 2) los que hicieron constar que obtuvieron 100 puntos o más en el examen TOEFL⁹⁹; y 3) los que declararon que habían tomado el examen TEM¹⁰⁰, a pesar de que no dieron a conocer sus notas.

Teniendo en cuenta que el examen IELTS se califica según una escala de

⁹⁸ IELTS es la sigla en inglés de *International English Language Testing System*. Es una prueba estandarizada internacional de dominio de la lengua inglesa para hablantes no nativos de dicho idioma. Es gestionado conjuntamente por el British Council, el IDP: IELTS Australia y el Cambridge Assessment English de la Universidad de Cambridge. Se trata de una de las dos principales pruebas del inglés en el mundo para educación superior; la otra es el TOEFL (véase la siguiente nota a pie de página).

⁹⁹ TOEFL es la sigla en inglés de *Test of English as a Foreign Language*. Al igual que el IELTS, el TOEFL es una prueba estandarizada internacional para medir la capacidad del idioma inglés de los hablantes no nativos que desean inscribirse en universidades de habla inglesa. La diseña y administra el Educational Testing Service de los Estados Unidos.

¹⁰⁰ TEM es la sigla en inglés de *Test for English Majors*. Se trata de una prueba nacional en China organizada por el Ministerio de Educación y orientada a estudiantes universitarios con especialidad en inglés para medir su dominio de esta lengua.

0 a 9, la puntuación de 7 o más supone un muy buen dominio de la lengua inglesa. Lo mismo se deduce en el caso de 100 puntos o más en el examen TOEFL, puesto que este se califica en una escala de 0 a 120. Dado esto, no equiparamos estos puntajes con el nivel superior de inglés en nuestra investigación. Por otra parte, el examen TEM es una prueba competente en el ámbito nacional de China, especialmente destinada a estudiantes universitarios de la especialidad de lengua inglesa. Por eso, en teoría, no se da acceso a los alumnos de otras especialidades que no sea la lengua inglesa para que tomen dicho examen. No obstante, en una de las tres universidades estudiadas en esta tesis, el programa de enseñanza establece la formación de los estudiantes orientada al dominio del español como especialidad y del inglés como subespecialidad. Por lo tanto, sus estudiantes de Filología Hispánica sí tienen acceso al examen TEM, pero con condiciones: tienen que sacar, como mínimo, 600 puntos en la prueba CET anteriormente comentada —un puntaje muy alto en una escala de 0 a 710 puntos—. Al ser así, se imagina que los informantes que afirmaron que habían tomado parte en el examen TEM disponían de muy buen nivel de inglés. No resultó fácil determinar el nivel de inglés de nuestros informantes, porque no había una pauta única que permitiera la forma de establecerlo, como hemos comentado en el apartado 4.3.1.2. De todas maneras, pudimos tener cuatro estratos en relación con el nivel de inglés de los informantes según el modo que hemos aclarado arriba.

Una vez aquí, vamos a ver cómo se han constituido las dos muestras con respecto a la variable del nivel de inglés. La Tabla 4.8 exhibe el número de informantes que tuvimos de cada uno de los cuatro niveles en la muestra de ELE y de CLM. Al igual que con las otras variables, no pudimos obtener números de informantes equivalentes entre un estrato y otro, ni entre una muestra y la otra. Las simples cifras no presentan, al parecer, tendencia o característica alguna sobre la constitución de las muestras, por lo que vamos a analizar la composición porcentual de cada muestra.

Nivel de inglés	ELE	CLM
Básico	98	64
Intermedio	94	49
Avanzado	88	57
Superior	8	2
Total	288	172

Tabla 4.8. Número de informantes en ELE y en CLM según el nivel de inglés

Pero antes de esto, hemos de llamar la atención sobre la poca cantidad de informantes de nivel de inglés superior en ambas muestras —8 en la de ELE y, aún más reducida, 2, en la de CLM—. Pensamos que es razonable, dado que el estrato de nivel máximo de cualquier clasificación en cualquier población supone una parte limitada.

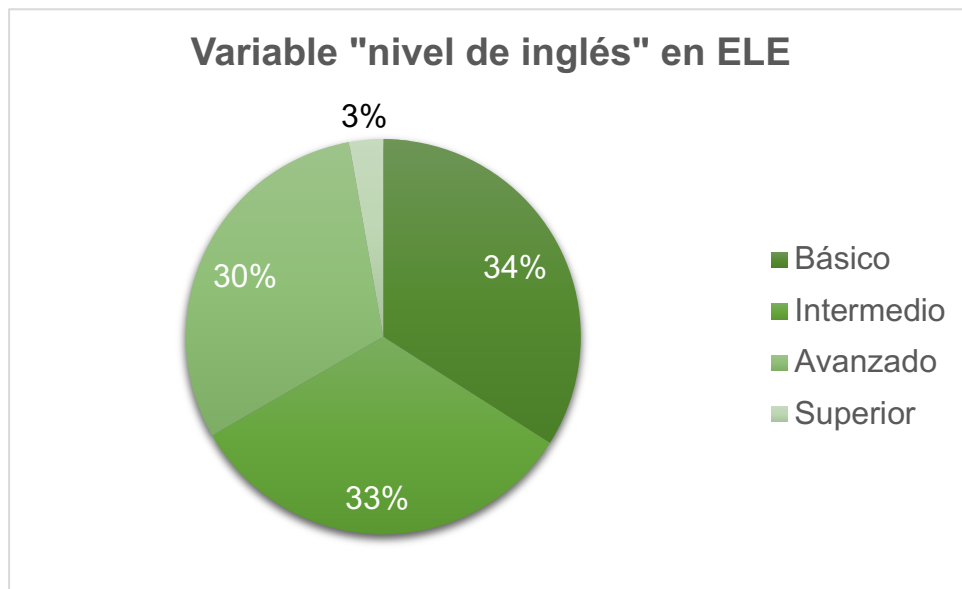


Gráfico 4.8. Distribución de los informantes de la muestra en ELE según el nivel de inglés

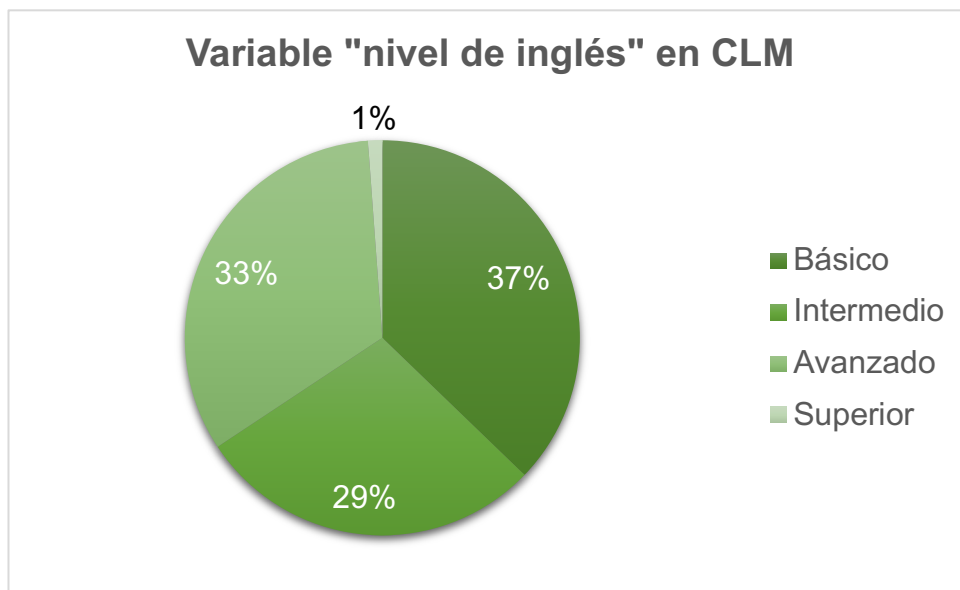


Gráfico 4.9. Distribución de los informantes de la muestra en CLM según el nivel de inglés

Al observar los Gráficos 4.8 y 4.9, pasamos al análisis sobre los porcentajes que ocupan los cuatro estratos en cuestión. En la muestra de ELE, salvo lo que supone el estrato del nivel superior —un reducido 3%—, se puede ver que los demás niveles implican, unánimemente, más o menos un tercio de los informantes de la muestra: el nivel básico reúne un 34% del alumnado, frente a un 33% del nivel intermedio y un 30% del nivel avanzado. De modo que hemos conseguido unas proporciones casi iguales en estos tres estratos excepto el de nivel superior. Si echamos una mirada a la muestra de CLM, podemos encontrar una situación parecida. Al exceptuar la limitada cuota del estrato de nivel superior de inglés —solo un 1%, en este caso—, descubrimos que los otros tres porcentajes fluctúan también alrededor del 30% del conjunto; aunque el estrato del nivel básico se aleja un poco más de esta cifra y se registra un 37%, en comparación con un 29% y un 33% que reflejan, respectivamente, el nivel intermedio y el avanzado. De esta manera, la comparabilidad de las dos muestras se basa en la proporcionalidad parecida entre ambas: bastante reducida proporción del nivel superior y alrededor de un tercio de informantes de cada uno de los tres niveles restantes.

4.4.5. Variable “tipo de centro”

El tipo de centro es una variable que las investigaciones hispánicas introdujeron en los estudios de disponibilidad léxica, con el objetivo de buscar divergencias en el léxico de los hablantes según pertenezcan a centros públicos o privados. Las investigaciones previas en el mundo hispánico han comprobado que esta variable es una de las que mejor pueden explicar las diferencias en el léxico disponible de sus informantes, es decir, consiste en una variable altamente influyente. A este respecto, los estudios del *Proyecto Panhispánico* coinciden en “destacar la importancia del factor ‘tipo de centro’ para dar cuenta de las diferencias de disponibilidad que se observan entre los sujetos encuestados” (Samper Padilla & Samper Hernández, 2006: 76-77). Por consiguiente, consideramos oportuno incluir esta variable en nuestra investigación, persiguiendo los mismos objetivos que los estudios previos. Además, también pretendemos, en esta tesis, averiguar cómo se manifiestan las diferencias tanto cuantitativas como cualitativas en nuestros informantes con respecto al tipo de centro y ver si el caso de Tianjin coincide con las investigaciones del *Proyecto Panhispánico* —puesto que casi todos los trabajos del proyecto apuntan a un promedio superior de los estudiantes de centros privados al de los de centros públicos—.

La distinción entre el carácter privado y público de las universidades chinas no es exactamente la misma que se da entre un centro privado y otro público de las investigaciones del *Proyecto Panhispánico*, lo cual ya hemos clarificado en el apartado 4.2. En síntesis, las universidades públicas referidas en esta tesis son tal y como debe ser un centro público en general; pero, por el contrario, la universidad privada de la que hablamos en esta investigación es una de aquellas que tienen que subordinarse a una universidad pública para servirse de los recursos académicos de esta, mientras que gestionan la administración de ellas mismas con fondos propios y privados. En principio, se supone que esta diferencia con los estudios anteriores, en lo que se refiere al perfil del centro privado, puede traer resultados en la disponibilidad léxica dispares de las conclusiones previas unánimes sobre la superioridad de un tipo

de centro sobre otro.

La Tabla 4.9 expone, en las dos muestras paralelas, cuántos informantes forman parte de cada uno de los estratos según sea público o privado el centro. La cuestión anteriormente tratada acerca de la falta de equivalencia cuantitativa entre los estratos y entre las muestras continúa con la variable del tipo de centro, toda vez que se observa en la tabla contraposiciones de unos pares de cifras desiguales. Por ejemplo, en el estrato de tipo de centro público, se enfrentan 201 alumnos en la muestra de ELE a solamente 126 en la de CLM. Esto, al igual que en los casos de las otras variables, lo justificamos con los porcentajes de cada composición.

Tipo de centro	ELE	CLM
Público	201	126
Privado	87	46
Total	288	172

Tabla 4.9. Número de informantes en ELE y en CLM según el tipo de centro

Tal y como revela el Gráfico 4.10, los informantes de centro público ocupan un 70% del conjunto de la muestra de ELE, dejando así el 30% restante representando a los alumnos de centro privado.

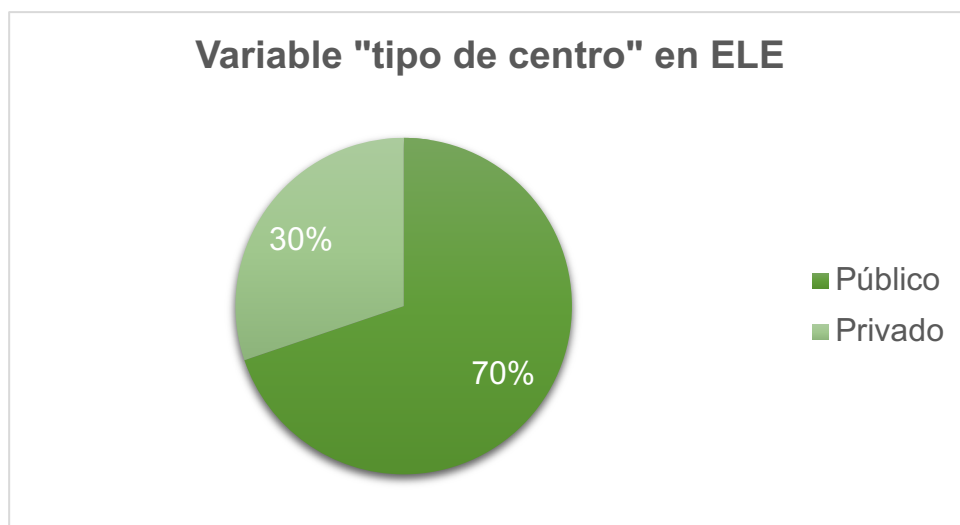


Gráfico 4.10. Distribución de los informantes en ELE según el tipo de centro

Si comparamos el Gráfico 4.11 con el 4.10, se puede ver una composición de proporcionalidad parecida, es decir, la muestra de CLM se compone, más o menos, de la misma manera que la muestra de ELE en cuanto a los porcentajes que suponen los diferentes estratos.

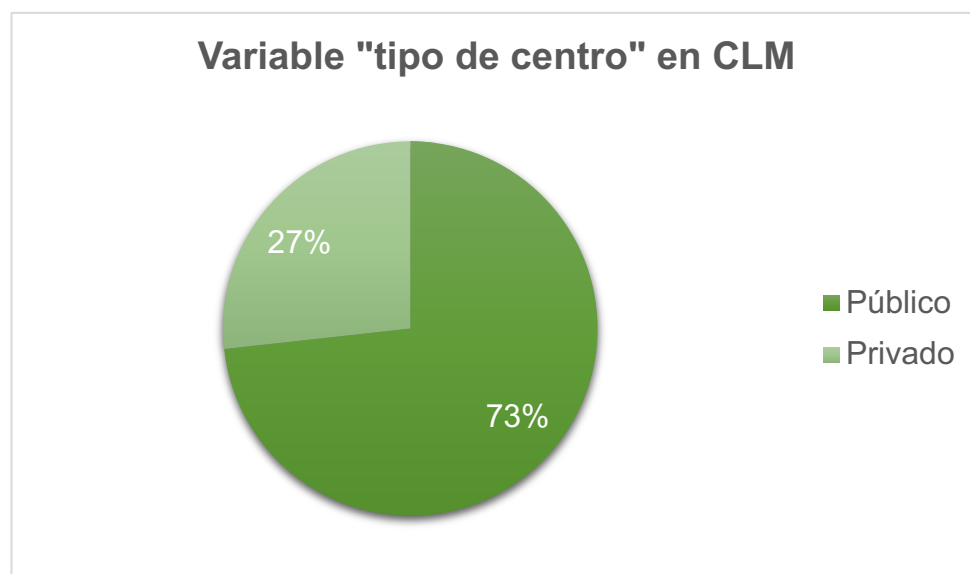


Gráfico 4.11. Distribución de los informantes en CLM según el tipo de centro

Concretamente, en la muestra de CLM, un 73% de los estudiantes proceden de un centro público y el 27% restante, de un centro privado —dos cifras que apenas se alejan del 70% y 30%, respectivamente, de la muestra de ELE—. Consiguientemente, tenemos dos muestras comparables por presentar carácter común con respecto a la estructura interna.

4.5. Criterios de edición

Después de recoger datos a través de las encuestas, teníamos a nuestra disposición las respuestas aportadas por los informantes. Estas constituían la base sobre la cual elaboramos las listas de disponibilidad léxica. Hemos comentado en el apartado 4.3.2 que, para llevar a cabo el procesamiento de los datos, había que someter las respuestas a unos criterios de edición e

insertar en el programa informático las palabras resultantes del tratamiento¹⁰¹. El presente apartado se dedica a exponer los criterios generales que hemos seguido para la edición de las respuestas y las pautas específicas en cada uno de los centros de interés.

Cabe mencionar que exponemos en este apartado solamente los criterios o pautas relativos a la edición de los datos en español como lengua extranjera obtenidos de nuestros informantes y no profundizamos en el tratamiento de los datos en lengua materna, aunque este también ha implicado normas por su parte. Tal decisión se debe a que la presente tesis se centra en la disponibilidad léxica en ELE de nuestros informantes y en las comparaciones sobre la base de ella. El estudio sobre la muestra de disponibilidad léxica en CLM de los mismos informantes sirve como análisis contrastivo, a partir del que sacamos informaciones sobre cómo es la disponibilidad léxica en ELE en comparación y cómo se puede mejorar. De modo que, en esta tesis, dejamos de explicar los criterios de edición sobre los datos en lengua materna de los estudiantes¹⁰².

4.5.1. Criterios generales

Los investigadores de disponibilidad léxica del español como lengua materna, sobre todo los del *Proyecto Panhispánico*, tienen cierto acuerdo en partir del año 1998 para homogeneizar los criterios de edición, puesto que tienen como objetivo comparar las distintas sintopías y las decisiones divergentes a la hora de editar los materiales pueden generar diferencias

¹⁰¹ Se aplican unos criterios de edición porque las respuestas, en bruto, pueden presentar, dentro de la misma palabra, variabilidad: además de eventuales errores ortográficos, si son sustantivos o adjetivos con moción de género y número, pueden estar escritos en masculino o femenino, en singular o plural; si son verbos, pueden aparecer en forma no personal o conjugados, en cualquier tiempo, modo, número y persona. Toda esta variabilidad dispersa las respuestas, por tanto, es imprescindible aplicar algún sistema para uniformar. Asimismo, no hay que olvidar que los estudios de disponibilidad ofrecen siempre un diccionario de léxico disponible, por lo que el vocabulario precisa de un tratamiento similar al que los lexicógrafos aplican a los lemas de un diccionario de lengua.

¹⁰² A pesar de ello, creemos conveniente destacar que hemos mantenido, en la edición de los datos tanto de ELE como de CLM, los mismos criterios en lo que se refiere al tratamiento sobre la amplitud de las relaciones asociativas registradas en cada centro de interés —cosa a la que dedicamos un espacio considerable en adelante al exponer las características específicas de cada campo léxico—.

artificiales entre una sintopía y otra. Han decidido eliminar en la medida de lo posible las discrepancias en las pautas empleadas para editar las palabras y estas pautas comúnmente adoptadas por las investigaciones hispánicas fueron recogidas en un trabajo de Samper Padilla (1998). La presente tesis se ha adherido a esos criterios generales porque ello va a permitir futuras comparaciones con muchas investigaciones que los han seguido.

Sin embargo, siendo nuestros informantes estudiantes de español como lengua extranjera, las respuestas que se han de editar, seguramente, cuentan con algunas peculiaridades que los estudios del español como lengua materna no han tomado en consideración. Por lo tanto, también hemos seguido otras normas propuestas por Carcedo González (2000) y Samper Hernández (2001) en relación con informantes extranjeros, dos investigadores pioneros en disponibilidad léxica sobre estudiantes de ELE, los cuales ingeniaron sus consideraciones al tratar los datos aportados, respectivamente, por los alumnos finlandeses en su país de origen y por los estudiantes extranjeros inmersos en los contextos lingüísticos de la Universidad de Salamanca.

Además de las tres fuentes de referencia arriba mencionadas y, asimismo, frecuentemente citadas en muchas otras investigaciones, también hemos tenido en cuenta las pautas empleadas en los trabajos de Gómez Devís (2004), Serrano Zapata (2014) y Sánchez-Saus Laserna (2016). Estas autoras han desarrollado de manera explícita el tema de los criterios generales y Sánchez-Saus Laserna, sobre todo, se ha centrado “en las pautas más propiamente relacionadas con la disponibilidad léxica aplicada a estudiantes de español como lengua extranjera y en cómo ha solucionado las dificultades que plantea la creación de una base de datos de disponibilidad léxica a partir de datos obtenidos de tales informantes” (Sánchez-Saus Laserna, 2016: 47).

Para consultas con respecto a la ortografía, morfología y lexicología, hemos utilizado, con relación al español, el *Diccionario de la Lengua Española* (en adelante, DLE) de la Real Academia Española de versión electrónica 23, de acceso a través de la página web <https://dle.rae.es>. Y como obras de

referencia complementarias, también hemos recurrido al *Diccionario de Uso del Español* (en adelante, DUE), de María Moliner (1998), y al *Diccionario General de la Lengua Española Vox* (en adelante, Vox). Por lo que se refiere a la lengua inglesa, para la edición de las palabras en inglés detectadas en las listas, hemos tenido como referencia fundamental el *New Oxford American Dictionary* (en adelante, NOAD) y el *Oxford Dictionary of English* (en adelante, ODE).

Como resumen, los criterios generales que hemos seguido para la edición de nuestros datos son los siguientes.

1) Eliminación de los signos de puntuación. Hemos suprimido todos los signos de puntuación que los informantes escribieron junto con las palabras, con excepción de aquellos que son propios de las unidades léxicas, como el guion (-) de *kung-fu* y *ping-pong*. Hemos encontrado muchos casos en los que los alumnos terminaron alguna palabra con un punto (.). Se entiende que estos informantes apoyaban la punta de la pluma haciendo un punto en señal de haber acabado una línea, una lexía o una serie de palabras. También hemos hallado algunas comas, que utilizaron unos alumnos con la finalidad de poner en una misma línea palabras diferentes pero relacionadas, como las comas en “*sol, luna*” y “*pelo, cabello*”. Hemos suprimido asimismo los dos puntos (:) que fueron empleados para enumerar los elementos de un conjunto, como en “*miembros: brazos, piernas*”. Del mismo modo hemos tratado algunas pocas comillas (“”) a las que unos estudiantes recurrieron porque no estaban seguros de la grafía de una palabra o de si se podía escribir una palabra.

2) Separación de las respuestas diferentes en una misma línea. En este caso, siempre hemos colocado la segunda respuesta en la línea de abajo y, si había más, hemos puesto las diferentes respuestas en sendas líneas debajo, según su orden, de la línea original en la que habían escrito la secuencia de palabras. De esta manera han sido separadas las palabras en algunos casos ya comentados en el apartado sobre los signos de puntuación, una vez eliminados estos. Además, hemos detectado otros casos en los que

los alumnos emplearon un “y” para enlazar las palabras, como en “*cuchillo y cuchara*”. Al ser así, hemos suprimido el “y” antes de separar las respuestas.

3) Eliminación de las respuestas repetidas. Las palabras o unidades léxicas repetidas en un mismo centro de interés aportadas por un mismo estudiante han de ser consideradas como descuido del informante. Se conserva la aparición de una misma respuesta en la posición más alta de la lista, es decir, la segunda e, incluso, las siguientes apariciones han de ser suprimidas. Algunas respuestas que no son exactamente iguales, pero que resultan ser iguales después de someterse a los demás criterios, también son consideradas como repetidas.

4) Neutralización de las variantes flexivas. Esta es una pauta comúnmente acordada por los trabajos de disponibilidad léxica, con el fin de evitar las inútiles repeticiones de una palabra en diferentes formas meramente flexivas. Hemos unificado las unidades léxicas en su forma no marcada del paradigma, siguiendo la manera como un diccionario recoge las entradas. En concreto, si no se trata de sintagmas —*ir de compras*, por ejemplo, con los que tenemos otras consideraciones que se presentan en adelante— los sustantivos y adjetivos se han insertado en el corpus de disponibilidad en la forma masculina y singular, mientras los verbos, lo han hecho en infinitivo.

En los casos de aquellos sustantivos que suelen usarse en forma plural como *pantalones*, *gafas*, *tijeras* y *vacaciones*, también los hemos singularizado porque la gran mayoría de nuestros informantes los escribieron en singular y porque en el DLE igualmente se recoge su forma en singular; a excepción de *afueras* cuya forma en plural hemos mantenido, porque todas las respuestas registradas en nuestras encuestas de esta palabra están en plural. Sin embargo, las palabras que contienen un plural lexicalizado —*lavaplatos*, *lavavajillas*, etc.—, claro está, las hemos admitido con la -s al final.

Las circunstancias que presenta el género son más complejas. Las palabras que hacen referencia a las profesiones que suelen ejercer las mujeres

—*enfermera, azafata*, etc.—, las hemos transcrito en forma masculina, lo cual se aplica también en los demás estudios de disponibilidad léxica. No obstante, la forma femenina la hemos conservado al tratarse de palabras cuya forma femenina supone lexema distinto, es decir, en los heterónimos —*madre, mujer, yegua*, etc.— o presenta variación irregular en la grafía —*gallina, reina, actriz*, etc.—. Por último, en lo referente a los nombres de algunos aparatos, cuyas formas tanto masculina como femenina son de uso habitual y son recogidas en el DLE, hemos incluido en las listas de palabras ambos géneros en entradas distintas. Esto es lo que ha pasado con vocablos como *computador - computadora, secador - secadora, tostador - tostadora* y *aspirador - aspiradora*. Esta última decisión la hemos tomado teniendo en cuenta que no va a influir en el total de palabras ni el promedio, toda vez que en esta investigación ningún informante escribió en ningún centro de interés ambas formas de este tipo de palabras. Como siempre preferían una de las dos formas, sería interesante mirar cómo se manifiesta la disponibilidad del masculino y del femenino de estas palabras.

5) Unificación de los derivados diminutivos y aumentativos. Las variantes morfológicas de este tipo las hemos unificado, siempre que no implican alteración alguna del significado léxico. Esto último quiere decir que solamente hemos conservado aquellos derivados diminutivos o aumentativos ya lexicalizados. Por ejemplo, los aumentativos *tazón* y *cucharón* no hacen referencia a los mismos objetos que *taza* y *cuchara* —estas cuatro palabras aparecen en el DLE como entrada independiente—, por lo que las primeras han sido incluidas en las listas de disponibilidad al igual que las segundas. Lo mismo ha sucedido con los diminutivos como *mesilla, mesita* y *tortilla*. El resto, que pertenece a derivaciones apreciativas, como *casita* y *conejito*, se ha unido a su respectiva palabra base: siguiendo con los ejemplos, a *casa* y *conejo*.

6) Corrección ortográfica y gráfica. Por un lado, al igual que en otras sintopías, los errores ortográficos de carácter fónico o morfológico son numerosos. Por ejemplo, los más frecuentes en nuestras encuestas han sido: i) la falta o uso erróneo de la tilde; ii) la confusión entre *p - b - v*; iii) la confusión

entre *t - d*; iv) la confusión entre *c - qu - g*; v) la confusión entre *z - c*; vi) la confusión entre *g - j*; vii) la confusión entre *r - l*; viii) la sustitución del diptongo por la vocal radical; y ix) la sustitución de *m* por *n* ante *b* o *p*. Todos estos errores ortográficos hay que corregirlos y las palabras ya corregidas las admitimos en el corpus léxico de disponibilidad.

No obstante, por otro lado, dado que nuestros informantes son aprendientes del español como lengua extranjera y sus niveles son variados, hemos encontrado innumerables errores meramente gráficos. Las razones por las que cometieron estos errores están más allá del desconocimiento del sistema ortográfico y es difícil atribuirles alguna explicación¹⁰³. Generalmente, los errores gráficos se manifiestan como ausencia o, por el contrario, adición de letras o sílabas y como cambio de una letra o sílaba por otra. Por ejemplo, con respecto a la respuesta *refrigerador* en el centro de interés *Los muebles de la casa*, hemos hallado varias formas erróneas como *refrisgrador*, *refregeriador*, *refregerator*, *refrigrador*, *refregeredor*, *refrigerator*, *refriegador*, *refriesgodor*, *refregirador*, *refregitora*, *resfragidor*, *refrigador*, *refrigetor*, etc. Al tratar estos errores de grafía defectuosa, hemos adoptado el criterio planteado por Samper Hernández (2001), es decir, los hemos corregido y, después, admitido en las listas de palabras, siempre que podemos adivinar el equivalente en español de la palabra aun errónea, teniendo en cuenta el centro de interés en que esta se encuentra y el contexto en que aparece. Así que las respuestas que aportan poca información, en cuanto a la grafía, sobre su relación con alguna palabra española han sido eliminadas.

7) Unificación de la ortografía en los extranjerismos. Para comenzar,

¹⁰³ A pesar de ello, Sánchez-Sauz Laserna (2016) ha hecho grandes contribuciones al respecto. La autora se dedicó a clasificar los errores registrados en las respuestas aportadas por los estudiantes de ELE en las universidades andaluzas y resumió de manera muy explícita las faltas gráficas propias de sus informantes. Este estudio suyo es una fuente de referencia importante en el ámbito del análisis de errores en el léxico disponible de alumnos de ELE. Pero no hemos podido aplicar completamente sus clasificaciones en la presente tesis, debido al perfil diferente de nuestros informantes, toda vez que su lengua materna (diferente a las analizadas por Sánchez-Saus, todas ellas europeas) y su contexto lingüístico (en el país de origen y no en inmersión lingüística) han generado problemas distintos.

llamamos la atención sobre el hecho de que no estamos hablando de la unificación de la ortografía de las palabras propias del español, sino de los extranjerismos en esta lengua. Decimos esto porque en la mayoría de los estudios de disponibilidad léxica en español como lengua materna, cuando el investigador se enfrenta con el caso de que los informantes escriben correctamente una palabra con varias formas, han optado por unificarlas y mantener una forma. A este respecto, en relación con las palabras propias de la lengua española, hemos hallado en nuestras encuestas solamente un caso, que es el de *hierba* - *yerba*. Sin embargo, hemos decidido no unificar las dos formas y mantener ambas, porque nuestro objetivo es distinto. A nuestro juicio, es bueno saber que nuestros estudiantes de español como lengua extranjera conocen ambas formas de esta palabra y, asimismo, nos interesa averiguar cuál de las dos formas actualizaron mejor los informantes, o sea, qué forma tiene una disponibilidad más alta. Esta decisión no va a alejar nuestros criterios de los comunes, puesto que se trata de un caso singular.

En cuanto al resto de las palabras que pueden tener más de una forma, o sea, que tiene variación gráfica —lo cual es destacado en los extranjerismos—, la decisión que hemos tomado es, tal y como resume el epígrafe de esta parte, la unificación de la ortografía. En el tratamiento de estos préstamos lingüísticos, hemos distinguido varios tipos. En primer lugar, están aquellos extranjerismos que se han adaptado a las reglas fonéticas y gráficas del español. Si existen varias formas de una palabra, pero el DLE solo recoge una, hemos incluido en nuestro estudio la forma recogida por el diccionario académico. Por ejemplo, para *bidé* - *bidet* y *koala* - *coala*, solo hemos admitido las formas del DLE *bidé* y *koala*. Si existen varias formas de una palabra y el DEL también recoge más de una, hemos mantenido la forma que con más frecuencia apareció en nuestras encuestas. Por ejemplo, para *kung-fu* - *kungfú*, *taichí* - *taichi*, *fútbol* - *futbol* y *voleibol* - *vóleibol*, todos recogidos en el DLE, hemos unificado las formas manteniendo solamente *kung-fu*, *taichí*, *fútbol* y *voleibol*, porque estas fueron aportadas por más informantes que las otras formas.

En segundo lugar, en lo que se refiere a los extranjerismos no adaptados

a las reglas del español y que conservan su forma de la lengua de origen, hemos realizado la edición según sea inglés o no su lengua de origen. En caso de que esta no sea en inglés —tenemos muy pocos casos de esta índole y citamos como ejemplo *pizza*, que es del italiano—, hemos incluido la palabra solo si también es recogida por el DLE y, además, solo hemos admitido la única forma recogida en el diccionario académico pese a que pueda tener varias formas en su lengua de origen. Por otra parte, en caso de que la lengua de origen de la respuesta sea el inglés, hemos procesado su edición de acuerdo con el siguiente criterio: puesto que los extranjerismos del inglés no adaptados a las reglas del español y que conservan su forma de la lengua de origen son palabras propias de la lengua inglesa, se escribirán exactamente igual que se escriben en inglés.

8) Tratamiento de las palabras del inglés. En relación con las palabras propias de la lengua inglesa, la decisión que hemos tomado es incluirlas todas en nuestras listas de palabras. Se puede distinguir dos tipos de vocabulario propio del inglés. Por un lado, hemos detectado en nuestras encuestas muchos vocablos de este idioma que el DLE también recoge en sus volúmenes. Estos forman parte de los préstamos lingüísticos en la lengua española que han conservado su forma original, por lo que no hemos tenido duda alguna en admitirlos en nuestro corpus de disponibilidad. De este tipo de palabras son, por ejemplo, *jean, jazz, footing, camping, marketing* y *stock*.

Por otro lado, están las palabras inglesas no recogidas por el DLE. Para el tratamiento de este tipo de palabras, Sánchez-Saus Laserna (2016), entre otros investigadores, optó por incluir en las listas de disponibilidad solamente aquellas que se cree que están muy extendidas en el español, como *wifi*. Se puede deducir que estos investigadores han visto la importancia de algunos vocablos ingleses en la comunicación diaria en español. Sin embargo, nos parece difícil delimitar cuál está más extendida que otra y las hemos incluido todas con motivo de los objetivos particulares de nuestra investigación. La consideración a partir de la cual tomamos esta decisión se puede consultar en el apartado 4.4.4, donde explicamos el papel que desempeña la lengua inglesa

en la adquisición del español de nuestros informantes. En síntesis, el inglés es la segunda lengua extranjera de estos alumnos e incluso puede considerarse como la primera lengua extranjera para los estudiantes de niveles de español bajos. El hecho de que escribieran una palabra inglesa en las encuestas no se debe a que supieran que esta estaba incluida en un diccionario español. Por lo tanto, incluir las palabras inglesas recogidas por el DLE o las más extendidas y excluir el resto es seleccionar artificialmente, del vasto vocabulario del inglés de los alumnos, una pequeña parte que por casualidad coincide con los criterios del diccionario español o con cierto juicio subjetivo del investigador. Dado esto, a lo que se añade nuestra intención de estudiar en el futuro la incidencia del nivel de inglés de los estudiantes, hemos decidido mantener todas las palabras actualizadas en este idioma, aunque no estén incluidas en el DLE. La unificación de formas con respecto a estas palabras —a pesar de que estas respuestas aportadas por los informantes presentan poca variación gráfica— la hemos basado en el NOAD y el ODE.

9) Unificación de los acortamientos y las formas plenas. Al igual que en casi todas las demás investigaciones de disponibilidad léxica, hemos optado por unificar los acortamientos de algunas palabras y sus formas plenas, utilizando el paréntesis, de modo que evitemos la repetición de estas palabras en dos formas e indiquemos la registrada y correcta eliminación de parte de ellas. Tales casos en nuestras listas de palabras se han limitado a vocablos como *bolí(grafo)*¹⁰⁴, *tele(visión)*, *bici(cleta)* y *foto(grafía)*.

Cabe mencionar que contamos, en esta investigación, con un caso particular en relación con la palabra *autobús*, la cual se suele incluir en los demás estudios como *(auto)bús*. Es verdad que en nuestras encuestas se han registrado tanto *autobús* como *bus*, formas que se deberían unificar como *(auto)bús* siguiendo el presente criterio. Sin embargo, nos hemos dado cuenta de que el *bus* que ha aparecido en las listas también puede ser simplemente inglés y no tratarse del acortamiento en el español. Por otra parte, también

¹⁰⁴ La palabra editada —*bolí(grafo)*— ha seguido las reglas de acentuación del español, aunque en el acortamiento —*boli*— no esté presente la tilde.

hemos detectado apariciones de *auto*, que es acortamiento de *automóvil*. La edición de *auto* en diferentes investigaciones ha seguido criterios heterogéneos: en unas, se ha incluido como *auto(móvil)* indicando la aparición de *auto* y *automóvil*; en otras, simplemente como *automóvil* sin aludir a la existencia de *auto*; y en el resto, están incluidos tanto *auto* como *automóvil* por separado. Así pues, como resumen, hemos entendido que los informantes nuestros, al realizar la prueba asociativa, no estaban seguros de que i) *bus* era acortamiento de *autobús* y escribieron en inglés dicha palabra; y ii) *auto* era acortamiento de *automóvil* y no de *autobús*. Dadas estas circunstancias de *auto - bus - autobús - automóvil* y la falta de homogeneidad para la edición de *automóvil*, hemos conservado en esta investigación las tres formas sin usar el paréntesis: *auto*, *autobús* y *automóvil*.

10) Unificación de la combinatoria léxica y su elipsis. Esta también es una pauta establecida por el *Proyecto Panhispánico* y comúnmente aceptada en casi todos los estudios de disponibilidad léxica. De igual modo que el criterio anterior, hemos usado el paréntesis para unificar, en este caso, las combinaciones léxicas que unas veces aparecen con parte de sus componentes y otras veces con la combinación completa. Por ejemplo, en el centro de interés *Partes del cuerpo humano*, algunos informantes escribieron *columna* en referencia a la columna vertebral y otros anotaron, de manera explícita, *columna vertebral*. Hemos unido los dos casos en *columna (vertebral)*. Combinaciones léxicas parecidas y recabadas en nuestras listas de disponibilidad las hay como *campana (extractora)*, *(cuarto de) aseo* y *(cuarto de) baño*. Hemos de enfatizar que esta última, es decir, la unificación de *baño* y *cuarto de baño* en *(cuarto de) baño*, solo se ha dado en el centro de interés *Partes de la casa (sin muebles)*, puesto que la aparición de *baño* en este centro hace clara referencia al cuarto que desempeña tal función, pero, en el resto de los centros, no es necesariamente así y no hemos hecho la unificación.

11) Eliminación de los modificadores. Por lo general, hemos eliminado los modificadores que acompañan a las unidades léxicas, incluidos principalmente los artículos y, asimismo, los adjetivos y adverbios, con la

finalidad de evitar que se repitan los mismos sustantivos.

Acerca de estos modificadores, llamamos la atención sobre el hecho de que son muy numerosos los artículos que hemos detectado en las encuestas, lo cual coincide con el estudio que Samper Hernández (2001: 279) llevó a cabo con los alumnos extranjeros de Salamanca. Esta autora señaló la aparición extendida de los artículos en sus encuestas, considerándolo de gran interés. Aquí intentamos dar algunas explicaciones al respecto, a partir de los casos observados en nuestras encuestas. Suponemos, por un lado, que los estudiantes se aprenden el género de algunos sustantivos recurriendo a un artículo antepuesto. Por ejemplo, *el mapa* y *la luz* tienen una frecuencia manifiestamente superior a *mapa* y *luz*. Por otro lado, sospechamos que los alumnos adquieren el sustantivo de aquellos objetos que suelen presentarse en pareja también anteponiendo un artículo para recordar dicha pluralidad. Por ejemplo, *los ojos*, *las orejas* y *las gafas* casi siempre van acompañados de artículo. Además, para aquellas palabras que, en la mayoría de las situaciones, el artículo es requerido, nuestros informantes las han actualizado casi siempre con artículo. Por ejemplo, *el sol* y *la luna* aparecen bastante más veces que *sol* y *luna*. A pesar de ello, nos parece conveniente eliminar todos estos artículos, siendo el presente estudio de carácter léxico. No obstante, hemos mantenido los artículos, si aparecen dentro de un sintagma que, según alguno de los otros criterios establecidos, se ha de conservar. Citamos como ejemplo *todo el mundo* y *remover la tierra*.

En cuanto a los adjetivos y adverbios que acompañaban a los sustantivos que actualizaban de acuerdo con el centro de interés, los hemos suprimido igual, pero con excepción de aquellos que suponen una alteración en el significado de la respuesta. Por ejemplo, con respecto a *ropa interior*, *pantalón corto*, *coche eléctrico*, *colegio mayor* y *mover la cabeza afirmativamente*, creemos que es necesario conservar los adjetivos o adverbios, toda vez que, con estos, la combinación de palabras implica cosas bastante distintas en comparación con el uso aislado del sustantivo. En adelante, veremos más casos de este tipo en cada centro de interés e iremos indicando sus respectivas

razones específicas.

12) Tratamiento de los sintagmas. En primer lugar, con respecto a los sintagmas preposicionales y nominales registrados en nuestras encuestas, hemos decidido conservarlos con tal de que tengan estructura sintáctica fija o hagan referencia semántica única, porque, de ser así, o bien son unidades fraseológicas o bien son de uso común. De lo contrario, hemos de eliminarlos o simplificarlos conservando solamente el sustantivo que aparece en el grupo sintáctico. Por ejemplo, hemos admitido *a pie, a tiempo, a distancia, lámpara de descarga, mesita de noche, palillo de dientes, habitación de matrimonio, fruta del tiempo, lugar de interés, energía limpia, agua potable*, etc.; mientras tanto, hemos conservado, de *traje de moda y salas de la casa*, solo los núcleos *traje* y *sala*.

Cabe mencionar que, en lo que se refiere a los sintagmas nominales, parte de ellos —los constituidos por un sustantivo más un adjetivo— coinciden con los casos del criterio anterior sobre el tratamiento de los modificadores adjetivos; es decir, los ejemplos *energía limpia* y *agua potable* arriba figurados bajo el presente criterio son lo mismo que algunos otros en el criterio anterior: *ropa interior, pantalón corto, coche eléctrico y colegio mayor*. Estos casos particulares se pueden someter a ambos criterios, pero, sin cambiar el resultado de la edición.

En segundo lugar, también tenemos en las encuestas un motón de sintagmas verbales. Con respecto a estos, la decisión que hemos tomado es mantenerlos en la medida de lo posible. Por un lado, en relación con los sintagmas con verbo que se consideran unidades fraseológicas —por ejemplo, *tomar el sol* e *ir de compras*—, no hay duda alguna para incluirlos en nuestro corpus de disponibilidad. Junto con ellos, hemos incorporado también aquellos que, sin ser unidades fraseológicas, son comúnmente aceptados en la lengua española. Citamos, como ejemplos de este tipo, *trabajar la tierra, preparar la comida, apagar la luz* y *montar a caballo*.

Por otro lado, ante la medida de algunos otros investigadores que insertaron, por ejemplo, los sintagmas *escuchar música* y *jugar al tenis* como *música* y *tenis* para que no se calculara la disponibilidad de las palabras “música” y “tenis” en dos casos distintos¹⁰⁵, creemos que es conveniente mantener los verbos para nuestra propia investigación¹⁰⁶. Dado que nuestros informantes son estudiantes de español como lengua extranjera, es muy posible que uno que puede emplear *música* o *tenis* en ciertas circunstancias comunicativas no sea capaz de actualizar los verbos “escuchar” o “jugar” de manera independiente o hablando de la música o del deporte, respectivamente, o bien que no sepa usar correctamente el compuesto sintagmático *jugar al tenis*. Por eso, la aparición de *escuchar música* y *jugar al tenis* en nuestras listas de palabras es bastante significativa, en el sentido de que informa del dominio y la disponibilidad tanto de los sustantivos como de los verbos y de la sintaxis de los compuestos sintagmáticos.

Además, dentro de los sintagmas verbales registrados en nuestras encuestas, hemos detectado algunas combinaciones de verbo y sustantivo que los hablantes nativos no utilizan habitualmente, pero que no resultan incorrectas desde el punto de vista de la morfosintaxis. A este respecto, hemos seguido lo que hicieron Carcedo González (2000) y Samper Hernández (2002), incorporando esos sintagmas tal y como los informantes los aportaron (Samper Hernández, 2001: 283). Nos referimos a casos, sobre todo del centro de interés

¹⁰⁵ En estas investigaciones, preferían eliminar el verbo y mantener solamente el sustantivo siempre que el verbo describiera la acción directamente relacionada con el sustantivo. Aquí citamos más ejemplos de este tipo que aludió Sánchez-Saus Laserna (2016: 53): *ver la televisión*, *hacer la compra*, *hacer deporte*, *ir a la universidad*, *hacer sol*, *hacer calor* y *hacer frío*, los cuales se unificaron en dichas investigaciones bajo *televisión*, *compra*, *deporte*, *universidad*, *sol*, *calor* y *frío*. Sin embargo, cuando se trataba de combinaciones sintagmáticas en las que el verbo realizaba una acción no esperable, algunos otros estudios —por ejemplo, en Gómez Devís (2004), Serrano Zapata (2014) y la mayoría de los estudios del *Proyecto Panhispánico*— sí mantuvieron el verbo, como en los casos de *tocar un instrumento*, *escribir poesía*, *componer música*, *actuar en televisión*, etc.

¹⁰⁶ Con respecto a la decisión de mantener los verbos de los sintagmas, además de nuestra investigación, Samper Hernández (2001) también los mantuvo, pero parcialmente. Concretamente, incorporó separados aquellos casos en los que el verbo o el complemento directo variaban: por ejemplo, la investigadora mantuvo los verbos en los siguientes casos: *escuchar música* - *oír música*, *ver televisión* - *mirar televisión*, *ir en bicicleta* - *montar en bicicleta*, *leer libros* - *leer novelas*, etc., con la finalidad de observar la vigencia de cada construcción.

Trabajos del campo y del jardín, como poner agua, matar los insectos, quitar hierbas, alimentar animales y recoger frutas.

13) Tratamiento de los nombres de marcas o establecimientos comerciales. En lo referente a incluir o no los nombres de marcas o establecimientos comerciales, los investigadores de disponibilidad léxica no han llegado a un acuerdo. Algunos insisten en incluir solamente aquellos nombres que ya están totalmente lexicalizados —los que en ciertas situaciones comunicativas pueden sustituir el nombre de los productos de su especie, por ejemplo, *coca cola* en sustitución de refrescos con gas—. Mientras tanto, otros dan por adecuado incluirlos todos porque resulta prácticamente difícil determinar si cierto nombre está lexicalizado o en qué grado puede considerarse su lexicalización. Para evidenciar dicha divergencia, Serrano Zapata (2014: 266) enumeró a los autores de cada una de las dos corrientes de opinión y se encontraron, en ambas, una buena cantidad de partidarios.

Teniendo en cuenta la particularidad de nuestros informantes, los cuales estudian el español como lengua extranjera, pensamos que, además de aprender los nombres comerciales con el fin de emplearlos como sustitución de los productos de la especie, la adquisición de dichos nombres todavía les sirve para algo más: el aprenderse esos nombres comerciales como unidades léxicas puede instruirles en el conocimiento propio del léxico acerca de la fonología, la morfología, etc., toda vez que esos nombres, como palabras, intrínsecamente contienen rasgos comunes que comparten algunos vocablos de la lengua estudiada. Por lo tanto, a nuestro juicio, la disponibilidad léxica que tienen los informantes de ELE de los nombres comerciales es, asimismo, la disponibilidad de algunas normas lingüísticas, sobre todo fonológicas y morfológicas como se ha comentado. De esto no debe haber duda, porque seguro que no todos los alumnos saben transcribir correctamente dichos nombres en el sistema de escritura, a pesar de que conozcan las marcas o, incluso, sean capaz de pronunciarlas. Dado esto, creemos conveniente conservar todos los nombres de marcas o establecimientos comerciales que aparecen en nuestras listas de palabras y así lo hemos hecho en la edición.

14) Tratamiento de las relaciones asociativas ampliadas. La prueba asociativa del vocabulario pide a los informantes que escriban el mayor número posible de palabras que conocen relacionadas con el tema de cada centro de interés. Como consecuencia, se encuentran en casi todas las investigaciones bastantes palabras que resultan no estar estrechamente relacionadas con la temática. Esto se conoce como la amplitud de las relaciones asociativas que intentamos tratar en este apartado. En nuestro caso, es aún mayor el alejamiento del campo léxico de algunas palabras aportadas por los alumnos de ELE. A este respecto, hemos seguido el criterio de Samper Hernández (2001) y Sánchez-Saus Laserna (2016), quienes trabajaron también con informantes de ELE. Esto quiere decir que hemos conservado, en la medida de lo posible, las respuestas dadas por los estudiantes.

En concreto, ante alguna palabra que, al parecer, se aleja considerablemente del significado léxico que expresa el centro de interés correspondiente, hemos tratado de buscar explicaciones que puedan justificar su inclusión en la lista y hemos hallado que la mayoría de esos casos no resultan ser completamente inexplicables, puesto que su aparición relativa a un centro de interés determinado puede informar de connotaciones psicológicas o socioculturales. La exclusión de estas informaciones del corpus de disponibilidad sería lamentable, porque, por un lado, no existe en realidad una definición fija del límite de las relaciones asociativas y, por otro lado, los estudios de disponibilidad léxica constituyen, de por sí, una fuente de referencia para las investigaciones de tipo psicológico y sociocultural. En síntesis, hemos dado cabida a las relaciones asociativas ampliadas¹⁰⁷ y solo hemos eliminado de las listas aquellas respuestas de relación del todo indescifrables.

¹⁰⁷ Esta cuestión se va a tratar largamente en los apartados que siguen.

4.5.2. Características específicas en los centros de interés

A continuación, vamos a ver los rasgos específicos que presenta cada uno de los dieciséis centros de interés en lo referente a la edición de los datos. Explicaremos las soluciones a los problemas particulares, sobre todo a la amplitud de relaciones asociativas por parte de nuestros informantes de ELE en cada campo léxico.

4.5.2.1. Partes del cuerpo humano

En este centro de interés hemos singularizado un montón de palabras que escribieron en plural, aquellos sustantivos que se refieren a las partes que tiene el cuerpo humano en pareja o en plural como *ojos, orejas, hombros, brazos, manos, piernas, rodillas, pies, dientes, dedos, uñas*, etc. Como hemos comentado en el apartado anterior, muchas de estas palabras van acompañadas de un artículo que hemos de eliminar como en *los ojos, las orejas, los brazos, los labios y las cejas*. Además, hemos unificado *los dedos del pie y dedo de pie* en *dedo del pie*. Con respecto al uso del paréntesis, hemos unificado *columna y columna vertebral* en *columna (vertebral)*.

Hemos hallado muchos verbos en este centro, de los cuales hemos conservado aquellos relacionados con alguna parte del cuerpo y aquellos que describen acciones que realiza el cuerpo utilizando alguna de sus partes o actuando sobre alguna de sus partes: *ver, leer, mirar, cepillar, caminar, correr, reír, cantar, lavar, fregar, limpiar, llegar, salir, desayunar, tomar, traer*, etc. Se han registrado bastantes verbos reflexivos, los cuales hemos unificado bajo la forma infinitiva de los mismos verbos: *bañarse, peinarse, sentarse, levantarse, arrodillarse*, etc.

Han aparecido, bajo esta área temática, algunas palabras que se espera que se evoquen en el siguiente centro de interés *Ropa y complementos*, como son *máscara, guante, chaqueta, falda, cinturón*, etc. Hemos optado por conservarlas porque el mecanismo de nuestra prueba asociativa hizo que los

alumnos no supieran que se les solicitaría tales palabras más tarde (véase el apartado 4.3.1.3) y porque, de verdad, estas respuestas sobre la ropa o sus complementos pueden ser actualizadas en situaciones comunicativas relativas al cuerpo humano. Por razones parecidas a esta última, hemos mantenido algunos adjetivos que describen alguna parte del cuerpo humano como *fuerte, liso, largo, blanco, negro, sano y gordo*; algunos sustantivos que hacen referencia a objetos con los que el cuerpo puede tener contacto como *gel, crema, sofá, alfombra, cama y lecho*; y, asimismo, otros sustantivos cuya relación asociativa con el cuerpo es fácil de enlazar como *beso, respiración, tos, dolor, operación, cirujano, médico, sentimiento, salud y deporte*.

En lo que se refiere a los diminutivos, hemos conservado *barbilla, codillo, calzoncillo y zapatilla*, que constituyen entradas independientes en el diccionario DLE. El único sintagma verbal registrado en este centro de interés ha sido *mover la cabeza afirmativamente*, el cual hemos mantenido, como se ha comentado en los criterios generales, por su referencia semántica única. Un caso particular que no ha ocurrido en otras sintopías es que hemos incluido el sintagma *cinco sentidos*, porque su traducción correspondiente en la lengua materna de los informantes —o sea, el chino— consiste en una palabra única e independiente y de uso frecuente que abarca semánticamente la vista, el oído, el olfato, el gusto y el tacto.

4.5.2.2. Ropa y complementos

Al igual que en el área temática anterior, han aparecido en este centro de interés numerosas palabras en plural que hacen referencia a objetos que suelen presentarse en pareja: *pantalones, zapatos, calcetines, guantes, gafas, pendientes, lentillas*, etc. Todas estas las hemos unificado manteniendo su forma en singular, con excepción de *gafas de sol*. En cuanto a los sintagmas, hemos conservado solamente los sustantivos en *broche alargado, traje de moda, falda de gala y anillo de boda*; mientras que hemos mantenido la forma completa de los siguientes sintagmas: *ropa interior, pantalón chino, zapato de tacón y zapato de cuero*, cada uno de los cuales se refiere a una subespecie

distintiva de la especie que denota el sustantivo, sea con respecto al modelo o al material. La unificación de las variantes la hemos hecho en *biquini* - *bikini*, uniéndolos bajo *biquini*.

Hemos encontrado gran cantidad de palabras de color que representan la variedad de colores de la ropa. En los casos en los que el adjetivo que indica el color tiene moción de género, se ha mantenido su forma masculina como en *rojo* y *blanco*. Todas las formas en plural se han singularizado. El género gramatical femenino se ha detectado en las parejas de palabras como *hombre* - *mujer*, *chico* - *chica* y *niño* - *niña*, donde hemos conservado ambos géneros porque lo que querían los informantes era distinguir la ropa para hombres y para mujeres. En lo referente a los verbos registrados en las encuestas, hemos incluido aquellos que suelen tener como complemento directo alguna pieza de ropa: *usar*, *poner*, *llevar*, *limpiar*, *lavar*, etc. y aquellos directamente relacionados con la ropa como *vestir*, *abrigar* y *enguantar*.

Han sido muy amplias las asociaciones que aportaron nuestros informantes respecto a este campo léxico y hemos aceptado la mayoría de ellas. Ha habido palabras que aluden a los diferentes materiales que se emplean para confeccionar la ropa y sus complementos: *oro*, *plata*, *piel*, *seda*, *textil*, etc.; ha habido sustantivos relativos a los adornos en la ropa: *estrella*, *flor*, *pluma*, *esmeralda*, *diamante*, etc.; ha habido adjetivos que describen, desde distintas perspectivas, la ropa o a la persona que usa alguna prenda de vestir: *lujoso*, *sencillo*, *grosero*, *caro*, *limpio*, *sucio*, *bonito*, *feo*, *confortable*, *clásico*, *moderno*, *gordo*, *estrecho*, etc.; y ha habido palabras que entran en relación con la compraventa de la ropa y con el comercio de sectores afines: *dinero*, *adquirir*, *supermercado*, *tienda*, *cadena*, *perfume*, *cosmético*, *joyería*, *artesanía*, etc. Por último, los nombres de empresas y marcas del sector de la ropa como *Zara*, *Mango* e *Inditex* los hemos incluido todos en nuestro corpus de disponibilidad.

4.5.2.3. Partes de la casa (sin muebles)

En este centro de interés ha habido menos casos de singularización en comparación con los dos anteriores. La mayoría sucede con las partes de la casa que, de por sí, son varias y, por ello, se piensa en plural cuando se las evoca como *habitaciones, escaleras, ventanas, puertas* y *esquinas*. Como se ha comentado en los criterios generales, hemos utilizado el paréntesis en *(cuarto de) baño* y *(cuarto de) aseo*. Además, con respecto a los diminutivos, hemos insertado *ventanilla* tal cual.

Lo particular de la presente área temática consiste en que nuestros informantes aportaron algunos sintagmas nominales con “sala” o “cuarto” como núcleo aludiendo el uso a que se dedica cierta parte de la casa: *sala de lectura, sala del libro, cuarto de ropa, cuarto para estudiar, cuarto para dormir*. Estos sintagmas los hemos incluido en nuestras listas tal y como los escribieron los estudiantes. Por razones similares, es decir, para referirse a la función que asume determinada parte de la casa, los alumnos también aportaron bastantes verbos como *vivir, descansar, comer, cocinar, lavar, divertir* y *jugar*, los cuales hemos admitido en su forma infinitiva.

Nos hemos dado cuenta de que nuestros informantes han entendido por parte de la casa también sus componentes no materiales, por lo que escribieron palabras como *familia, padre, madre, niño, hijo, abuelo, nieto, pareja, novio, perro, gato*, etc. Estas respuestas relativas a los miembros de la familia las hemos admitido en el corpus sin problema alguno. De modo parecido, los alumnos todavía apuntaron objetos intangibles que se perciben en la casa, por ejemplo, palabras como *aire, luz* y *sol*, que igualmente hemos conservado.

Algunas asociaciones que han realizado los estudiantes giran alrededor de la construcción de la casa o de sus materiales de construcción. De ahí surgen respuestas como *desaguar, construir, madera, cristal, barro, escombros*, etc., que hemos mantenido en las listas de disponibilidad. Otro tipo de

asociación corriente entre los informantes es aquella que tiene vinculación con la ubicación de cierta parte de la casa o la relación de ubicación entre una parte y otra, lo cual también hemos aceptado. Citamos como ejemplo *ubicar, centro, lado, izquierda, derecha, norte, sur, este, oeste, dentro, debajo, encima, etc.*

Los adjetivos que expresan la cualidad de la casa y los sustantivos que tienen la misma raíz que algunos de estos adjetivos son otro tipo de respuestas que hemos hallado en las encuestas y aceptado en nuestras listas finales. Se trata de palabras como *grande, pequeño, cómodo, limpio, sucio, suciedad, arreglado, desarreglado, desorden, lujoso, humilde, vacío, lleno, alto, bajo, luminoso, oscuro, cuadrado, etc.*

Por último, hemos eliminado de la lista de este campo léxico numerosas palabras que han de aparecer en el siguiente centro de interés¹⁰⁸: *estantería, sofá, mesa, silla, cama, televisor, coche, cortina, lámpara, ordenador, teléfono, cepillo, escoba, fregona, etc.* La aparición de estas palabras se debe o bien a que los informantes no sabían que se les solicitaría más tarde palabras sobre los muebles, o bien a que ellos, en realidad, no tenían idea clara sobre qué objetos de la casa no son, seguramente, muebles. Aun así, hemos de admitir que el presente centro de interés y el que sigue comparten algunas palabras que tienen alta frecuencia en ambas listas.

4.5.2.4. Los muebles de la casa

Bajo este centro de interés, hemos empleado el paréntesis para unificar los acortamientos de *televisión* y *fotografía* —*tele* y *foto*— y sus formas plenas, insertándolos como *tele(visión)* y *foto(grafía)*. Hemos mantenido los diminutivos *mesita* y *mesilla* porque aparecen en el diccionario como entrada independiente y tienen referencia diferente de la simple mesa. Con respecto a

¹⁰⁸ La razón es que en el nombre del centro de interés ya se explicita que no se han de incluir muebles. No obstante, la decisión no ha sido fácil porque no existe una línea divisoria clara entre *Las partes de la casa (sin muebles)* y *Los muebles de la casa*. Con motivo de la demarcación bastante ambigua de estos dos centros de interés, Sánchez-Saus Laserna (2016) prefirió unirlos en uno, que resultó ser *La casa*.

las parejas *computador - computadora* y *secador - secadora*, hemos conservado las cuatro palabras como se ha establecido en los criterios generales. Se han registrado dos abreviaciones —aunque de origen inglés— en esta área temática: *TV* y *CD* y las hemos mantenido en el corpus.

De la palabra “*mesa*” y sus diminutivos han surgido varios compuestos sintagmáticos que tratan de aludir diferentes tipos de mesa, los cuales hemos conservado por su referencia semántica estable: *mesa de estudio*, *mesa de comer*, *mesa de té*, *mesa de tele(visión)*, *mesa de centro*, *mesilla de noche* y *mesita de noche*. Los verbos también tienen considerable frecuencia en el presente centro de interés y solamente hemos conservado aquellos que tienen relación con la función que asume algún mueble: *sentar*, *acostar*, *dormir*, *fregar*, *barrer*, *limpiar*, *enchufar*, *comer*, *leer*, *estudiar*, *colgar*, *adornar*, etc. La amplitud de relaciones asociativas también se da en los adjetivos, de los cuales hemos mantenido los que, en general, pueden modificar los muebles: *eléctrico*, *grande*, *pequeño*, *caro*, *barato*, *limpio*, *moderno*, etc. Además, hemos encontrado y conservado unas palabras que explican el material del que se hacen los muebles: *vidrio*, *madera*, *hierro*, *bronce*, etc. Lo mismo sucede con otras palabras que hablan del mantenimiento de los muebles: *reparar*, *mantener*, *arreglado*, etc.

La decisión más difícil, en lo que se refiere a conservar o no ciertas respuestas, se ha planteado al tratar aquellas relaciones asociativas que se realizan en este centro de interés, pero que, en principio, han de tener lugar en *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios*. Estamos otra vez ante la cuestión que hemos planteado en el área temática anterior. Los informantes, sin saber que en adelante se les solicitaría palabras de otras áreas y no conociendo la demarcación exacta de los muebles, escribieron en la lista del presente centro de interés palabras como *vaso*, *copa*, *botella*, *taza*, *cuchillo*, *cuchara*, *olla* y *sartén*. No hemos podido asegurar que estas respuestas constituyan relaciones asociativas no debidas y, finalmente, las hemos incluido en la lista de *Los muebles de la casa*, aunque podrán reaparecer en los otros dos centros de interés mencionados. La razón la

explicamos a continuación. Junto con dichas respuestas que plantean problemas, también han aparecido otras palabras como *bombilla*, *espejo*, *mapa*, *foto(grafía)*, *colchón* y *sábana*. Estas no presentan problemas si se incluyen en *Los muebles de la casa*, porque no existe otro centro de interés a que, en principio, hayan de pertenecer. Tanto aquellas primeras como estas segundas respuestas son objetos movibles de la casa y no manifiestan diferencias a este respecto. Por eso, con respecto al campo léxico *Los muebles de la casa*, consideramos inadecuado eliminar las primeras y conservar las segundas solo porque las primeras pertenezcan a otras áreas temáticas. Como resultado, la decisión nuestra ha sido conservar ambos grupos de palabras sin discriminación, toda vez que las primeras que plantean problemas no son muchas.

4.5.2.5. Alimentos y bebidas

En este centro de interés hemos singularizado muchas palabras que se usan habitualmente en plural, como *tapas*, *fideos*, *tallarines*, *macarrones*, *calamares*, *verduras* y *cereales*¹⁰⁹. En cuanto a la unificación de las variantes gráficas, tenemos el ejemplo de *raviole* - *ravioli* y hemos optado por el segundo; en cambio, hemos mantenido tanto *pimiento* como *pimienta* por su diferencia semántica. En comparación con otras áreas temáticas, hay en esta más diminutivos y aumentativos que hemos conservado por su significado distinto de la palabra no derivada: *tortilla*, *burrito*, *panecillo*, *natilla*, *cochinillo*, *camarón*, *salchichón*, etc.

En muchas otras investigaciones de disponibilidad léxica, se ha llamado la atención sobre la aparición de gran cantidad de marcas comerciales en este campo léxico. Pero, en nuestras listas, solo hemos detectado los siguientes nombres: *coca cola*, *pepsi cola*, *fanta*, *sprite*, *KFC* y *McDonald's*, los cuales hemos conservado sin discriminación. Cabe mencionar que, a estos nombres,

¹⁰⁹ Dentro de estas palabras, nos encontramos con unas que el DLE indica que se usan más en plural —*fideos*, *tallarines*, *macarrones*—. No obstante, nos decidimos por la singularización de ellas porque dicho diccionario, a pesar de todo, presenta un lema en singular.

todavía se suma *Maotai*, que es una marca conocida de licor chino. Además, hemos encontrado otras respuestas que raras veces se ven en las demás sintopías: *pastel de luna*, *tienda de conveniencia* y *raíz de loto*. Son compuestos sintagmáticos poco frecuentes en el español, pero tienen la estructura sintáctica correcta y hacen referencia a objetos muy habituales en el país de origen de los informantes, siendo calco o traducción literal de la lengua china. La primera se trata de un dulce tradicional que el pueblo chino come en la Fiesta del Medio Otoño, festividad que se celebra el 15 del octavo mes del calendario lunar, especialmente en la noche de luna llena. La segunda se refiere a las pequeñas tiendas donde ofrecen todo tipo de productos usuales —incluidos alimentos y bebidas preparados o envasados— y que se abren las 24 horas para facilitar la compra de la vecindad. La última, como su nombre indica, consiste en la raíz de dicha planta, que se consume en China como verdura.

El mayor problema que hemos hallado en este centro de interés es la aparición de numerosos sintagmas nominales cuyo núcleo es un sustantivo que también aparece solo en las listas. Estamos hablando de unidades léxicas como *agua - agua mineral - agua con gas*, *vino - vino tinto - vino blanco - vino rosado*, *café - café con leche - café solo*, *patata - patata frita - patata brava*, etc. En principio, podríamos unir los sintagmas nominales con los sustantivos independientes, conservando así solo los sustantivos, pero, claro está, la unificación no resultaría del todo justificable. Revisando las encuestas, nos hemos dado cuenta de que, además de los citados compuestos sintagmáticos, es común que los estudiantes escribieran nombres concretos de alimentos o bebidas. Por ejemplo, podemos encontrar *pollo asado*, *pescado frito*, *merluza a la romana*, *arroz a la cubana*, *gazpacho*, *consomé*, *sopa de marisco*, *sangría*, *zumos de naranja*, *té rojo*, *fanta limón*, *macedonia*, *tiramisú*, *arroz con leche*, etc. Estando así las cosas, suponemos que estas palabras del campo léxico sobre alimentos y bebidas las adquieren nuestros informantes con un objetivo preciso: utilizarlas en situaciones reales como en bares o restaurantes para poder pedir exactamente lo que quieren consumir. Es decir, si lo que quieren es, por ejemplo, vino rosado o merluza a la romana, no basta con decir simplemente

vino o merluza; tienen que expresarse lo más explícitamente posible. En otras palabras, estamos frente a unas respuestas que son las realmente actualizadas en situaciones comunicativas dadas, tal y como se define la disponibilidad léxica. Por lo tanto, hemos decidido mantener todas las respuestas que consisten en nombres concretos de alimento o bebida, incluidas las arriba citadas¹¹⁰.

Por último, como amplitud de relaciones asociativas, hemos aceptado otras respuestas que, de seguro, tienen que ver con los alimentos y bebidas. Citamos como ejemplo los verbos *tomar, cocinar, asar, hervir, freír, desayunar, pedir, invitar* y *vivir*; los adjetivos *tierno, jugoso, sabroso, rico, dulce, picante, agrio, amargo, frito, caliente* y *frío*; los sustantivos *cocina, restaurante, almuerzo, comedor, plato, carta, hambre* y *sed*; y el sintagma *todos los días*.

4.5.2.6. Objetos colocados en la mesa para la comida

En este campo asociativo, igual que en los centros de interés anteriores, la variación de número se da mayoritariamente en las palabras que aluden a objetos que aparecen más de uno en la mesa, como *platos, palillos, tazas, cuchillos, cucharas, tenedores, cubiertos, vasos, copas* y *utensilios*, cuya singularización ha sido definitiva. En cuanto a los derivados, hemos detectado y mantenido los diminutivos *platillo* y *cucharilla de postre* y los aumentativos *platón*¹¹¹, *cucharón* y *tazón*.

El tratamiento sobre las relaciones asociativas extendidas es el mayor problema que hemos de resolver en este centro de interés. En síntesis, hemos

¹¹⁰ Esta es una cuestión que podría convertirse en un criterio general para la edición de palabras en un corpus de ELE en el centro de interés *Alimentos y bebidas*, dado lo que hemos explicado sobre los sintagmas nominales que aluden a nombres concretos de alimentos o bebidas.

¹¹¹ Recordamos que *platón* es un americanismo. Aunque este término no es usual en el español de España, es una palabra que se usa en varios países de Hispanoamérica y que los informantes han podido tener contacto con ella a través de las lecturas o del profesorado, dado que no es infrecuente la presencia, en las universidades chinas, de profesores nativos hispanoamericanos que, naturalmente, utilizan su variedad de español en el aula y fuera de ella.

conservado los siguientes tipos de asociaciones ampliadas. En primer lugar, aparecen bastantes nombres de alimentos y bebidas que suelen haber en la mesa a la hora de comer o de tomar algo. Nos hemos dado cuenta de que los informantes querían referirse a los diferentes recipientes en los que se colocan los alimentos y bebidas, pero aludieron a estos mismos como sustitutos de ellos porque no conocían las palabras españolas que designan tales recipientes. Esto se debe a que, en la lengua china, los sustantivos que hacen referencia a los recipientes combinan el nombre del objeto que estos contienen con la palabra que indica la categoría de ellos mismos. Por ejemplo, al indicar una *vinagrera* en la lengua materna, los estudiantes dicen algo así como *botella* o *frasco de vinagre*, que constituye una sola palabra en chino. Al haber escrito ya *botella* o *frasco* en líneas anteriores, los informantes se deciden por *vinagre* pensando en *botella* o *frasco de vinagre*. Lo mismo sucede con muchos otros sustantivos para que no tuvieran que repetir *botella de esto* o *botella de lo otro*. Respuestas de este tipo las tenemos en *vinagre, sal, pimienta, salsa, queso, agua, leche, sopa, café, té, vino, mermelada, limón, pan, arroz*¹¹², etc.

En segundo lugar, hemos detectado algunas palabras que hacen referencia a objetos que sí existen sobre la mesa, pero que no son estrictamente para la propia comida. Por ejemplo, hemos incorporado *flor, florero, vela, bombilla, luz, trapo, mugre, grasa*, etc. En tercer lugar, aparecen sustantivos que se refieren a las personas sentadas a la mesa para la comida. En estos casos, hemos admitido las palabras relativas a comensales y de significado común como *persona, gente y hombre*, y otras que constituyen razones por las que la gente se sienta a la mesa para comer juntos como *invitado y amigo*; en cambio, hemos suprimido las palabras que aluden a relaciones personales más concretas como *hijo, madre y esposo*.

En cuarto lugar, hemos notado muchas alusiones a los materiales de los que se hacen los objetos que hay sobre la mesa. Por ejemplo, en las listas

¹¹² En la tradición china, el arroz en las comidas desempeña el mismo papel que el pan en Occidente. Es decir, si tenemos en cuenta su presencia constante en la mesa de los chinos, ya no se trata de un plato preparado sino de una especie de acompañante a los guisos.

finales se han figurado *material, metal, plástico, cristal, tela, vidrio, porcelana, madera, hierro*, etc. En quinto lugar, no han faltado adjetivos que describen la cualidad de los objetos para la comida, como *limpio, mojado, liso, eléctrico, sucio, mugriento, grasoso*, etc. Todos ellos los hemos mantenido. Finalmente, también hemos incorporado algunos pocos verbos que se relacionan con las acciones en una comida: *comer, beber, invitar, servir y charlar*.

En lo que se refiere a los sintagmas, se han registrado en esta área temática una cantidad limitada. Hemos incluido en el corpus de disponibilidad los siguientes compuestos sintagmáticos: *plato hondo, ropa de mesa, rollo de papel, olla a presión, cucharilla de postre y palillo de dientes*. Cabe mencionar que, a estos, también se ha añadido *servilleta de papel*, puesto que un mismo informante aportó tanto *servilleta* como *servilleta de papel*. Suponemos que dicho alumno entiende *servilleta* por la de tela, lo que puede generalizarse en los demás estudiantes. Por eso, creemos conveniente distinguir las dos entradas, admitiéndolas ambas.

4.5.2.7. La cocina y sus utensilios

Este campo asociativo no presenta grandes problemas de unificación. Las cuestiones que hemos de tratar en él son muy similares a las del área temática anterior. Por lo próximos que están *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios* en lo que se refiere a las relaciones asociativas —puesto que la mesa para la comida puede estar en la propia cocina o muy cerca de ella—, los dos centros de interés tienen que compartir bastantes vocablos que hemos considerado oportuno incorporar en ambos, dado el criterio general sobre la amplitud de relaciones asociativas aplicado en todos los campos léxicos.

Hemos unificado las formas plurales y singulares de los nombres de utensilios que se ven en varias unidades en la cocina: *platos, ollas, sartenes, utensilios, aparatos, botellas, cajas, guantes, servilletas, cuchillos, cucharas, tenedores, tazas*, etc. Hemos recurrido al paréntesis para unir *campana* y

campana extractora en *campana (extractora)*. En cuanto a los nombres de aparatos que pueden terminar tanto en *-dor* como en *-dora*, hemos seguido los criterios en los centros anteriores manteniendo la pareja *tostador - tostadora*, mientras que hemos incorporado solamente *secador* y *aspiradora* por la ausencia de *secadora* y *aspirador* en las listas de la presente área temática.

Las respuestas asociativas de este centro de interés presentan también una amplitud de relaciones. Con respecto a los nombres de alimentos y bebidas, solo hemos conservado aquellos que se suelen guardar en la cocina, sean ingredientes y condimentos o materias primas y productos para la cocina: *aceite, harina, ajo, sal, vinagre, azúcar, canela, carne, pescado, verdura, patata, maíz, huevo, arroz, pan, jamón*, etc. Además, hemos incorporado otros nombres relativos a cosas que jamás faltan en la cocina: *agua, gas, gas natural, fuego, llama, luz, electricidad, vapor, humo, ceniza, calor, basura*, etc. Los verbos son otro tipo de amplitud de relaciones asociativas, de los cuales hemos incluido todos aquellos que implican relaciones con los variados aparatos y acciones realizadas en la cocina: *cortar, calentar, freír, hervir, picar, cocinar, enchufar, encender, apagar, lavar, limpiar, arreglar, barrer*, etc. Igual que en los demás centros de interés, todavía hemos encontrado y mantenido gran cantidad de adjetivos que describen, desde diferentes perspectivas, los utensilios y alimentos guardados en la cocina: *eléctrico, usado, nuevo, ordenado, arreglado, desarreglado, sucio, mojado, frito, cocido, congelado*, etc. Los sustantivos que explican el material de los utensilios de la cocina son otra magnitud de respuestas que hemos incorporado: *plástico, cerámica, porcelana, metal, cristal, madera*, etc.

En lo referente a los sintagmas que hemos incluido en nuestras listas, cabe mencionar que hemos mantenido *fuego lento*, porque se trata de una una de las modalidades de cocción o freidura importante en la gastronomía china; hemos admitido *máquina de arroz*, porque es sintácticamente correcto y hace referencia a la “olla arrocera” que está presente en la cocina de casi todas las casas chinas; y hemos incorporado *marmita al vacío*, tal y como escribieron los informantes, aunque parece más adecuado decir “marmita de cocción al vacío”.

4.5.2.8. La escuela: sus muebles y materiales

En comparación con los estudios del *Proyecto Panhispánico*, este centro de interés en nuestra investigación presenta una diferencia marcada. A la hora de hacer las pruebas asociativas, nuestros informantes entendieron la temática no solo en relación a la escuela sino también a la universidad, ya que eran todos estudiantes universitarios. De modo que las respuestas de este campo léxico recabadas en la presente tesis parten de los contornos tanto de la escuela como de la universidad. Para empezar, ha habido en esta área temática menos casos de unificación con respecto a la variación de número y de género. Con relación a la unificación de los acortamientos con sus formas plenas, hemos usado el paréntesis en cuatro palabras: *bolí(grafo)*, *tele(visión)*, *bici(cleta)* y *foto(grafía)*. En cuanto a la variación gráfica, hemos conservado la pareja *computador - computadora*. Además, en este centro de interés se han registrado más siglas que en los anteriores: hemos mantenido *CD*, *PPT*, *PDF* y *USB*.

Igual que en otras sintopías, es destacable la amplitud de relaciones asociativas que se da en esta área temática. Además de los propios muebles y materiales de la escuela o universidad, hemos encontrado y mantenido respuestas relativas a las dependencias de la escuela o del campus universitario. Por ejemplo, hemos conservado *instituto*, *biblioteca*, *comedor*, *gimnasio*, *estadio*, *cafetería*, *restaurante*, *jardín*, *residencia*, *colegio mayor*, *copistería*, *librería*, *laboratorio*, *supermercado*, *tienda*, *campo de deportes*, *lago*, *museo*¹¹³, *hospital*¹¹⁴, etc. e, incluso, otras respuestas que aluden a cosas que

¹¹³ En una de las universidades de la que proceden nuestros informantes, se encuentra un museo de historia natural dentro del campus.

¹¹⁴ Es común que las universidades chinas cuenten con una sección de servicio sanitario para atender específicamente a su profesorado y alumnado. Además, una de las universidades estudiadas en esta tesis dispone de una escuela de medicina, la que, por su parte, tiene un hospital público universitario. Cabe destacar, además, que los campus universitarios chinos son verdaderas ciudades en pequeño, puesto que todos los estudiantes que no son de la localidad en la que se encuentra la universidad, que son la mayoría, viven en alojamientos dentro del campus, al igual que el profesorado extranjero. El campus se autoabastece, y tiene todo tipo de tiendas, además de restaurantes o instalaciones deportivas.

existen en dichos sitios como *oficina, despacho, libro, revista, árbol, habitación, dormitorio, sala de lectura, salón de actos, conferencia, televisor, teléfono, piscina*¹¹⁵, etc.

Por otra parte, las alusiones a las personas relacionadas con la escuela o universidad es un tema problemático. En muchos de los estudios del *Proyecto Panhispánico*, han rechazado los términos relativos a elementos humanos, a pesar de que algunos alcanzan altos índices de frecuencia. Consideramos, por nuestra parte, que es inadecuado eliminarlos teniendo en cuenta sus numerosas apariciones y elevado grado de disponibilidad, a lo cual se añade el hecho de que sus relaciones asociativas con el epígrafe del centro son del todo comprensibles. Dado que nuestros objetivos centrados en los alumnos de ELE no son exactamente los mismos que los de las investigaciones del *Proyecto Panhispánico* y que también existen trabajos en los que se han admitido los elementos humanos —por ejemplo, en Gómez Devís (2004), Serrano Zapata (2014) y Sánchez-Sauz Laserna (2016)—, hemos decidido mantener palabras como *profesor, maestro, bibliotecario, estudiante, alumno, compañero, rector* e, incluso, *decano, colega, jefe, amigo, novio*¹¹⁶, *abuelo*¹¹⁷, *médico*¹¹⁸, etc. De igual modo, hemos incorporado los nombres relativos a las variadas asignaturas como *español, inglés, chino, lengua, fonética, gramática, comprensión auditiva, expresión oral, conversación, redacción, retórica, literatura, filología, asignatura*, etc., los cuales ponen en evidencia el perfil de nuestros informantes como estudiantes de lengua extranjera.

Otro tipo de asociaciones de gran peso son los diversos verbos que entran en relación con las diferentes actividades que se realizan en el aula o fuera de

¹¹⁵ Dos de las universidades donde nuestros informantes estudian tienen una piscina en el gimnasio ubicado en su campus.

¹¹⁶ Es corriente que los estudiantes empiecen una relación amorosa en la universidad, dado que la mayoría se pasa toda la semana en el campus (día y noche), cuando no, todo el mes o todo el lapso que hay entre periodo vacacional y periodo vacacional, de modo que el contacto entre las personas se convierte fácilmente en estrecho.

¹¹⁷ La inclusión de esta palabra se debe a que los alumnos les llaman cariñosamente “abuelo” o “abuela” a sus profesores veteranos de edad mayor. Se trata de una diferencia cultural.

¹¹⁸ Se trata del personal de la sección de servicio sanitario que hay en todas las universidades.

las clases. Por ejemplo, figuran en nuestras listas finales palabras como *tomar apuntes, escribir, borrar, teclear, escuchar, preguntar, recordar, memorizar, asistir, descansar, jugar, aprobar, suspender, representar, imprimir, graduar, etc.* En fin, creemos que el presente centro de interés se caracteriza por una extensión notable de las relaciones asociativas porque lo que aportan los informantes son palabras que reflejan los contornos próximos a ellos mismos. Por lo tanto, hemos aceptado la mayoría de las respuestas que hemos visto justificables según la propia experiencia del autor de esta tesis como antiguo estudiante de ELE en China.

4.5.2.9. Iluminación y calefacción

Este centro de interés no presenta grandes problemas de edición. Hemos tenido que eliminar el artículo que va acompañado, con alta frecuencia, de los siguientes sustantivos: *el sol, la luna y la luz*. Con respecto a la corrección gráfica que hemos procedido en esta área temática, llaman la atención las innumerables variantes erróneas de *aire acondicionado* generadas por la supuesta deficiencia en el aprendizaje, las cuales hemos corregido y unificado bajo la forma correcta. Como ejemplo, citamos *aire condición, aire condicionado, aire condicionar, aire acondicionar, aire acondicionador, aire acondicional, aire acondicionación, aire condonante, aire condinación, aerocondición, etc.*

Las relaciones asociativas que se dan en este campo léxico son muy variadas. Nos hemos dado cuenta de que, aunque el epígrafe en cuestión habla solamente de la iluminación y la calefacción, nuestros informantes han realizado asociaciones también en torno a la ventilación y refrigeración, es decir, lo opuesto a la calefacción. Considerando que, en muchos de los estudios del *Proyecto Panhispánico*, los investigadores han ampliado el marco referencial de esta área temática añadiendo “los medios para airear un recinto”, hemos decidido conservar las respuestas que tratan sobre ventilación y refrigeración. En concreto, por un lado, hemos incorporado, sin duda alguna, los medios artificiales: *lámpara, calefacción, fuego, vela, linterna, bombilla, ventilador, etc.* Por otro lado, hemos hallado y mantenido las palabras relativas

a medios naturales: *sol, luna, estrella, aire, viento, nieve, llover*, etc., a pesar de que estas últimas respuestas no están incluidas en las otras investigaciones de disponibilidad léxica.

No obstante, debido a la pobreza léxica de nuestros informantes con respecto a la temática del presente centro de interés, se escribieron muchas unidades léxicas que se caracterizan por la contigüidad y extensión del tema. En primer lugar, hemos incorporado las palabras de ropa y sus accesorios, así como de sus materiales, con los cuales uno se calienta o refresca: *ropa, abrigo, chaqueta, guante, bufanda, manta, jersey, sombrero, lana, algodón*, etc. En segundo lugar, hemos aceptado las respuestas que se refieren a medios de uso particular y de eficacia limitada: *sudar, frotar, correr, aliento, abrazo, agua caliente, termo, chocolate, helado, móvil*¹¹⁹, etc. En tercer lugar, se han registrado respuestas, referentes a fechas determinadas o al tiempo y clima, que giran sobre la necesidad de iluminación o cambio de temperatura: *día, noche, primavera, verano, otoño, invierno, navidad, noviembre, ahora*¹²⁰, *tiempo, clima*, etc., las cuales hemos admitido igualmente.

En cuarto lugar, hemos incluido también las palabras que indican sitios o recintos donde hay gran demanda de iluminación, calefacción o ventilación: *hogar, casa, dormitorio, habitación, (cuarto de) baño, universidad, colegio, instituto, biblioteca, hospital, restaurante, cocina, tren, metro*, etc. En quinto lugar, hemos dado cabida a respuestas de diferente índole que hacen referencia o bien a los métodos de iluminación, calefacción y ventilación, o bien al entorno o ambiente donde se satisfacen o no tales necesidades: *eléctrico, sistema electrónico, tecnología, moderno, cómodo, confortable, agradable, espléndido, deslumbrante, fuerte, débil, claro, claridad, oscuro, oscuridad, humoso, ventoso, hacer buen día, sin luz*, etc. En sexto lugar, hemos

¹¹⁹ Los alumnos se refieren a dar iluminación con la linterna del teléfono móvil.

¹²⁰ Se admite *ahora* porque el informante se refirió a que las fechas de la prueba coincidieron con el momento en que se inició la calefacción colectiva. Fue en noviembre cuando aplicamos entre los estudiantes la prueba asociativa del vocabulario, mes del año en que empieza a funcionar el sistema de calefacción colectiva en todo el municipio de Tianjin.

considerado conveniente aceptar las respuestas que implican la instalación de los aparatos de iluminación, calefacción o ventilación y el espacio donde se instalan: *tubería, enchufe, válvula, conducto, cable, circuito, pila, techo, pared, suelo, instalación*, etc. En séptimo lugar, hemos admitido palabras que se relacionan con el consumo de energía por los aparatos: *recurso*¹²¹, *petróleo, leña, energía solar, gasto, gastar, cobrar, tarifa, precio, factura*, etc.

Además, hemos incluido en las listas finales tres respuestas sobre la ubicación geográfica: *norte, sur y noreste de China*. Las primeras dos tienen gran sentido porque, sin hallarse ni *este ni oeste* en las listas, denotan los distintos medios y hábitos de calefacción y ventilación generados por la geografía del país, puesto que el norte y el sur se diferencian en muchos detalles al respecto. Mientras tanto, *noreste de China* fue aportado por un estudiante procedente del noreste del país, zona que se conoce como la más fría de todo el territorio y que goza de la prolongación más larga de calefacción colectiva¹²² en el ámbito nacional. Por último, cabe mencionar que hemos admitido el apellido del inventor estadounidense a quien se debe la lámpara de incandescencia: *Edison*.

4.5.2.10. La ciudad

Este campo léxico no se destaca por las tareas de neutralización, unificación ni corrección. Plantea pocas dudas a ese respecto, una vez que hemos seguido los criterios generales ya comentados. Solo hay que mencionar que hemos empleado el paréntesis para la unificación gráfica en *bici(cleta)*. Sin embargo, el epígrafe de este centro de interés conduce, de por sí, a todas las posibilidades de relaciones asociativas, puesto que no parece tener límites demarcados. Por lo tanto, hemos optado por mantener el mayor número de

¹²¹ Se refiere a los recursos naturales que el hombre aprovecha como fuente de energía.

¹²² En España, la calefacción colectiva es la que se tiene en algunos edificios con un único sistema de calefacción para todo el edificio. En cambio, en las ciudades del norte de China, un sistema parecido funciona de igual manera, pero conecta todos los edificios de la ciudad entera y se gestiona por el gobierno de los municipios. Cada municipio puede determinar las fechas en que empieza y deja de funcionar, dependiendo del clima de su zona.

respuestas posibles, lo que se traduce en una mayor riqueza léxica que ya ha sido comprobada en casi todos los estudios de disponibilidad léxica.

En general, hemos detectado y conservado numerosas palabras que designan distintas realidades que existen en las ciudades. Por ejemplo, 1) hay entidades físicas privadas o públicas como *empresa, universidad, hospital, escuela, tienda, supermercado, biblioteca, restaurante, colegio, mercado, hotel, cafetería, bar, etc.*; 2) hay construcciones e infraestructuras como *edificio, calle, metro, casa, parque, tren, tráfico, camino, zona verde, plaza, jardín, estación, aeropuerto, transporte, carretera, autopista, etc.*; 3) hay poderes y administraciones públicas como *policía, comisaría, alcalde, gobierno, país, habitante, ciudadano, ayuntamiento, pueblo, capital, población, zona, barrio, distrito, región, provincia, municipal, etc.*; 4) hay tipos de actividades humanas como *economía, cultura, sociedad, comunidad, comercio, negocio, industria, turismo, política, religión, etc.*; 5) hay paisaje y naturaleza como *árbol, flor, hierba, hoja caída, río, aire, mar, cielo, bosque, lago, campo, playa, pájaro, lluvia, costa, isla, etc.*; 6) hay cualidades y problemas propios de las ciudades como *desarrollo, tecnología, urbanización, internacional, conveniente, medio ambiente, calidad del aire, contaminación, basura, ruido, bullicio, atasco, congestión, hora punta, prisa, soledad, populoso, ritmo rápido, ocupado, cansado, atareado, alto precio, caro, lujoso, próspero, etc.*; y, además, una amplia gama de asociaciones secundarias concretas o abstractas generadas a partir de dichas realidades. Finalmente, añadimos que, como casos peculiares, hemos mantenido el nombre del sistema de pago por móvil y por internet: *Alipay*; y la emblemática noria situada en el centro del municipio de Tianjin: *Ojo de Tianjin*.

A diferencia de muchos estudios del *Proyecto Panhispánico* que prescinden de los nombres propios de ciudades, hemos optado, en esta investigación, por mantenerlos, siguiendo la decisión de Sánchez-Saus Laserna (2016). La razón por la que hemos incluido los nombres de ciudades es la misma que la de incorporar los nombres de marcas o establecimientos comerciales, tal y como se ha mencionado en los criterios generales. En

síntesis, la disponibilidad entre los informantes de los nombres de ciudades en lengua extranjera es la disponibilidad de algunas normas fonológicas y morfológicas de las palabras, cosa muy significativa para los alumnos en proceso de adquisición de la lengua. Y esos nombres reflejan, igual que los demás vocablos disponibles, el conocimiento de las palabras o, incluso, la competencia léxica o geográfica de los estudiantes.

Como resultado, hemos admitido todos los nombres de ciudades en español y, asimismo, todos aquellos (no muchos) que han actualizado en inglés. Pero, con respecto a los nombres de ciudades en chino transcritos al alfabeto latino, solo hemos mantenido los de frecuencia alta, que resultan ser tres: *Beijing, Tianjin y Shanghai*. Como caso particular, hemos incluido también el nombre de la población desaparecida: *Numancia*, porque los estudiantes conocieron en el aula la ciudad heroica y la famosa resistencia numantina.

4.5.2.11. El campo

En este centro de interés, hemos tratado numerosos casos de singularización en los sustantivos referentes a seres o cosas que se caracterizan por su cantidad múltiple debido a las circunstancias propias del campo o de la zona rural. Citamos como ejemplos *hijos, hermanos, niños, ancianos, abuelos, animales, gallinas, ovejas, huertos, cultivos, productos, montañas, ríos, senderos, etc.* El plural en *hijos* y *hermanos* se refiere a que las familias de las zonas rurales suelen ser familias de más de un hijo. Las respuestas *niños, ancianos* y *abuelos* en plural hace referencia a un montón de niños y ancianos enfrentados al problema social de que los padres de las familias rurales tienen que abandonar la casa para ir a buscar trabajo a la ciudad, dejando así solos a los menores o mayores que difícilmente se valen por sí mismos. A pesar de ello, la unificación bajo las formas singulares ha sido definitiva en las listas finales.

Con respecto a la riqueza léxica de esta área temática, coincidimos con los demás estudios de disponibilidad léxica, al afirmar que este es otro centro

de interés que, después de *La ciudad*, suscita una mayor cantidad de asociaciones subjetivamente ampliadas. Además, en este centro de interés figuran bastantes respuestas que también aparecen en el centro anterior, tal y como han advertido Samper Padilla (1998), Samper Hernández (2001) y Gómez Devís (2004). Estas repuestas las hemos respetado basándonos en el criterio de mantener las relaciones asociativas siempre que sean justificables. Por lo general, hemos detectado y conservado los siguientes tipos de evocaciones léxicas.

En primer lugar, tenemos las palabras relativas al paisaje del campo como *río, bosque, árbol, flor, aire, cielo, lago, verde, paisaje, bonito, naturaleza, estrella, sol, viento, fuente, arroyo, monte, hermoso, maravilla*, etc. En segundo lugar, hemos mantenido las diversas especies de animal cuya aparición en el campo no es extraña, aunque se les solicita palabras de animales en un centro posterior: *perro, cerdo, vaca, caballo, gato, pollo, gallo, oveja, pájaro, pez, ave, gallina, conejo, ratón, toro, buey*, etc. En tercer lugar, hemos dado cabida a respuestas relacionadas con las actividades de la agricultura y de la ganadería como *verdura, arroz, trigo, huerto, huerta, tierra, cultivar, maíz, fruta, fresco, cereal, cosecha, ganado, agrícola, granja, camión, leche, plantar, carne, huevo, alimento, harina, patata, hortaliza*, etc.

Luego, también hemos registrado y conservado las palabras relativas a las diferentes facetas de la vida rural como *tranquilo, cómodo, divertido, contento, simple, libre, interesante, alegría, gustar, encantar, silencioso, feliz, longevo, difícil, pobre*, etc. Además, hemos incorporado bastantes sintagmas que tratan de explicar detalladamente las situaciones naturales o sociales características del campo. Hemos mantenido la forma plena de dichos compuestos sintagmático, porque tal característica pierde una vez eliminado algún elemento de la unidad entera. Nos referimos a sintagmas como *cielo azul, aire fresco, tierra fértil, calidad de vida, animal salvaje, animal doméstico, camino fangoso, camino áspero, gente simpática, gente amable, piel morena, mano seca, falta de educación, falta de dinero, salario bajo, menos desarrollado*, etc.

4.5.2.12. Medios de transporte

En este centro de interés, igual que en los demás, hemos recurrido al paréntesis en el caso de *bici(cleta)*, para unificar el acortamiento y la forma plena. Pero esto no lo hemos aplicado a *moto - motocicleta*, sino que hemos mantenido ambas respuestas porque existen informantes que aportaron ambas palabras en la lista de este centro. Luego, tal y como hemos explicado en los criterios generales, hemos conservado como palabras independientes *auto - autobús - automóvil* y, en cambio, *bus* como palabra del inglés. En lo que se refiere a la unificación del género, hemos respetado la aparición por separado de *motor - motora* y de *barco - barca*, ya que cada palabra hace referencia a una realidad diferente. Con relación a los derivados diminutivos, hemos mantenido *carrito*, dado que es recogido por el diccionario DLE como entrada independiente.

Frente al tratamiento sobre *pie - a pie*, pareja que unos estudios de disponibilidad léxica optan por unificar bajo *pie* y que otros prefieren distinguir entre sí, nuestra decisión ha sido esto último. Es decir, hemos mantenido ambas respuestas, principalmente debido a que también figuran en las listas *de pie* y *en pie* como amplitud de relaciones asociativas. Con el objetivo de diferenciar las tres locuciones adverbiales, hemos optado por respetar la aparición de la preposición en *a pie*, *de pie* y *en pie*. Lo mismo hemos hecho en *a bordo*, cuya preposición hemos conservado. No obstante, en cuanto a otras locuciones construidas con preposición más sustantivo como *a caballo*, al no encontrarlas en las listas, no ha habido problemas para unificarlas bajo el sustantivo correspondiente..

En cuanto a los nombres de marcas comerciales, siguiendo el criterio aplicado en los demás centros de interés, hemos incorporado en las listas finales, por un lado, cinco palabras que designan compañías de transporte: *Renfe*, *BlaBlaCar*, *Uber*, *Mobike* y *ofo*, todas con la primera letra en mayúscula excepto la última, porque la propia marca se escribe en minúscula y los

informantes también aportaron la palabra así. Por otro lado, hemos dado cabida a cuatro términos referentes a marcas de coche: *Ferrari, Tesla, Audi* y *BMW*. Con respecto a las siglas, hemos mantenido, sin duda alguna, *AVE*. Y, además, hemos encontrado y conservado *CIF* y *FOB*, dos términos de comercio internacional que los estudiantes adquirieron en la asignatura del español de los negocios.

Como excepción, hemos permitido en esta área temática una voz que el diccionario no recoge: *tricicleta*. Evidentemente, los informantes crearon la palabra a partir de la estructura interna de “bicicleta” y querían aludir al “triciclo”. Hemos creído conveniente mantenerla porque tiene una frecuencia de aparición bastante alta, mientras su referente “triciclo” no cuenta con ni una aparición en ninguna de las listas. A esto se añade que, en algunos estudios de disponibilidad léxica, también se recogen creaciones de los informantes que alcanzan alta frecuencia de aparición y que, por consiguiente, se consideran como innovaciones lingüísticas (Serrano Zapata, 2014: 268).

Con respecto a las relaciones asociativas ampliadas, hemos hallado las referentes a motivos u objetivos del uso de los medios de transporte: *ir, salir, visitar, despedir, regresar, trabajar, comprar, llevar, llegar, carga, estiba, encuentro, trabajo, casa, escuela, etc.*; las relativas a diversas cualidades de los medios de transporte: *rápido, rapidez, velocidad, fácil, conveniente, ventaja, desventaja, seguro, seguridad, peligroso, peligro, orden, urbano, a tiempo, ganar tiempo, etc.*; las relacionadas con los efectos del transporte: *contaminación, polución, humo, ruido, hora punta, atasco, embotellamiento, retraso, etc.*; y gran cantidad de asociaciones secundarias a partir de los servicios de transporte.

4.5.2.13. Trabajos del campo y del jardín

Al igual que en los demás campos léxicos, en el presente, la singularización de los sustantivos en plural la hemos hecho en todos los casos, aunque se trate de objetos que implican una pluralidad lógica como *tijeras* y

guantes, que hemos insertado como *tijera* y *guante*. Con respecto a la variación de género, hemos mantenido las parejas *fruto - fruta* y *huerto - huerta*, por la alteración en la semántica. También hemos conservado, como se ha comentado en los criterios generales, la pareja *hierba - yerba*, a pesar de la posibilidad de unificación. Cabe mencionar que este es el centro de interés en que se recoge la mayor cantidad de verbos, puesto que el título de la temática insinúa acciones que se efectúan a la hora de trabajar en el campo o el jardín. A todos los verbos registrados, sean pronominales o estén conjugados, hemos unificado bajo sus formas infinitivas.

Aparte de los verbos, hemos mantenido bastantes sintagmas verbales, aunque los mismos verbos de los sintagmas ya estén incluidos en las listas finales. Esto se debe, por un lado, a que de verdad designan acciones características que se llevan a cabo en el campo y el jardín, cuyo significado los verbos solos no llegan a especificar; y, por otro lado, a que los compuestos sintagmáticos hablan del conocimiento sintáctico de que disponen nuestros informantes. Siguiendo los criterios generales ya comentados, hemos incorporado, por ejemplo, *trabajar la tierra, labrar la tierra, remover la tierra, arrancar hierbas, quitar hierbas, matar los insectos, abrigar los árboles, plantar árboles, cortar las ramas, aplicar abono, poner agua, regar flores, recoger frutas, alimentar animales*, etc.

Además de los verbos y sintagmas verbales, entre las relaciones asociativas más frecuentes que se detectan en esta área temática, hemos hallado 1) utensilios, herramientas y maquinaria en los trabajos del campo y del jardín como *herramienta, tijera, azada, tractor, guante, arado, coche, piedra, hoz, máquina, grúa, herbicida, insecticida, pesticida, mano, hacha, rastro, clavo*, etc.; 2) distintos lugares en los que o en cuyos alrededores se realizan los trabajos: *campo, jardín, bosque, huerto, huerta, patio, terraza, río, parque, césped, granja, parcela*, etc.; 3) objetos o productos que constituyen el fin de los trabajos como *flor, árbol, hierba, semilla, agua, tierra, arroz, planta, fruta, verdura, cosecha, maíz, cultivo, insecto*, etc.; 4) profesionales que realizan los trabajos o tipos de personas que suelen hacer algunas pequeñas faenas en su

tiempo de ocio: *campesino, jardinero, granjero, florista, agricultor, labrador, trabajador, pastor, obrero, basurero, albañil, mano de obra, hombre, anciano, abuelo, padre, madre, etc.*; 5) animales que ayudan en los trabajos o que se relacionan con estos: *perro, burro, caballo, buey, toro, vaca, oveja, pez, cerdo, mariposa, rana, ratón, bicho, etc.*; 6) características de las faenas sea como profesión o como afición: *cansado, cansar, cansancio, ocupado, agotado, esforzar, rendimiento, pesado, difícil, sudor, músculo, atezar, daño, fuerte, polvo, sucio, suciedad, barro, fango, duro, belleza, bonito, contento, feliz, alegría, etc.*

4.5.2.14. Los animales

La unificación de las variaciones bajo la forma no marcada se ha aplicado a todos los términos registrados en este centro de interés, con excepción de los heterónimos o de aquellos sustantivos que tienen sufijos distintos para el masculino y el femenino. Por ejemplo, la inserción por separado se da en las parejas *caballo - yegua, toro - vaca y gallo - gallina*. La misma unificación la hemos aplicado a los derivados diminutivos que los informantes emplean con frecuencia como expresión de afecto hacia los animales que crían como mascota: *perrito, pajarito, conejito, pollito, cerdito, etc.*, los cuales hemos uniformado bajo las palabras de las que se derivan.

Cabe señalar que, siguiendo nuestro criterio de mantener las palabras del inglés, hemos hallado una respuesta de *dragon* sin tilde y la hemos aceptado tal como la aportó el estudiante, porque dio respuestas en inglés antes y después de *dragon* y, así, suponemos que esta también es una palabra inglesa que no debe llevar tilde. Además, frente a la aparición tanto de *coala* como de *koala*, hemos optado por esta última forma recogida por el diccionario DLE, tal y como se ha explicado en los criterios generales.

Como se ha demostrado en otras investigaciones, este centro de interés, por el carácter concreto de su referencia semántica, plantea muchos menos problemas con respecto a la amplitud de relaciones asociativas. Aparte de los

nombres de animales, que son respuestas realmente esperables y que suponen la mayoría de las palabras recabadas en esta área temática, hemos encontrado otros tipos de asociaciones lógicas. 1) Hemos conservado las respuestas que designan las partes del cuerpo animal como *pata, cuerpo, piel, mano, pierna, cuerno, lomo, pelo*, etc.; 2) hemos admitido las palabras relacionadas con los productos de origen animal como *carne, pescado, huevo, chuleta, jamón, leche, sabroso, alimento, comida*, etc.; 3) igualmente están incluidas en las listas finales las palabras que denotan los rasgos y atributos de diversos animales: *feroz, carnicero, peligroso, horrible, fuerte, perezoso, ágil, cuco, rápido, leal, cariño, bonito, adorable, simpático, amigo, compañero*, etc.; 4) también hemos incorporado otras respuestas que implican las acciones propias de los animales o las que el ser humano toma sobre ellos: *reproducción, vivir, correr, volar, comer, picar, cazar, alimentar, proteger, adoptar*, etc.; 5) por último, se han aceptado los referentes a los espacios de libre movimiento o lugares de residencia de los animales: *tierra, cielo, mar, océano, río, agua, zoo, reserva, árbol, campo, montaña, bosque, jardín, parque, casa, naturaleza, África*, etc.

4.5.2.15. Juegos y distracciones

En este centro de interés hemos utilizado el paréntesis para la unificación de los acortamientos con sus formas plenas en los casos de *tele(visión)* y *bici(cleta)*. Siguiendo el criterio general de unificación de la ortografía de los extranjerismos, hemos conservado en el corpus de disponibilidad los que siguen: *mah-jong, kung-fu, karaoke, taekwondo, taichí* y *kárate*. De ahí se observa la peculiaridad de nuestros informantes, ya que todos estos préstamos son de origen asiático.

En cuanto a las abreviaciones, hemos aceptado la abreviatura *app*, que se refiere a la aplicación informática para teléfonos móviles inteligentes; y el acrónimo de origen inglés *KTV*, que hace referencia al tipo de entretenimiento musical interactivo llamado en inglés “Karaoke television”. Además, hemos incluido en las listas finales el nombre de una aplicación móvil que los

estudiantes emplean diariamente con alta frecuencia: *WeChat*, que es el equivalente chino del *Whatsapp*. Por otra parte, aparecen en este centro de interés varios nombres propios de juegos o sus siglas o acrónimos, los hemos incorporado tal como hemos admitido los nombres de las marcas comerciales. Se han registrado y mantenido los siguientes: *LoL*¹²³, *WoW*¹²⁴, *Dota*¹²⁵, *UNO*¹²⁶, *NFS*¹²⁷, *Super Mario*¹²⁸ y *Xbox*¹²⁹.

Al igual que en otras sintopías, el mayor problema que presenta esta área temática es el referido a los compuestos sintagmáticos. En nuestro caso, llama la atención la aparición de gran cantidad de sintagmas verbales. Como hemos comentado en los criterios generales, los hemos admitido en la medida de lo posible. Así pues, los sintagmas verbales aceptados en esta investigación tienen las siguientes características. En primer lugar, hemos mantenido las locuciones como *tomar el sol* y las construcciones idiomáticas formadas por “ir de”: *ir de compras*, *ir de vacaciones* e *ir de excursión*.

En segundo lugar, hemos mantenido diversos sintagmas verbales en los que la combinación del verbo y el sustantivo es del todo esperable, a pesar de que el sustantivo también aparece en las listas¹³⁰: *jugar al fútbol*, *jugar a las cartas*, *ir al gimnasio*, *tocar la guitarra*, *ver películas*, *leer libros*, *hacer deporte*, etc. Es decir, hemos separado en dos entradas distintas dichas respuestas y los simples sustantivos *fútbol*, *carta*, *gimnasio*, *guitarra*, *película*, *libro* y *deporte*, respectivamente, con el fin de que nuestras listas finales sean capaces de dar información sobre la disponibilidad no solo de los sustantivos, sino también de

¹²³ Se trata del videojuego de ordenador con el nombre completo de “League of Legends”.

¹²⁴ Se refiere al videojuego de ordenador llamado “World of Warcraft”.

¹²⁵ Representa una serie de videojuegos que comenzó con el lanzamiento de “Defense of the Ancients”.

¹²⁶ Es el nombre de un juego de cartas estadounidense.

¹²⁷ Es la sigla de “Need for Speed”, que consiste en una serie de aproximadamente 17 videojuegos de carreras de coches.

¹²⁸ Es el nombre de una serie de videojuegos de plataformas creados por la empresa desarrolladora Nintendo.

¹²⁹ Es cómo se llama la videoconsola de sobremesa producida por Microsoft.

¹³⁰ Para el tratamiento de estos casos, la mayoría de los estudios del *Proyecto Panhispánico* prefiere unificar dichos sintagmas verbales bajo los sustantivos que aparecen en ellos mismos.

los verbos y de la construcción correcta de las combinaciones sintagmáticas —esta última denota la competencia lingüística relativa a la sintaxis—.

En tercer lugar, hemos encontrado algunas combinaciones de verbo y complemento como amplitud de relaciones asociativas: *perder peso, subir la montaña, saltar a la comba, comprar por internet, jugar entre árboles, charlar con amigos, salir con amigos, comer con la familia, matar el hambre, salir de noche*, etc. Estas respuestas respetan las normas sintácticas y son construcciones absolutamente aceptables.

Por último, hay que mencionar que, como relaciones asociativas ampliadas, hemos incorporado unas palabras que tienen que ver con qué piensan los informantes de los juegos y distracciones, o cómo se sienten con ellas. Citamos como ejemplos *cansado, ocupado, alegría, alegre, feliz, divertido, importante, aburrido, aburrimiento, soledad, gustar, disfrutar, encantar, amar, obsesionar*, etc.

4.5.2.16. Profesiones y oficios

Las labores de edición en este campo asociativo se centran, por un lado, en la singularización de los sustantivos en plural y, por otro, en la unificación también de los sustantivos bajo la forma no marcada del masculino. La primera no plantea problema alguno, mientras que la segunda, sí. En primer lugar, dos palabras en forma femenina —*actriz* y *reina*— que tienen un sufijo distinto de su equivalente masculino las hemos mantenido tal como se escriben. En segundo lugar, en cuanto a la inserción de *ama de casa*, hemos respetado su forma femenina, porque el papel predominante de la mujer en esta actividad es del todo comprensible y, además, no se registra en nuestras listas ni un caso de su forma masculina. En tercer lugar, frente a las respuestas como *azafata, enfermera, secretaria* y *niñera*, estas las hemos unificado bajo su forma masculina, a pesar de que contienen las mismas implicaciones sociolingüísticas que *ama de casa*. La decisión parte del hecho de que la forma masculina de dichas respuestas también aparece en nuestras encuestas, de

modo que nos parece conveniente la unificación del género al igual que en otros sustantivos que no son de oficios.

En lo referente a la unificación de las variantes gráficas, se halla un caso único en este centro de interés, que consiste en la pareja *chofer* - *chófer*. Hemos mantenido esta última con tilde, puesto que es aportada por más informantes y, así, aparece con más frecuencia en las encuestas. Con respecto a las abreviaciones, hemos incluido la abreviatura *Dr.*, que representa el título de doctor. De igual modo hemos conservado algunas siglas de origen inglés: *DJ*, *CEO* y *COO*, que representan, respectivamente, “disc jockey”, “chief executive officer” y “chief operating officer”. Además, también hemos mantenido la sigla *MD*, de cuyo referente no estamos muy seguros, pero, como va después de *CEO* y *COO* en la lista de palabras, suponemos que se refiere a “managing director”.

Tal y como se ha señalado en otras investigaciones, en sus listas y también en las nuestras, están los nombres de los profesionales y trabajadores, de las actividades profesionales y de las carreras universitarias que dan acceso a una profesión. Sin embargo, a pesar de la concreción que presenta esta área temática, nuestros informantes aportaron, además, 1) los diversos sectores profesionales como *industria*, *gastronomía*, *ferrocarril*, *moda*, *arte*, *marketing*, *economía*, *educación*, *negocio*, *seguro*, etc.; 2) los lugares de trabajo como *restaurante*, *universidad*, *biblioteca*, *cafetería*, *oficina*, *banco*, *farmacia*, *empresa*, *tienda*, *hotel*, *fábrica*, etc.; 3) las acciones o actividades concretas que se realizan en el trabajo: *servir*, *cocinar*, *entrevista*, *bailar*, *traducir*, *enseñar*, *escribir*, *leer*, *colectar*, *editar*, *comentario*, *reportaje*, *poesía*, *novela*, *baile*, *canto*, *construcción*, etc.; 4) algunas otras unidades léxicas relativas a los trabajos o al mercado laboral: *trabajo*, *labor*, *admisión*, *convenio*, *contrato*, *obligación*, *ocupado*, *encargado*, *responsabilidad*, *competir*, *ambición*, *sin trabajo*, *en paro*, etc. Estas respuestas las hemos admitido como amplitud de relaciones asociativas.

5. ANÁLISIS CUANTITATIVO DE LOS DATOS

Este capítulo es donde la léxico-estadística funciona como un soporte técnico para la presente investigación de disponibilidad léxica. Las fórmulas matemáticas y el programa informático de los que hemos hablado en los capítulos anteriores (véase los apartados 3.2.2.1.3 y 4.3.2) permiten que extraigamos resultados cuantitativos a partir de los datos editados y codificados, resultados que la labor humana no alcanza a procesar. Se calculan y muestran, en adelante, los resultados cuantitativos más valorados y tratados en los estudios de disponibilidad. Concretamente, son, entre otros índices, el número de palabras, el número de vocablos, el promedio de palabras por informante, el número de vocablos con índice de disponibilidad superior a 0.02 y 0.1, el número de vocablos que suman una frecuencia acumulada del 75%, el número de vocablos mencionados por más del 25% de los informantes, la densidad léxica y el índice de cohesión.

El objetivo de este capítulo es, por un lado, presentar e interpretar dichos resultados cuantitativos extraídos de las dos muestras constituidas por estudiantes de Tianjin —una, del léxico en español como lengua extranjera (ELE) y la otra, del léxico en la lengua materna (CLM) de los informantes—, y, por otro lado, comparar cuantitativamente los resultados en ELE y en CLM, con el fin de observar cómo va la adquisición del léxico español, tomando como referencia la competencia léxica que los alumnos tienen de su lengua materna. Puesto que las respuestas en ambas muestras se organizan en torno a los mismos campos semánticos, la diferencia cuantitativa que se da entre las relaciones asociativas en ELE y en CLM es capaz de predecir la deficiencia por parte de los informantes en la adquisición del español.

5.1. Resultados cuantitativos generales

5.1.1. Índices globales en ELE y en CLM

Para empezar, creemos oportuno repetir la distinción entre *palabra* y

vocablo, de los que hemos ido hablando a lo largo de los capítulos anteriores. Esa distinción corresponde a la que se da entre *token* y *type*, dos conceptos que tienen amplia aplicabilidad en la lingüística. Los tokens se refieren a todos los casos individuales y los types, en cambio, hacen referencia a los tipos de casos. De ahí que, en los estudios de disponibilidad léxica, tenemos, por un lado, todas las ocurrencias de unidades léxicas que se registran en las listas de respuestas de los informantes; y, por otro lado, los tipos de unidad léxica diferentes. Es a las ocurrencias léxicas a las que se llama *palabras*, mientras que a los tipos léxicos se les denomina *palabras diferentes* o, simplemente, *vocablos*. Así pues, el *número de palabras*, que resulta de la suma de las actualizaciones de unidades léxicas, es capaz de evaluar la productividad léxica de los informantes, puesto que mide la cantidad total de palabras que ellos han producido en los cuestionarios. Por otra parte, el *número de vocablos*, al medir la cantidad de tipos de unidad léxica diferentes, hace saber la riqueza léxica, o lo que es lo mismo, la variedad léxica del corpus entero o bien de determinados centros de interés o bien de determinados tipos de informantes.

Tal y como hemos hablado en el capítulo correspondiente a la metodología, los datos obtenidos de las diversas sesiones de encuesta en ELE y en CLM permiten que se confronten los resultados cuantitativos de las dos muestras. En este apartado, nos vamos a fijar en los diferentes índices cuantitativos globales de cada una de las dos muestras en cuestión.

5.1.1.1. Lo primero que vamos a observar consiste en los índices cuantitativos en el conjunto de las encuestas en ELE y en CLM con respecto a las palabras.

	ELE	CLM
Número de informantes	288	172
Total de palabras	44525	44193
Promedio de palabras por centro de interés	2782.81	2762.06
Promedio de palabras por informante	154.60	256.94
Promedio de palabras por informante y por centro de interés	9.66	16.06

Tabla 5.1. Índices cuantitativos globales en ELE y en CLM con respecto a las palabras

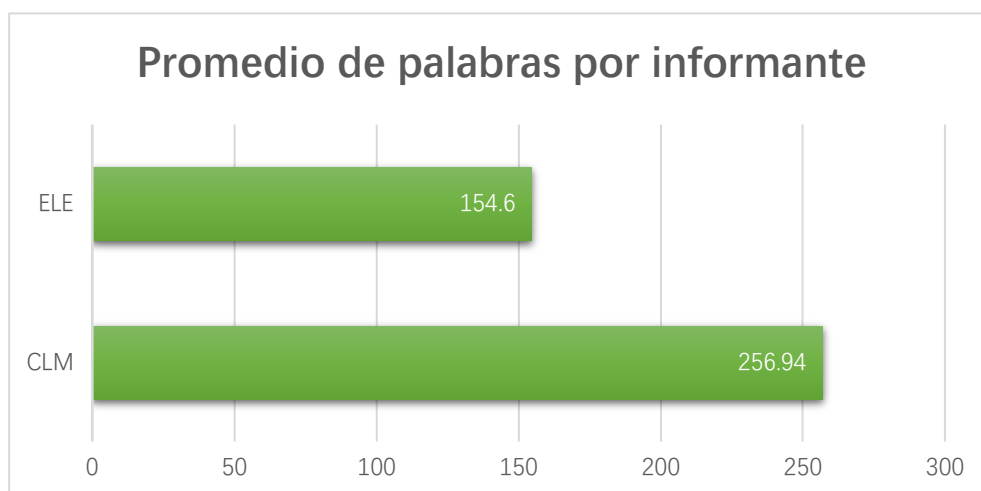


Gráfico 5.1. Promedio de palabras por informante en ELE y en CLM

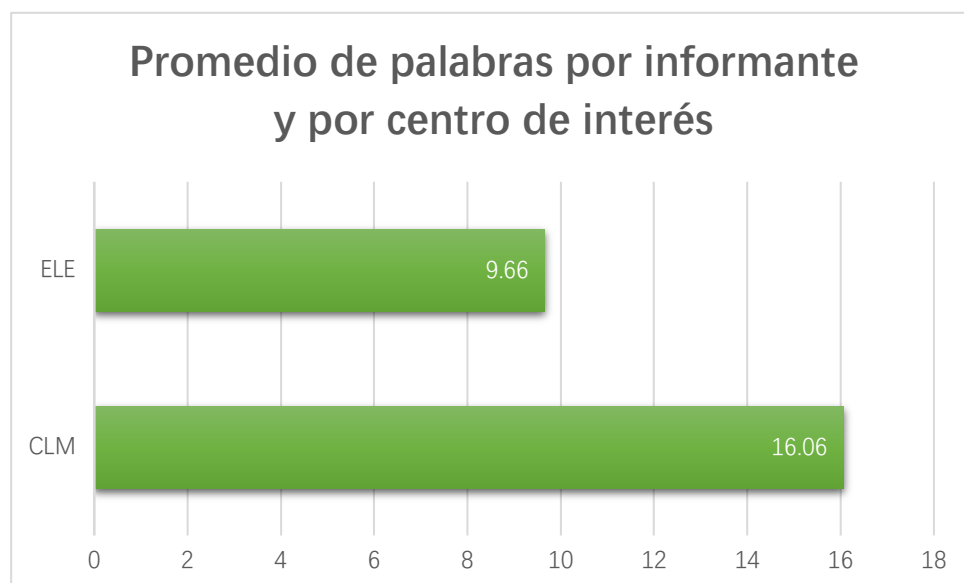


Gráfico 5.2. Promedio de palabras por informante y por centro de interés en ELE y en CLM

Como podemos ver en la Tabla 5.1, el total de palabras (44525, en ELE, y 44191, en CLM) resulta ser la suma de todas las respuestas proporcionadas por los informantes en los dieciséis centros de interés, y por eso es el número que indica el volumen del corpus de disponibilidad léxica. Está claro que el total de palabras viene profundamente afectado por el número de informantes (288, en ELE, y 172, en CLM) que forman parte de la muestra, toda vez que toma en cuenta cualquier respuesta de cada encuestado. Así pues, dado que el número de informantes en la muestra de ELE (288) es mayor de aquel en la de CLM (172) —el primero supone el 167.44% del segundo—, lo que se observa es que el corpus en ELE recoge más palabras (44525) que el corpus en CLM (44191), en concreto, 334 más. Esta diferencia en la productividad léxica repercute, del mismo modo, en el promedio de palabras por centro de interés (2782.81, en ELE, frente a 2762.06, en CLM), que resulta ser de 20.75 palabras más por área temática en la muestra en español. Este índice es la media de respuestas en cada centro dadas por todos los informantes. Se ven, a su vez, dos promedios, de ELE y de CLM, de valor semejante, a pesar de la diferencia que se da en el número de informantes. Esto implica que los dos corpus léxicos tienen un volumen real más o menos parecido, pero proporcionalmente los estudiantes disponen de más productividad léxica en el conjunto de las encuestas en CLM que en la totalidad de encuestas en ELE.

Nos aproximamos a la competencia léxica del estudiante medio de Tianjin, si pasamos a observar los últimos dos tipos de promedio expuestos en la Tabla 5.1; puesto que son dos medias que eliminan la influencia provocada por el número total de informantes. Por un lado, contamos con el promedio de palabras por informante, que significa el número de repuestas que el alumno medio produce en todos los centros de interés, o lo que es lo mismo, en todas las listas de palabras establecidas en el cuestionario entero. Así pues, el Gráfico 5.1 informa de la diferencia entre la muestra en ELE y la correspondiente en CLM al respecto. Como podemos ver, el alumno medio de las encuestas en ELE aporta 154,6 respuestas al completar el cuestionario, y el estudiante medio en CLM, en cambio, da 256,94 respuestas. A partir de ello, llegamos a

la conclusión de que los informantes tienen menor productividad léxica en ELE que en CLM y, además, la producción léxica de los alumnos en ELE supone el 60.17% de la que tienen en CLM.

Por otro lado, contamos con el promedio de palabras por informante y por centro de interés. Es el indicador que aproxima a los investigadores a la competencia léxica que el estudiante medio tiene en el campo semántico medio. A lo comentado también se añade otra ventaja de este índice, y es que, al ser dividido entre los informantes y entre los centros, llega a presentarse en un valor numérico pequeño (9.66 en ELE y 16.06 en CLM, por ejemplo, en la tabla 5.1) que es más fácil de entender e interpretar. Tal y como manifiesta el Gráfico 5.2, la producción léxica por informante y por centro de interés es de 9.66 palabras en las encuestas de ELE y 16.06, en las de CLM, como acabamos de comentar. Así es que la producción en ELE representa un poco más de la mitad de la producción en CLM. Se nota una diferencia de 6.4 palabras, es decir, los alumnos escriben, en promedio, 6.4 palabras menos en un centro de interés de la encuesta en ELE que en uno de la encuesta en CLM. No podemos más que afirmar que, en comparación con la competencia léxica que los estudiantes tienen en su lengua materna, se evidencia una deficiencia en la productividad léxica en español como lengua extranjera.

5.1.1.2. A continuación, vamos a ver los resultados cuantitativos sobre los vocablos en el conjunto de las encuestas en ELE y en CLM.

	ELE	CLM
Número de informantes	288	172
Total de vocablos	5650	6145
Promedio de vocablos por centro de interés	353.13	384.06
Porcentaje de vocablos sobre palabras	12.69%	13.90%

Tabla 5.2. Índices cuantitativos globales en ELE y en CLM con respecto a los vocablos

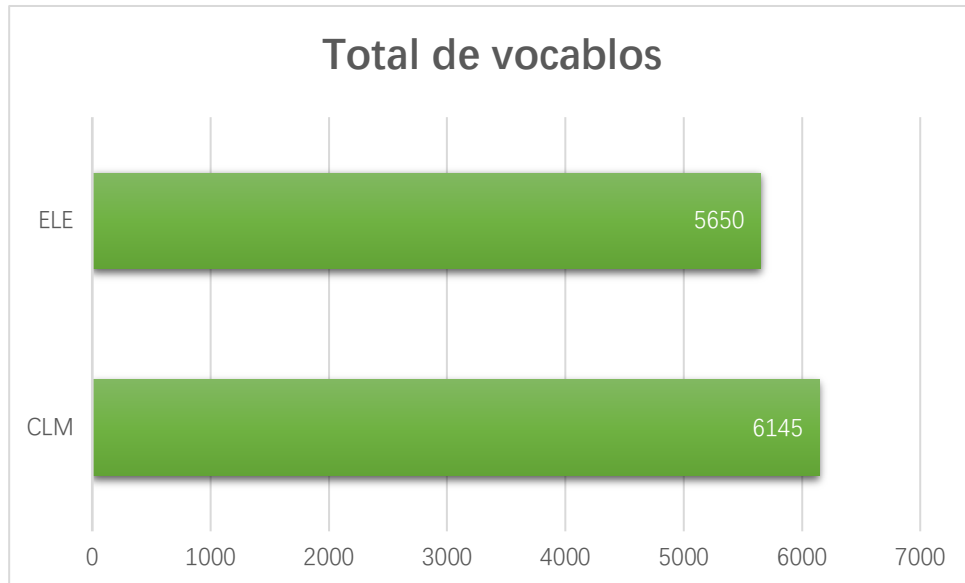


Gráfico 5.3. Total de vocablos en ELE y en CLM

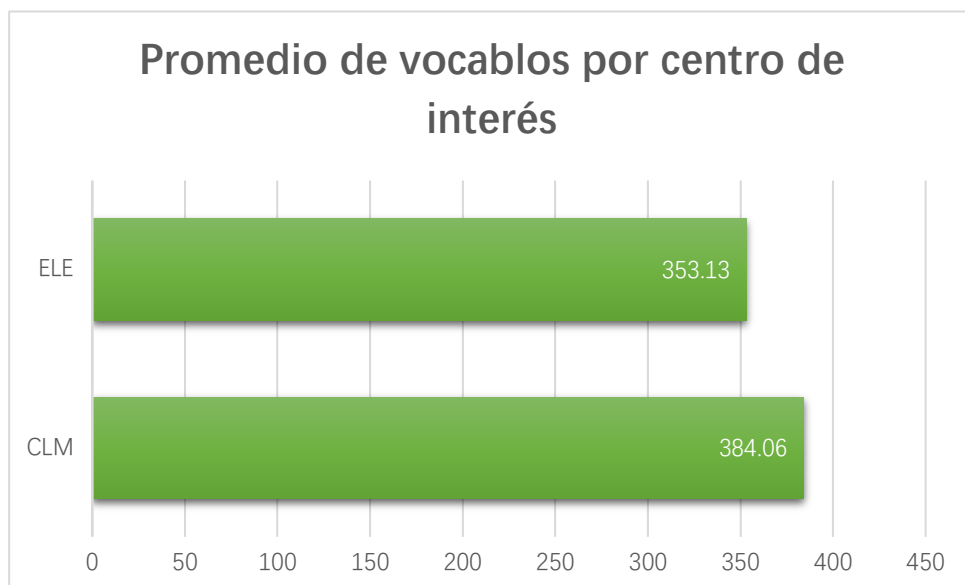


Gráfico 5.4. Promedio de vocablos por centro de interés en ELE y en CLM

Con respecto a los vocablos, los índices globales que se exponen en la Tabla 5.2 parecen señalar también la deficiencia leve de los datos obtenidos de las encuestas en ELE, en comparación con los obtenidos de las encuestas en CLM. Como podemos ver en la tabla, el total de vocablos en los dos corpus (5650 en ELE y 6145 en CLM) es la suma del número de palabras diferentes recogidas bajo cada uno de los dieciséis centros de interés, respectivamente; en consecuencia, el promedio de vocablos por centro de interés (353.13, en

ELE, y 384.06, en CLM) es el número medio de vocablos registrados por centro.

Cabe destacar que el número de vocablos de cada centro no tiene que ver necesariamente con el número de informantes, ni con la producción léxica de los mismos. Dicho de otra manera, no se establece la siguiente correlación: cuantos más informantes haya o cuanto mayor producción léxica se genere, más vocablos se dan. En realidad, el número de vocablos tiene que ver con el comportamiento de los informantes. Si, bajo un campo semántico, un estudiante escribe, por diversas causas, unas palabras muy distintas y alejadas de las aportadas por sus compañeros, contribuye al aumento del número de vocablos. De todos modos, los índices con respecto a los vocablos implican la riqueza o –lo que viene a ser lo mismo– la variedad léxica de un grupo de estudiantes ante cierto estímulo temático.

Así pues, el Gráfico 5.3 confronta el volúmen de los vocablos recabados de las encuestas en ELE (5650) y en CLM (6145), de lo que llegamos a afirmar que la cantidad de vocablos en ELE supone el 91.94% de la obtenida en CLM. Mediante el promedio de vocablos por centro de interés que muestra el Gráfico 5.4, podemos ver con mayor claridad la diferencia entre las dos lenguas en la riqueza léxica del campo semántico medio (353.13 en ELE y 384.06 en CLM). Se observa así una diferencia de 30.93 vocablos. Esto quiere decir que los alumnos escriben, en promedio, 30.93 palabras diferentes menos en un centro de interés de la encuesta en ELE que en uno de la encuesta en CLM. Como resultado, se observa una deficiencia, aunque leve, en la riqueza léxica en español como lengua extranjera por parte de los estudiantes, en comparación con la competencia léxica que tienen en su lengua materna.

Por otra parte, la Tabla 5.2 también informa del porcentaje de vocablos sobre todas las palabras (12.69%, en ELE, frente a 13.90%, en CLM), que resulta de dividir el total de vocablos entre el total de palabras. Dicho porcentaje se inspira en el *type-token ratio* (*TTR*) empleado en la lingüística computacional, puesto que la distinción entre *vocablo* y *palabra* es la misma que se da entre *type* y *token* (véase el comentario al inicio del apartado 5.1.1).

En nuestro caso, el porcentaje de vocablos sobre palabras indica cuántos vocablos se registran en cada cien palabras, sobre la base del conjunto de las encuestas en ELE y en CLM, respectivamente. Así pues, podemos tener una idea de la diferencia entre las dos lenguas, eliminando la influencia provocada por el número distinto de informantes. Llegamos a saber que, en las encuestas de ELE, los estudiantes aportan 12.69 vocablos cada vez que escriben cien palabras; mientras, en las encuestas en CLM, esta cifra es de 13.90 vocablos. Esto no contradice la conclusión anteriormente extraída sobre la deficiencia leve en ELE, sino que la ratifica, dado que la diferencia entre una lengua y otra es de 1.21 vocablos.

5.1.1.3. Vamos a analizar, de aquí en adelante, algunos otros índices acerca de los vocablos. Debido a que el número de vocablos tiene que ver con el comportamiento de los informantes —tal como se ha explicado más arriba—, parece que, al realizar estudios comparativos entre distintos conjuntos, no tiene sentido comparar aquellos vocablos que son aludidos por pocos informantes o, incluso, que son actualizados por un solo sujeto atípico. Por tanto, ante la gran cantidad de vocablos generados por un número reducido de informantes, se formula la pregunta de qué y cuántos vocablos se deben tener en cuenta para comparar el léxico disponible de dos o más conjuntos.

Las medidas de corte en el cotejo se basan normalmente en la frecuencia acumulada o en el índice de disponibilidad. La frecuencia acumulada es la suma de la frecuencia relativa de un vocablo más las frecuencias relativas de todos los vocablos que lo preceden en la lista de disponibilidad. La frecuencia relativa, a su vez, resulta de dividir el número de veces que un vocablo aparece en un centro de interés entre el número total de palabras del mismo centro. Por lo tanto, mientras la frecuencia relativa señala la aparición de cada vocablo entre el total de ocurrencias, la frecuencia acumulada denota la aparición de todos los vocablos desde el primero hasta el considerado —según su orden en la lista de disponibilidad— entre el total de ocurrencias. Por otra parte, el índice de disponibilidad mide, como su nombre indica, el grado de disponibilidad de cada vocablo para el informante, teniendo en cuenta sobre todo la frecuencia

de aparición del vocablo y su posición en las listas de la prueba asociativa (véase el apartado 3.2.2.1.3 para más información sobre su cálculo).

Así pues, por un lado, Samper Padilla (1999) sitúa el límite de los vocablos para el cotejo en el 75% de la frecuencia acumulada en cada centro de interés; y, por otro lado, Bartol Hernández (2001) propone que se comparen los vocablos que tienen un índice de disponibilidad igual o mayor que 0.02 en cada centro de interés. Es decir, en la actualidad, existen dos criterios sobre qué y cuántos vocablos de las listas de disponibilidad se deben usar para hacer las comparaciones entre distintos conjuntos: 1) los vocablos que tienen una frecuencia acumulada hasta el 75% y 2) los vocablos que tienen un índice de disponibilidad superior a 0.02. Presentamos, a continuación, los resultados correspondientes a ambos criterios que se obtienen de nuestros corpus de ELE y de CLM.

Además, a estos dos criterios, creemos oportuno añadir resultados de otros dos índices sobre los vocablos: 1) los vocablos que tienen un índice de disponibilidad superior a 0.1 y 2) los vocablos que son mencionados por más del 30% de los informantes. Los dos son índices que Carcedo González (2001) considera para seleccionar, respectivamente, los vocablos altamente disponibles y los altamente frecuentes, y son ampliamente empleados en las investigaciones de disponibilidad léxica. Cabe señalar que, debido al reducido número detectado en nuestros informantes de ELE con respecto al segundo índice, optamos por seguir a Gómez Devís (2004), analizando los vocablos mencionados por más del 25% de los informantes y no del 30% comentado.

Así pues, la Tabla 5.3 expone el número de vocablos de los dieciséis centros de interés de las encuestas en ELE y en CLM, en relación con los tres tipos arriba comentados: 1) el número de los vocablos tenidos en cuenta para hacer comparaciones entre distintos conjuntos (los que suman una frecuencia acumulada del 75% y los que cuentan con un I.D. superior a 0.02); 2) el número de vocablos altamente disponibles (los que poseen un I.D. superior a 0.1); y 3) el número de vocablos altamente frecuentes (los mencionados por más del 25%

de los informantes).

	ELE	CLM
Número de vocablos que suman una frecuencia acumulada del 75%	850	1203
Número de vocablos con I.D. > 0.02	791	1162
Número de vocablos con I.D. > 0.1	223	341
Número de vocablos mencionados por más del 25% de los informantes	139	242

Tabla 5.3. Índices cuantitativos globales en ELE y en CLM con respecto a la frecuencia acumulada, el índice de disponibilidad y la aparición de los vocablos

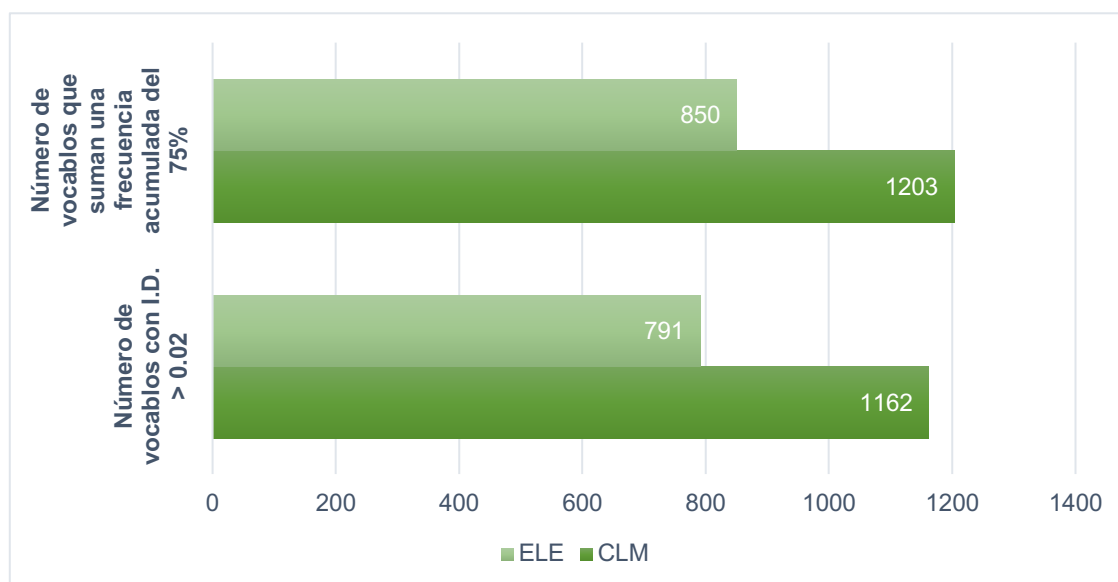


Gráfico 5.5. Número de vocablos que suman una frecuencia acumulada del 75% y número de vocablos con índice de disponibilidad superior a 0.02 en ELE y en CLM



Gráfico 5.6. Número de vocablos con índice de disponibilidad superior a 0.1 y número de vocablos mencionados por más del 25% de los informantes en ELE y en CLM

El Gráfico 5.5 manifiesta las diferencias en número de vocablos entre las dos muestras al establecerse las dos medidas de corte. Se observa que, al poner el límite de vocablos bien en el 75% de la frecuencia acumulada, bien en el 0.02 del I.D., se demuestra la inferioridad estadística que se da en el corpus de ELE en comparación con el de CLM, sea cuál sea la medida de corte. Además, en ambos criterios, las diferencias entre el número en ELE y el obtenido en CLM se presentan de modo similar. Concretamente, según el criterio de la frecuencia acumulada, el número de vocablos en ELE (850) supone el 70.66% del número de vocablos en CLM (1203), con una diferencia de 352 vocablos. Mientras tanto, según el criterio del I.D., dicho porcentaje es de 68.07% (791 en ELE frente a 1162 en CLM) y se nota una diferencia de 371 vocablos. Vemos que una medida de corte no dista mucho de la otra: las dos señalan porcentajes semejantes de vocablos en ELE sobre vocablos en CLM (70.06% y 68.07%, respectivamente) y diferencias parecidas de número de vocablos (353 y 371, respectivamente), y, por eso, ponen en evidencia la deficiencia de los estudiantes en los vocablos en ELE en comparación con sus vocablos en CLM.

El Gráfico 5.6 muestra, por una parte, la diferencia entre los dos corpus con respecto al número de vocablos con I.D. superior a 0.1 (223, en ELE, y 341, en CLM). Vemos que, en el ámbito de los vocablos altamente disponibles, los alumnos dominan 118 vocablos menos en ELE que en CLM, o sea, manejan en ELE el 65.40% del número de vocablos altamente disponibles en CLM. El mismo gráfico, por otra parte, informa de la diferencia en el número de vocablos mencionados por más del 25% de los informantes (139, en ELE, y 242, en CLM). Llegamos a saber que los estudiantes aportan 103 vocablos altamente frecuentes menos en ELE que en CLM. La cantidad de los vocablos altamente frecuentes en ELE supone el 57.44% de esta cantidad en CLM.

5.1.2. Índices por centro de interés en ELE y en CLM

En los siguientes apartados vamos a analizar por centro de interés los resultados cuantitativos obtenidos de las encuestas de ELE y de CLM. Además de comparar, al igual que en los apartados anteriores, los datos confrontados según la lengua de la encuesta, observaremos también las diferencias estadísticas entre los centros de interés, dado que estos, de por sí, constituyen variables lingüísticas, según las que covaría la producción léxica de los informantes. Está claro que, del mismo modo que las variables extralingüísticas¹³¹ influyen en los resultados de disponibilidad, los centros de interés implican campos semánticos que determinan las relaciones asociativas que los estudiantes son capaces de actualizar. Por eso, el análisis sobre las diferencias entre los centros nos permite ver la competencia lingüística de los alumnos ante distintas áreas nocionales.

5.1.2.1. Número de palabras

A continuación, vamos a ver los datos cuantitativos relativos a las palabras en cada centro de interés. La Tabla 5.4 muestra el número de palabras de cada

¹³¹ Se puede consultar el apartado 4.4 para ver qué variables extralingüísticas hemos tenido en cuenta en esta tesis. Sus influencias en la disponibilidad léxica las vamos a analizar en adelante.

centro en los corpus en ELE y en CLM. Los Gráficos 5.7 y 5.8, que figuran después de la misma, hacen saber de forma ilustrativa el volumen de las palabras que tiene cada campo léxico en relación con la producción total.

Centro de interés	Total de palabras	
	ELE	CLM
01 Partes del cuerpo humano	3496	3300
02 Ropa y complementos	2482	2974
03 Partes de la casa (sin muebles)	2415	2238
04 Los muebles de la casa	2570	2628
05 Alimentos y bebidas	4355	3555
06 Objetos colocados en la mesa para la comida	1740	1956
07 La cocina y sus utensilios	1906	2455
08 La escuela: sus muebles y materiales	3204	2877
09 Iluminación y calefacción	1613	2105
10 La ciudad	3938	3024
11 El campo	2904	2841
12 Medios de transporte	2596	2472
13 Trabajos del campo y del jardín	2201	2108
14 Los animales	3036	3765
15 Juegos y distracciones	2735	2760
16 Profesiones y oficios	3334	3135

Tabla 5.4. Total de palabras por centro de interés en ELE y en CLM

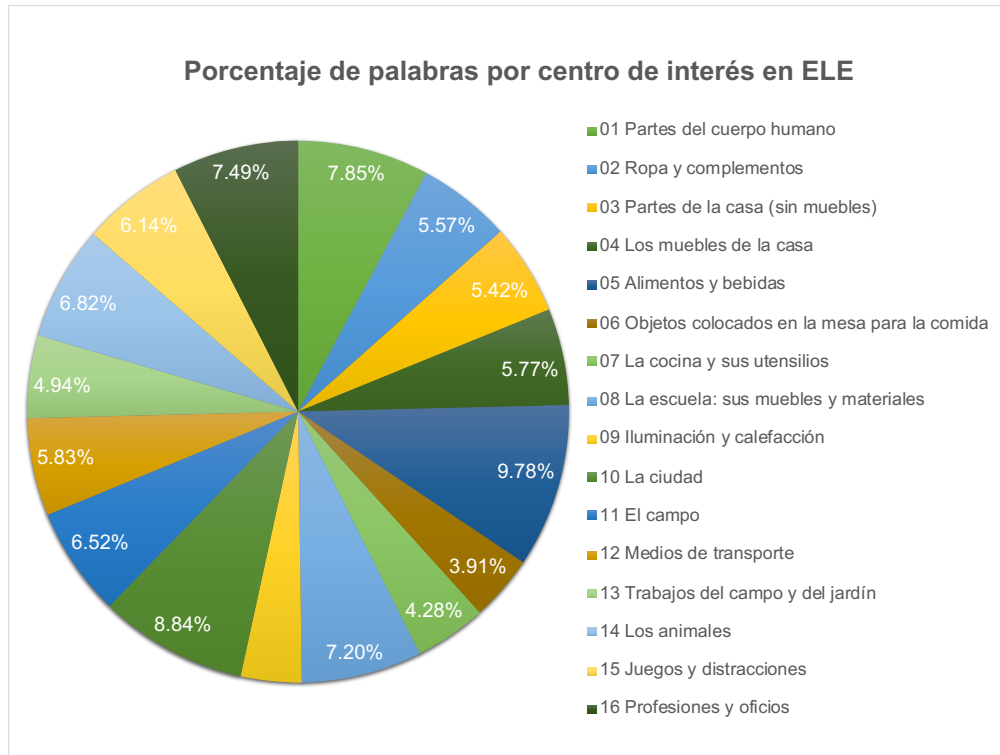


Gráfico 5.7. Porcentaje de palabras por centro de interés en ELE

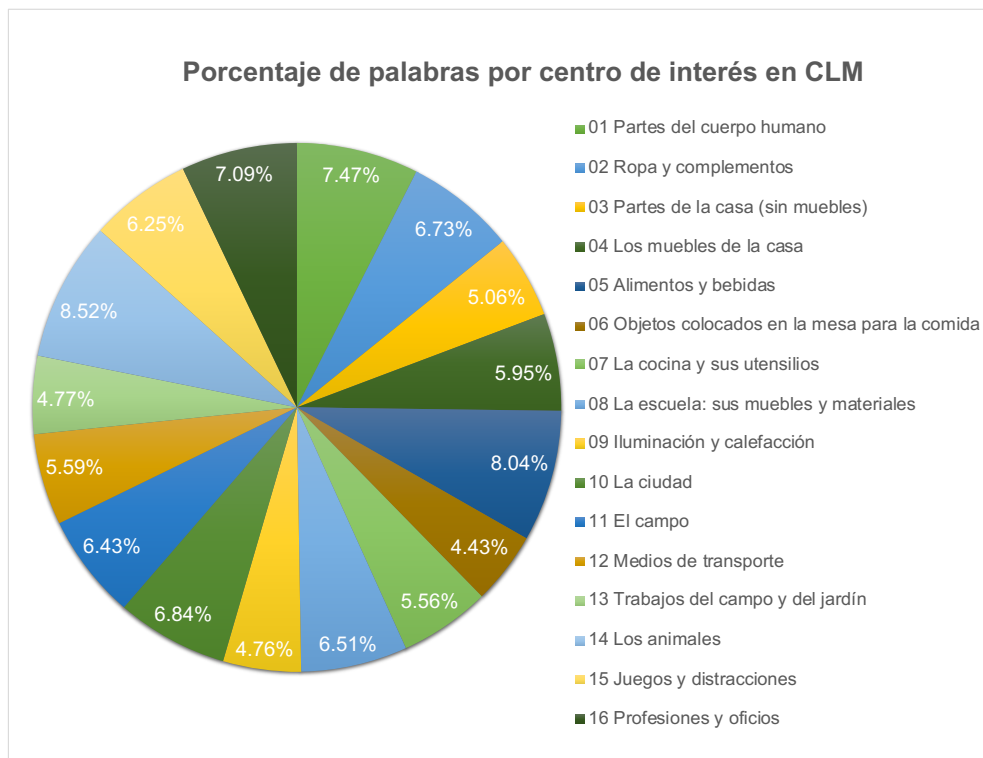


Gráfico 5.8. Porcentaje de palabras por centro de interés en CLM

Los datos expuestos arriba informan de las enormes diferencias en la

producción léxica entre los centros de interés, tanto en ELE como en CLM. Así comprobamos que los estudiantes muestran capacidades asociativas bastante desiguales ante un campo semántico u otro. La mayor distancia se da, en el corpus de ELE, entre *Alimentos y bebidas* (4355) e *Iluminación y calefacción* (1613), una diferencia de 2742 palabras. Mientras tanto, la mayor distancia en el corpus de CLM es de 1809 palabras, entre *Los animales* (3765) y *Objetos colocados en la mesa para la comida* (1956). De ahí, llegamos a suponer que, desde la perspectiva de cada corpus global, los centros de interés se diferencian en un mayor grado en ELE que en CLM con respecto a la producción de palabras. Ello se comprueba en los dos gráficos de porcentaje. En ELE, los dos extremos de los porcentajes señalan una distancia de un 6.16% del total (9.78%, de *Alimentos y bebidas*, y 3.62%, de *Iluminación y calefacción*), mientras que en CLM se da una diferencia de un 4.09% (8.52%, de *Los animales*, frente a 4.43%, de *Objetos colocados en la mesa para la comida*).

Puesto que las áreas temáticas influyen de manera considerable en la producción léxica, nos interesa averiguar, seguidamente, el mayor o menor grado de productividad que supone cada centro de interés. Las celdas en fondo gris de la Tabla 5.4 indican aquellos centros de interés cuyo total de palabras es superior a la media de los dieciséis centros, que era 2782.81, en ELE, y 2762.06, en CLM, como se consignaba en la Tabla 5.1. De ahí, se observa que, salvo *Ropa y complementos*, los centros de interés que superan la media con respecto al total de palabras son los mismos en el corpus de ELE que en el de CLM; una coincidencia en el grado de productividad de los centros independiente de la lengua de la encuesta. Concretamente, por encima de la media están los centros más productivos: *Partes del cuerpo humano*, *Alimentos y bebidas*, *La escuela: sus materiales y muebles*, *La ciudad*, *El campo*, *Los animales* y *Profesiones y oficios* de las dos lenguas, más *Ropa y complementos* en la CLM. Por el contrario, quedan por debajo de la media los centros menos productivos: *Partes de la casa (sin muebles)*, *Los muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Iluminación y calefacción*, *Medios de transporte*, *Trabajos del campo y del jardín* y *Juegos y distracciones* en los corpus de ambas lenguas, más

Ropa y complementos en ELE.

En la Tabla 5.5, se ordenan los centros de interés desde el rango más alto hasta el más bajo según la producción léxica. La media comentada, que divide los centros entre los más productivos y los menos productivos, se marca en esta tabla con una línea gruesa negra. Se observa otra vez que los centros que se hallan por encima y por debajo de ese valor son los mismos, excepto *Ropa y complementos*.

Rango	ELE	CLM
1	05 Alimentos y bebidas	14 Los animales
2	10 La ciudad	05 Alimentos y bebidas
3	01 Partes del cuerpo humano	01 Partes del cuerpo humano
4	16 Profesiones y oficios	16 Profesiones y oficios
5	08 La escuela: muebles y materiales	10 La ciudad
6	14 Los animales	02 Ropa y complementos
7	11 El campo	08 La escuela: muebles y materiales
8	15 Juegos y distracciones	11 El campo
9	12 Medios de transporte	15 Juegos y distracciones
10	04 Los muebles de la casa	04 Los muebles de la casa
11	02 Ropa y complementos	12 Medios de transporte
12	03 Partes de la casa (sin muebles)	07 La cocina y sus utensilios
13	13 Trabajos del campo y del jardín	03 Partes de la casa (sin muebles)
14	07 La cocina y sus utensilios	13 Trabajos del campo y del jardín
15	06 Objetos en la mesa para la comida	09 Iluminación y calefacción
16	09 Iluminación y calefacción	06 Objetos en la mesa para la comida

Tabla 5.5. Rango de los centros de interés en ELE y en CLM según el total de palabras

Como podemos ver en la Tabla 5.5, en ELE, el centro más productivo es *Alimentos y bebidas*, y este centro está en una posición también muy alta —la segunda— en el corpus de CLM. Mientras tanto, en CLM, el primer lugar en producción léxica lo ocupa *Los animales*, centro que cae en ELE a la sexta

posición. Es digno mencionar que *Alimentos y bebidas*, *La ciudad*, *Partes del cuerpo humano* y *Profesiones y oficios* son los cuatro centros que ostentan lugares predominantes en ambos corpus. En cambio, *Iluminación y calefacción* y *Objetos colocados en la mesa para la comida* son los dos centros que generan la menor cantidad de palabras en ambas lenguas.

Las celdas en fondo gris de la Tabla 5.5 señala los centros de interés que tienen un rango altamente inestable, o sea, los que presentan una mayor diferencia entre sus rangos en el corpus de ELE y de CLM. *La ciudad* cae de la segunda posición en ELE a la quinta en CLM; *Los animales*, de la primera en CLM a la sexta en ELE; y *Ropa y complementos*, de la sexta en CLM a la undécima en ELE. Por eso, en lo que se refiere a la productividad, estos son los centros ante los que los informantes se comportan de la manera más distinta en una lengua que en otra. Llama la atención el hecho de que, salvo los tres centros más inestables, el resto de centros presentan una notable coincidencia en el orden de sus rangos en ELE y en CLM. Concretamente, en ambas lenguas, el orden del rango va descendiendo desde *Alimentos y bebidas*, pasando sucesivamente por *Partes del cuerpo humano*, *Profesiones y oficios*, *La escuela: sus muebles y materiales* y *El campo*, hasta *Juegos y distracciones*. Luego siguen *Medios de transporte* y *Los muebles de la casa*, con una mínima diferencia entre las dos lenguas. Por detrás de estos, quedan *Partes de la casa (sin muebles)*, *Trabajos del campo y del jardín* y *La cocina y sus utensilios*, aunque alternan sus posiciones en las dos lenguas. Por último, se encuentran *Objetos colocados en la mesa para la comida* e *Iluminación y calefacción*.

Así pues, después de analizar las dos tablas de arriba, llegamos a la conclusión de que, a pesar de que la productividad léxica de un centro de interés difiere de la de otro, los campos léxicos son bastante coincidentes en ambas lenguas —o sea, independientemente de la lengua de la encuesta— 1) en su distribución a partir de la media de producción, es decir, en su división entre los más productivos y los menos productivos, y 2) en su escala del mayor al menor grado de productividad léxica.

5.1.2.2. Promedio de palabras por informante

El promedio de palabras por informante consiste en un índice muy fiable, porque elimina la influencia provocada por el número de informantes y habla de la producción léxica del estudiante medio. Por lo tanto, es altamente reconocido para hacer comparaciones, centro por centro, entre diferentes investigaciones. Vamos a ver a continuación el promedio de producción léxica en cada uno de los dieciséis centros de interés en ELE y en CLM. Como podemos ver abajo, la Tabla 5.6 enumera las cifras de dicho promedio de cada área semántica en los corpus de ELE y de CLM, e informa del rango de las áreas. El Gráfico 5.9 expone las diferencias cuantitativas tanto entre los centros de interés como entre las dos lenguas de la encuesta.

Centro de interés	Promedio de palabras		Rango	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	12.14	19.19	3	3
02 Ropa y complementos	8.62	17.29	11	6
03 Partes de la casa (sin muebles)	8.39	13.01	12	13
04 Los muebles de la casa	8.92	15.28	10	10
05 Alimentos y bebidas	15.12	20.67	1	2
06 Objetos en la mesa para la comida	6.04	11.37	15	16
07 La cocina y sus utensilios	6.62	14.27	14	12
08 La escuela: muebles y materiales	11.13	16.73	5	7
09 Iluminación y calefacción	5.60	12.24	16	15
10 La ciudad	13.67	17.58	2	5
11 El campo	10.08	16.52	7	8
12 Medios de transporte	9.01	14.37	9	11
13 Trabajos del campo y del jardín	7.64	12.26	13	14
14 Los animales	10.54	21.89	6	1
15 Juegos y distracciones	9.50	16.05	8	9
16 Profesiones y oficios	11.58	18.23	4	4

Tabla 5.6. Promedio de palabras por informante por centro de interés en ELE y en CLM y el rango de los centros

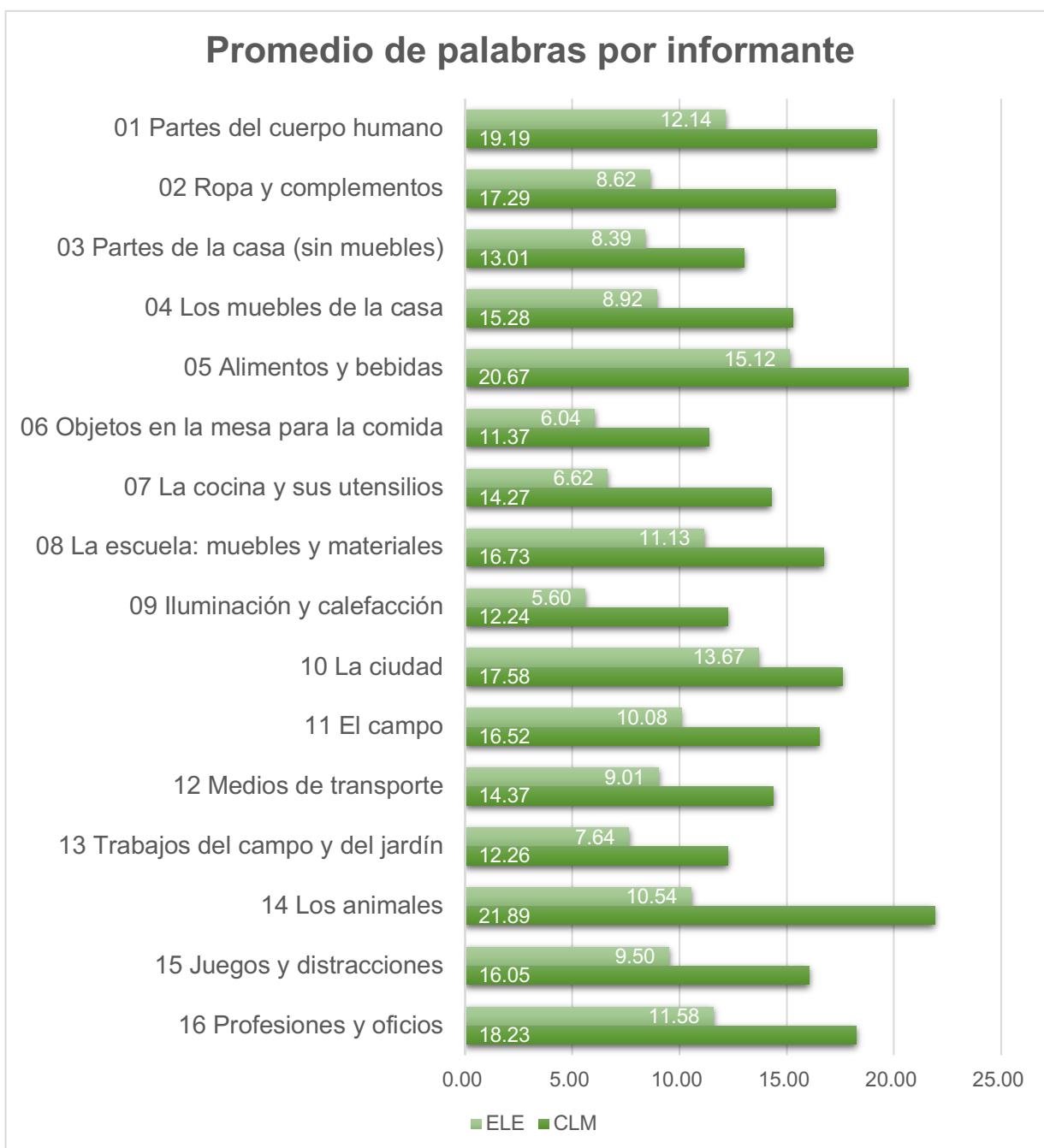


Gráfico 5.9. Promedio de palabras por informante por centro de interés en ELE y en CLM

Al igual que en el análisis sobre el total de palabras, el centro más productivo en ELE es *Alimentos y bebidas* (15.12), seguido por *La ciudad* (13.67) y *Partes del cuerpo humano* (12.14). Son los únicos tres centros en los que los alumnos evocan más de 12 palabras en español. En CLM, los campos léxicos con promedios más altos son *Los animales* (21.89), *Alimentos y*

bebidas (20.67) y *Partes del cuerpo humano* (19.19), tres medias que rondan las 20 palabras. Por otra parte, los centros de interés menos fecundos en ambas lenguas son *Iluminación y calefacción* (5.60 en ELE y 12.24 en CLM) y *Objetos colocados en la mesa para la comida* (6.04 en ELE y 11.37 en CLM). Cabe señalar que el promedio de palabras por informante en ELE es superado por dicho índice en CLM en todos los centros de interés sin excepción. Pero, además, el área semántica que alcanza el máximo promedio en ELE (*Alimentos y bebidas*, 15.12) tiene un resultado incluso inferior a la décima área en CLM (*Los muebles de la casa*, 15.28). En cambio, el centro de interés con la menor cantidad de producción en CLM (*Objetos colocados en la mesa para la comida*, 11.37) muestra un resultado hasta por encima del centro que ocupa la quinta posición en ELE (*La escuela: sus muebles y materiales*, 11.13).

Al analizar las diferencias cuantitativas entre los centros de interés de cada corpus en sí mismo, vemos que, por un lado, la mayor distancia en el corpus de ELE es de 9.52 palabras —la cual se da entre *Alimentos y bebidas* (15.12) e *Iluminación y calefacción* (5.60)—; por otro lado, la diferencia más enorme en el corpus de CLM es de 10.52 palabras —entre *Los animales* (21.89) y *Objetos colocados en la mesa para la comida* (11.37)—. Al parecer, los centros de interés en ELE se difieren en menor grado que en CLM. Sin embargo, si tenemos en cuenta el peso de la distancia máxima —entre el promedio máximo y el mínimo— en el total de promedios de los dieciséis centros, tenemos un resultado igual que el obtenido en el análisis sobre el total de palabras (véase el apartado 5.1.2.1). Es decir, la mayor diferencia en ELE supone el 6.16%¹³² del total de promedios y esta cifra en CLM alcanza el 4.09%¹³³. Por lo tanto, llegamos a afirmar —esta vez tomando como referencia el promedio de palabras por informante— que los centros de interés difieren entre sí en mayor grado en ELE que en CLM, en lo que se refiere a la

¹³² Resulta de dividir la mayor distancia detectada en ELE (9.52 palabras) entre la suma de los promedios de todos los centros en ELE expuestos en la Tabla 5.6 (154.60 palabras) (véase también el Gráfico 5.1).

¹³³ Resulta de dividir la mayor distancia detectada en CLM (10.52 palabras) entre la suma de los promedios de todos los centros en CLM expuestos en la Tabla 5.6 (256.94 palabras) (véase también el Gráfico 5.1).

producción léxica.

Además de las diferencias estadísticas entre los centros de interés, también nos interesa averiguar las discrepancias cuantitativas entre las dos lenguas de la encuesta. Por eso, elaboramos la Tabla 5.7, que presenta por centro de interés las diferencias entre las producciones léxicas en ELE y en CLM —o sea, entre las cifras de cada lengua expuestas en la Tabla 5.6—. Al mismo tiempo, el Gráfico 5.10 muestra, de forma ilustrativa, el mayor o menor grado de dichas diferencias.

Centro de interés	Diferencia de promedio entre ELE y CLM
01 Partes del cuerpo humano	7.05
02 Ropa y complementos	8.67
03 Partes de la casa (sin muebles)	4.62
04 Los muebles de la casa	6.36
05 Alimentos y bebidas	5.55
06 Objetos colocados en la mesa para la comida	5.33
07 La cocina y sus utensilios	7.65
08 La escuela: sus muebles y materiales	5.60
09 Iluminación y calefacción	6.64
10 La ciudad	3.91
11 El campo	6.44
12 Medios de transporte	5.36
13 Trabajos del campo y del jardín	4.62
14 Los animales	11.35
15 Juegos y distracciones	6.55
16 Profesiones y oficios	6.65

Tabla 5.7. Diferencia de promedio de palabras entre ELE y CLM por centro de interés

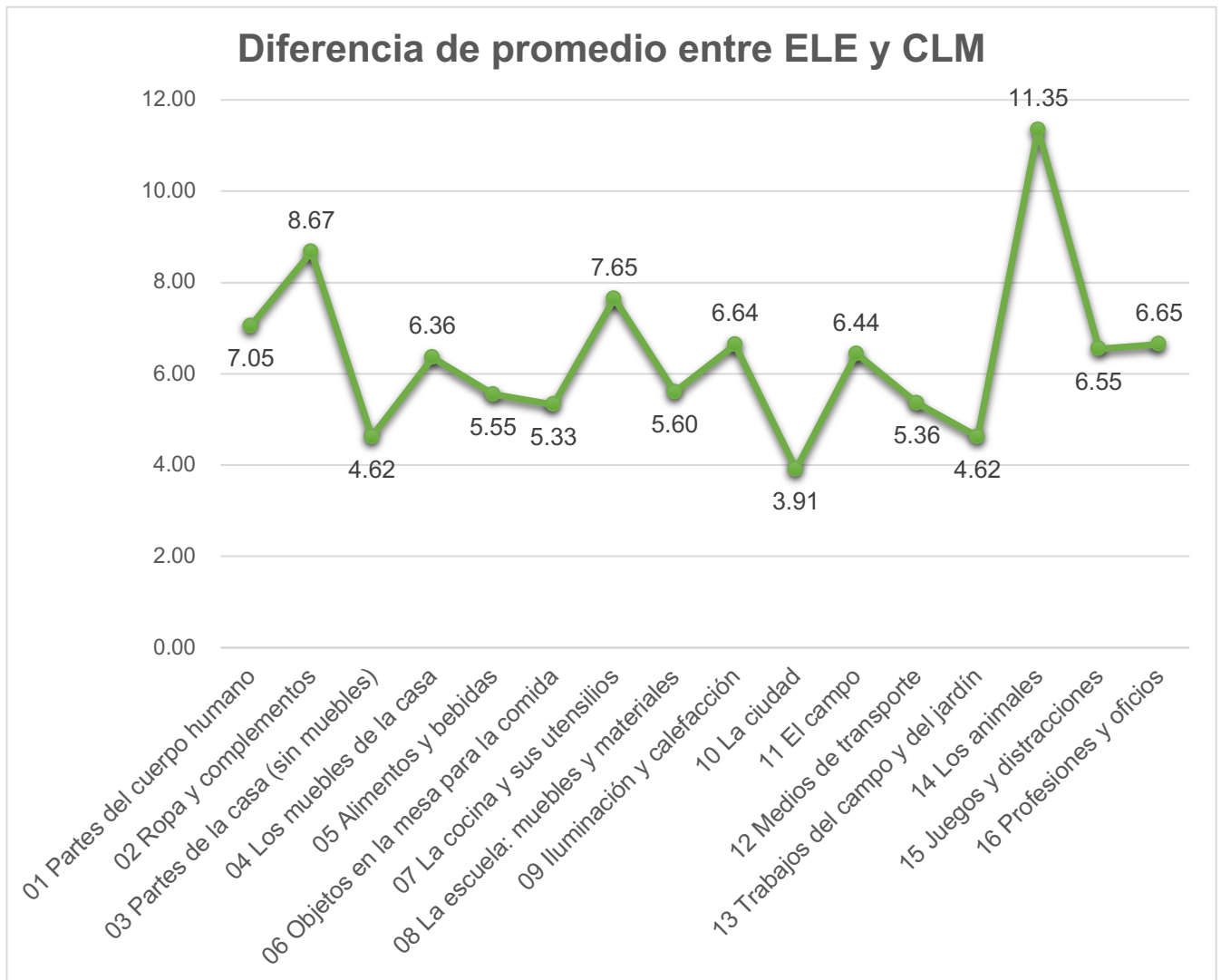


Gráfico 5.10. Diferencia de promedio de palabras entre ELE y CLM por centro de interés

Como ya sabemos, la producción léxica por informante en ELE es inferior a la expresada en CLM en todos los centros de interés sin excepción. Así pues, la Tabla 5.7 pone de manifiesto cuántas palabras menos hay en cada centro del corpus en ELE en comparación con su producción correspondiente en CLM. A continuación, vamos a analizar estas diferencias desde dos perspectivas, con el fin de dar a saber cuáles son los campos léxicos que implican mayores divergencias de producción léxica entre las dos lenguas.

En primer lugar, si pretendemos comparar las cifras de la Tabla 5.7 entre ellas mismas, podemos pasar a observar el Gráfico 5.10. En este, la mayor divergencia se da en *Los animales* (11.35). Es decir, en promedio, los

estudiantes escriben bajo este ámbito nocional 11.35 palabras menos en ELE que en CLM. Es seguido por *Ropa y complementos* (8.67), *La cocina y sus utensilios* (7.65) y *Partes del cuerpo humano* (7.05). Y, en concreto, se trata de los únicos cuatro centros de interés que tienen una diferencia cuantitativa por encima de siete palabras. Por lo tanto, llegamos a la conclusión de que, al compararse entre sí las diferencias de promedio de todos los centros de interés, *Los animales*, *Ropa y complementos*, *La cocina y sus utensilios* y *Partes del cuerpo humano* manifiestan mayores divergencias de producción léxica entre las dos lenguas. En segundo lugar, las filas marcadas en fondo gris en la Tabla 5.7 señalan aquellos campos léxicos cuya diferencia de promedio entre ELE y CLM es superior al propio promedio en ELE (para comprobar cuál es este, véase a la Tabla 5.6); o lo que es lo mismo, son aquellos centros de interés en los que el promedio de palabras por informante en ELE no alcanza ni la mitad del promedio en CLM. De modo que, en relación con los promedios en ELE y en CLM de la Tabla 5.6, *Ropa y complementos*, *La cocina y sus utensilios*, *Iluminación y calefacción* y *Los animales* son las áreas semánticas que muestran mayores diferencias de producción léxica entre las dos lenguas.

Las dos conclusiones extraídas arriba a partir de puntos de partida distintos no se desvían mucho, puesto que coinciden en tres centros de interés: *Los animales*, *Ropa y complementos* y *La cocina y sus utensilios*. Así pues, nos enteramos de que, en estas tres áreas nocionales, la competencia léxica adquirida —la relativa a la producción léxica— de los estudiantes en ELE se alejan mucho de su competencia léxica en CLM.

5.1.2.3. Número de vocablos

En este apartado, vamos a ver los datos cuantitativos relacionados con los vocablos de cada centro de interés en las dos lenguas. Como hemos comentado en los apartados anteriores, los índices sobre los vocablos denotan el grado de riqueza o variedad léxica del conjunto de respuestas producidas por los informantes. Sin embargo, volvemos a advertir de que dichos índices están fuertemente afectados por el comportamiento de los estudiantes, sobre

todo, de los alumnos atípicos que actualizan vocablos poco o incluso nada comunes; porque estos, a pesar de ser poco habituales, también contribuyen al aumento de la riqueza léxica. A continuación, la Tabla 5.8 presenta el número de vocablos de cada centro de interés del corpus en ELE y en CLM. Detrás de ella, los Gráficos 5.11 y 5.12 muestran, de forma ilustrativa, el volumen de los vocablos que tiene cada área semántica con respecto al conjunto del vocabulario.

Centro de interés	Total de vocablos	
	ELE	CLM
01 Partes del cuerpo humano	286	250
02 Ropa y complementos	320	353
03 Partes de la casa (sin muebles)	241	236
04 Los muebles de la casa	217	274
05 Alimentos y bebidas	399	680
06 Objetos colocados en la mesa para la comida	165	186
07 La cocina y sus utensilios	242	243
08 La escuela: sus muebles y materiales	343	322
09 Iluminación y calefacción	245	246
10 La ciudad	700	789
11 El campo	569	720
12 Medios de transporte	329	313
13 Trabajos del campo y del jardín	478	394
14 Los animales	290	326
15 Juegos y distracciones	404	395
16 Profesiones y oficios	422	418

Tabla 5.8. Total de vocablos por centro de interés en ELE y en CLM

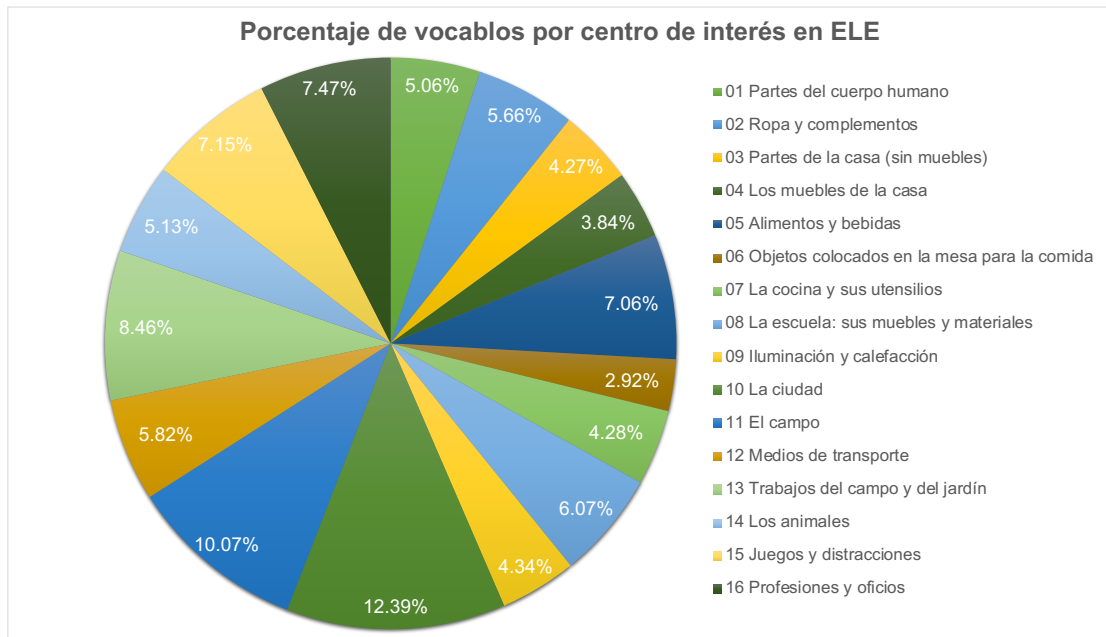


Gráfico 5.11. Porcentaje de vocablos por centro de interés en ELE

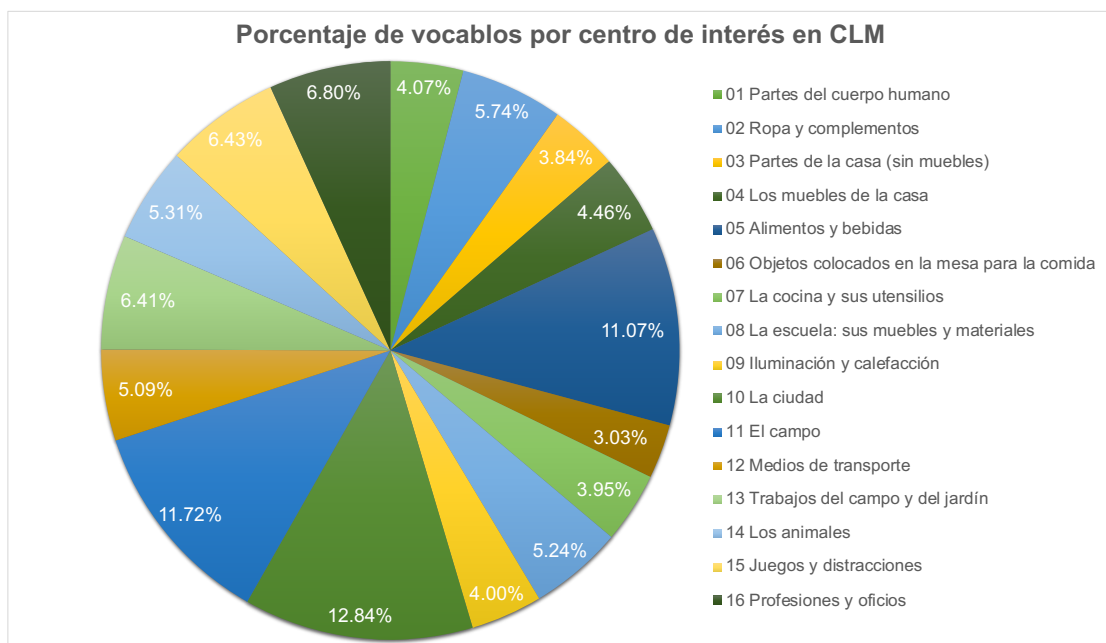


Gráfico 5.12. Porcentaje de vocablos por centro de interés en CLM

Como podemos ver, los resultados de los dieciséis centros de interés señalan enormes diferencias entre un centro y otro. Por eso, llegamos a saber que las distintas áreas temáticas, a pesar de generar estímulo ante los mismos sujetos, implican grados de variedad léxica bastante desiguales. A partir de ahí, nos enteramos de que la variedad de los tipos de relaciones asociativas que

los estudiantes son capaces de evocar depende mucho del área tratada. La mayor diferencia en el número de vocablos en ELE se detecta entre *La ciudad* (12.39%) y *Objetos colocados en la mesa para la comida* (2.92%) —tal como manifiesta el Gráfico 5.11—, una distancia que marca un 9.47% del conjunto del vocabulario. Si pasamos a examinar los datos expuestos en el Gráfico 5.12, llegamos a ver que la mayor divergencia al respecto en CLM cae entre los mismos centros de interés, que consiste en un 9.81% (12.84%, en *La ciudad*, frente a 3.03%, en *Objetos colocados en la mesa para la comida*) del léxico en su conjunto. Las dos cifras de diferencia mayor no se alejan mucho la una de la otra. Por lo tanto, llegamos a la conclusión de que, al considerar de manera global cada corpus, los centros de interés difieren entre sí en un grado parecido en ELE y en CLM con respecto a la riqueza léxica.

Puesto que se evidencian las divergencias entre los campos léxicos con relación a la riqueza léxica, tratamos de averiguar, en adelante, cuáles son los centros de interés más ricos en el vocabulario y cuáles, menos. Si echamos un vistazo a la Tabla 5.8, encontramos unas celdas marcadas en fondo gris. Estas señalan aquellas áreas temáticas que tienen un número total de vocablos por encima de la media de las dieciséis áreas, que es de 353.13, en ELE, y de 384.06, en CLM, como se recogía en la Tabla 5.2. Por consiguiente, vemos que existe una coincidencia absoluta entre las dos lenguas en los centros que suponen una mayor riqueza léxica. Estos son *Alimentos y bebidas*, *La ciudad*, *El campo*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*. En cambio, el resto de centros, a saber, *Partes del cuerpo humano*, *Ropa y complementos*, *Partes de la casa (sin muebles)*, *Los muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La escuela: sus muebles y materiales*, *Iluminación y calefacción*, *Medios de transporte* y *Los animales*, constituyen los que tienen un número de vocablos inferior a la media y, por eso, son los campos semánticos con una menor riqueza léxica, en comparación con los anteriores.

En la siguiente tabla, se ordenan los centros de interés de acuerdo con el rango que alcanzan según el número de vocablos. La media comentada, que

divide los centros de interés en los más ricos y los menos ricos en vocabulario, se marca en la Tabla 5.9 por una línea gruesa negra (la más gruesa en posición horizontal). Podemos observar que, en ambas lenguas, la división en referencia a la riqueza léxica se da entre la sexta y la séptima posición del rango. Además, se comprueba aquí que, tanto por encima de la media como por debajo de ella, se encuentran los mismos centros de interés en cada una de las dos lenguas.

Rango	ELE	CLM
1	10 La ciudad	10 La ciudad
2	11 El campo	11 El campo
3	13 Trabajos del campo y del jardín	05 Alimentos y bebidas
4	16 Profesiones y oficios	16 Profesiones y oficios
5	15 Juegos y distracciones	15 Juegos y distracciones
6	05 Alimentos y bebidas	13 Trabajos del campo y del jardín
7	08 La escuela: muebles y materiales	02 Ropa y complementos
8	12 Medios de transporte	14 Los animales
9	02 Ropa y complementos	08 La escuela: muebles y materiales
10	14 Los animales	12 Medios de transporte
11	01 Partes del cuerpo humano	04 Los muebles de la casa
12	09 Iluminación y calefacción	01 Partes del cuerpo humano
13	07 La cocina y sus utensilios	09 Iluminación y calefacción
14	03 Partes de la casa (sin muebles)	07 La cocina y sus utensilios
15	04 Los muebles de la casa	03 Partes de la casa (sin muebles)
16	06 Objetos en la mesa para la comida	06 Objetos en la mesa para la comida

Tabla 5.9. Rango de los centros de interés en ELE y en CLM según el total de vocablos

Como podemos ver en la Tabla 5.9, los corpus en ELE y en CLM coinciden perfectamente, por un lado, en los dos campos léxicos más ricos (la primera y segunda posición) —*La ciudad* y *El campo*— y, por otro, en el campo léxico menos rico (la última posición) —*Objetos colocados en la mesa para la*

comida—. Salvo *Los muebles de la casa*, que presenta posiciones enormemente alejadas en ELE (la decimoquinta) y en CLM (la undécima), el resto de centros de interés muestran rangos parecidos.

La semejanza en el rango que alcanza el resto de centros la podemos observar a través de los tres compartimentos encuadrados con líneas gruesas en la Tabla 5.9. En primer lugar, de la tercera a la sexta posición del rango se encuentran *Trabajos del campo y del jardín*, *Profesiones y oficios*, *Juegos y distracciones* y *Alimentos y bebidas*. De estos, *Profesiones y oficios* y *Juegos y distracciones* coinciden totalmente en su rango, mientras *Trabajos del campo y del jardín* y *Alimentos y bebidas* alternan sus lugares en las dos lenguas. En segundo lugar, de la séptima a la décima posición, están *La escuela: sus muebles y materiales*, *Medios de transporte*, *Ropa y complementos* y *Los animales*, los cuales manifiestan en las dos lenguas rangos diferentes, pero no alejados. En tercer lugar, de la undécima a la decimoquinta posición, podemos ver que las cuatro áreas semánticas —*Partes del cuerpo humano*, *Iluminación y calefacción*, *La cocina y sus utensilios* y *Partes de la casa (sin muebles)*— muestran un orden de sucesión coincidente en el rango de las dos lenguas, aunque, de ELE a CLM, cada área desciende a la posición inmediatamente siguiente.

De todo lo anterior se deduce que, a pesar de que la riqueza léxica de un centro de interés difiere de la de otro, las áreas semánticas son bastante coincidentes en ambas lenguas —o sea, independientemente de la lengua de la encuesta— 1) en su distribución por la media de número de vocablos, es decir, en su división entre las más ricas y las menos ricas y 2) en su escala del mayor al menor grado de riqueza léxica.

5.1.2.4. Frecuencia acumulada, índice de disponibilidad y aparición

Como hemos hablado a lo largo de este capítulo, el número de vocablos viene fuertemente afectado por el comportamiento de los informantes atípicos cuando estos traen a la mente relaciones asociativas poco comunes en

comparación con el resto de respuestas aportadas por sus compañeros. De modo que, en el apartado 5.1.1.3, siguiendo a otros investigadores, planteamos aplicar medidas de corte en las listas de palabras para someter a las comparaciones entre distintas muestras solo aquellos vocablos que alcanzan cierto grado de disponibilidad o de frecuencia. Así conseguimos evitar la influencia generada por el comportamiento de los sujetos atípicos. En este apartado, vamos a aplicar, en el cotejo por centro de interés, los mismos criterios de corte que hemos tomado como referencia al hacer las comparaciones globales (véase el apartado 5.1.1.3).

Lo primero que analizamos es el número de vocablos al ponerse el límite en el 75% de la frecuencia acumulada. La Tabla 5.10 presenta dicho número de cada uno de los centros de interés en los dos corpus de lenguas diferentes, y expone el rango que el número de cada centro alcanza entre los dieciséis de su corpus correspondiente. Mientras tanto, en el Gráfico 5.13 se evidencian las diferencias cuantitativas entre los números de la Tabla 5.10.

Centro de interés	Número de vocablos que suman una frecuencia acumulada del 75%		Rango	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	22	31	14	14
02 Ropa y complementos	33	56	10	7
03 Partes de la casa (sin muebles)	20	30	15	15
04 Los muebles de la casa	24	33	13	13
05 Alimentos y bebidas	57	137	5	3
06 Objetos en la mesa para la comida	15	19	16	16
07 La cocina y sus utensilios	39	51	8	9
08 La escuela: muebles y materiales	49	54	7	8
09 Iluminación y calefacción	34	41	9	11
10 La ciudad	147	233	1	1
11 El campo	115	186	2	2
12 Medios de transporte	27	38	12	12
13 Trabajos del campo y del jardín	111	95	3	4
14 Los animales	32	42	11	10
15 Juegos y distracciones	68	77	4	6
16 Profesiones y oficios	57	80	5	5

Tabla 5.10. Número de vocablos que suman una frecuencia acumulada del 75% por centro de interés en ELE y en CLM y el rango de los centros

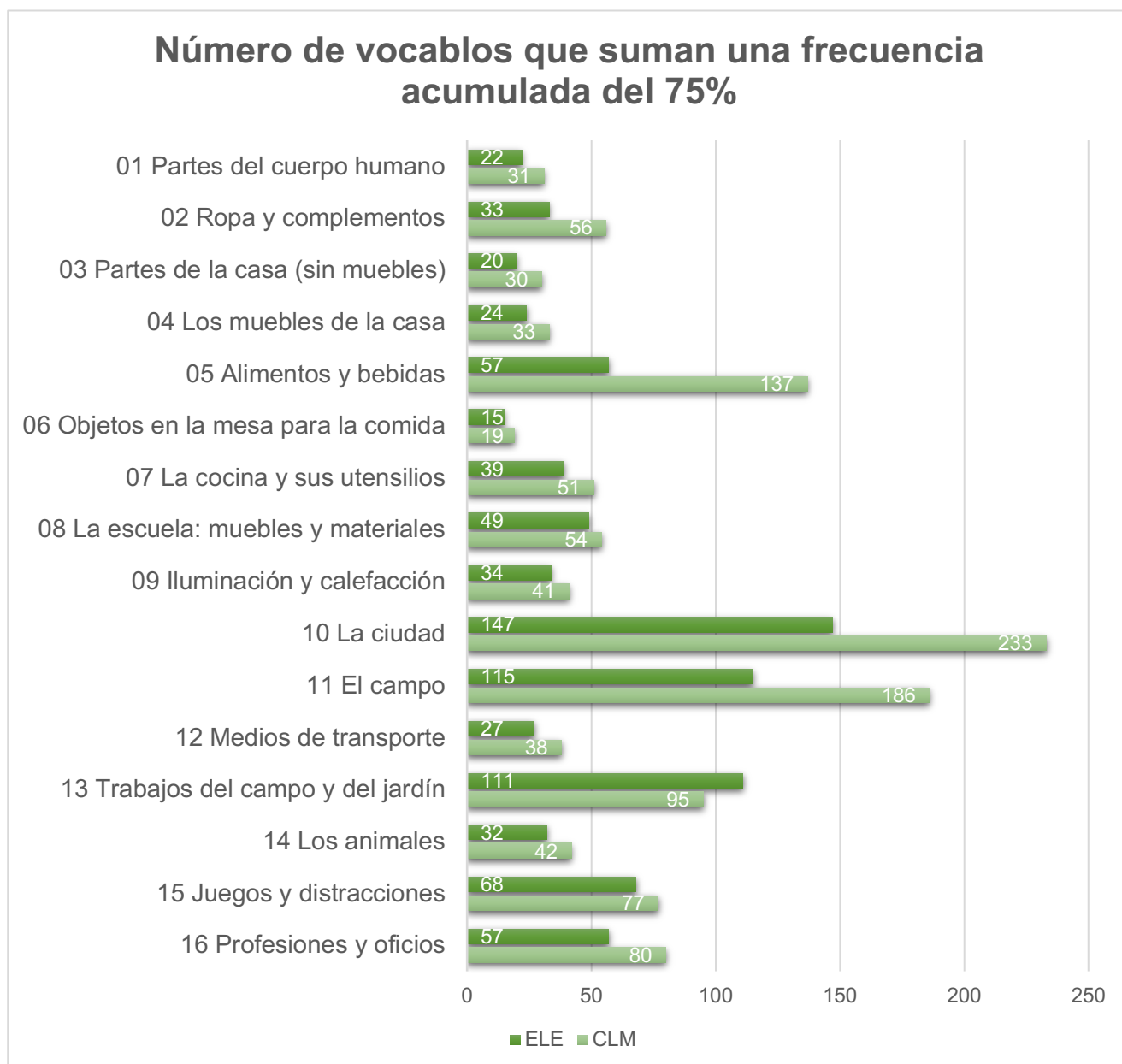


Gráfico 5.13. Número de vocablos que suman una frecuencia acumulada del 75% por centro de interés en ELE y en CLM

Las celdas en fondo gris en las columnas del número de vocablos de la Tabla 5.10 señalan los centros de interés cuya cantidad de vocablos frecuentes (hasta el corte en el 75% de la frecuencia acumulada) está por encima de la media de los dieciséis centros (53.13, en ELE, y 75.12, en CLM). Vemos que los centros que tienen la cifra superior a la media en ELE y en CLM son los mismos; y, por tanto, los que tienen esa cifra inferior a la media en las dos

lenguas también son los mismos. Así es que los dos corpus en ELE y en CLM coinciden en todas las áreas temáticas según su capacidad para generar mayor o menor cantidad de vocablos frecuentes en referencia a la media. Concretamente, las áreas que albergan una mayor cantidad de vocablos frecuentes en las dos lenguas son *Alimentos y bebidas*, *La ciudad*, *El campo*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*. Por el contrario, las que contienen una menor cantidad de vocablos frecuentes en ambas lenguas son el resto de centros de interés. Cabe destacar que las áreas con un número de vocablos frecuentes (hasta el 75% de la frecuencia acumulada) superior a 100 en ELE son *La ciudad*, *El campo* y *Trabajos del campo y del jardín*, mientras que, en CLM, las áreas de esta índole son *La ciudad*, *El campo* y *Alimentos y bebidas*.

Por otra parte, los corpus en ELE y en CLM coinciden en la mitad de los centros (los marcados en fondo gris en las columnas del rango) con respecto al rango que los centros alcanzan según el número de vocablos frecuentes (hasta el corte en el 75% de la frecuencia acumulada). Además, para el resto de campos semánticos, el rango que un campo alcanza en una lengua no se aleja mucho del rango que alcanza en la otra lengua. Esto quiere decir que los dieciséis centros de interés muestran afinidad entre las dos lenguas en su escala de mayor a menor magnitud para estimular vocablos frecuentes. En ambas lenguas, el campo léxico que da cabida al mayor número de vocablos frecuentes es *La ciudad* (la primera posición), seguido por *El campo* (la segunda posición). Por el contrario, el área semántica que contiene la menor cantidad de vocablos frecuentes, en ambas lenguas, es *Objetos colocados en la mesa para la comida* (la última posición), seguido por *Partes de la casa (sin muebles)*, *Partes del cuerpo humano*, *Los muebles de la casa* y *Medios de transporte* (las cuatro áreas, de la decimoquinta a la duodécima posición del rango).

Si echamos un vistazo a las diferencias cuantitativas entre las dos lenguas en cada campo semántico, tal y como muestra el Gráfico 5.13, llegamos a observar que, salvo *Trabajos del campo y del jardín*, en todos los centros de

interés la cantidad de los vocablos más frecuentes (hasta el 75% de la frecuencia acumulada) en ELE se halla en desventaja con respecto a la cantidad en CLM. En *Trabajos del campo y del jardín*, aunque dicho número de vocablos en ELE aventaja al que se recoge en CLM, se trata de una diferencia pequeña. Las mayores desigualdades se dan, de mayor a menor grado y en detrimento del corpus en ELE, en *La ciudad* (86 vocablos de diferencia), *Alimentos y bebidas* (80 de diferencia) y *El campo* (71 de diferencia). Es en estos tres campos léxicos en los que los estudiantes disponen de muchos menos vocablos frecuentes (hasta el 75% de la frecuencia acumulada) en ELE que en CLM.

A continuación, vamos a analizar el número de vocablos al ponerse el límite en el 0.02 del índice de disponibilidad. La Tabla 5.11 expone, por un lado, dicho número de vocablos de cada área temática en los corpus de ELE y de CLM y, por otro lado, el rango de las áreas en cada corpus según el mismo número de palabras. Después de ella, el Gráfico 5.14 presenta, de manera ilustrativa, las diferencias cuantitativas entre los números expuestos en la Tabla 5.11.

Centro de interés	Número de vocablos con I.D. > 0.02		Rango	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	39	59	12	11
02 Ropa y complementos	43	79	10	6
03 Partes de la casa (sin muebles)	36	47	13	15
04 Los muebles de la casa	40	51	11	13
05 Alimentos y bebidas	77	110	2	1
06 Objetos en la mesa para la comida	29	32	15	16
07 La cocina y sus utensilios	44	74	9	7
08 La escuela: muebles y materiales	57	71	6	9
09 Iluminación y calefacción	28	59	16	11
10 La ciudad	88	106	1	2
11 El campo	62	105	4	3
12 Medios de transporte	31	50	14	14
13 Trabajos del campo y del jardín	46	74	8	7
14 Los animales	48	64	7	10
15 Juegos y distracciones	58	93	5	4
16 Profesiones y oficios	65	88	3	5

Tabla 5.11. Número de vocablos con índice de disponibilidad superior a 0.02 por centro de interés en ELE y en CLM y el rango de los centros

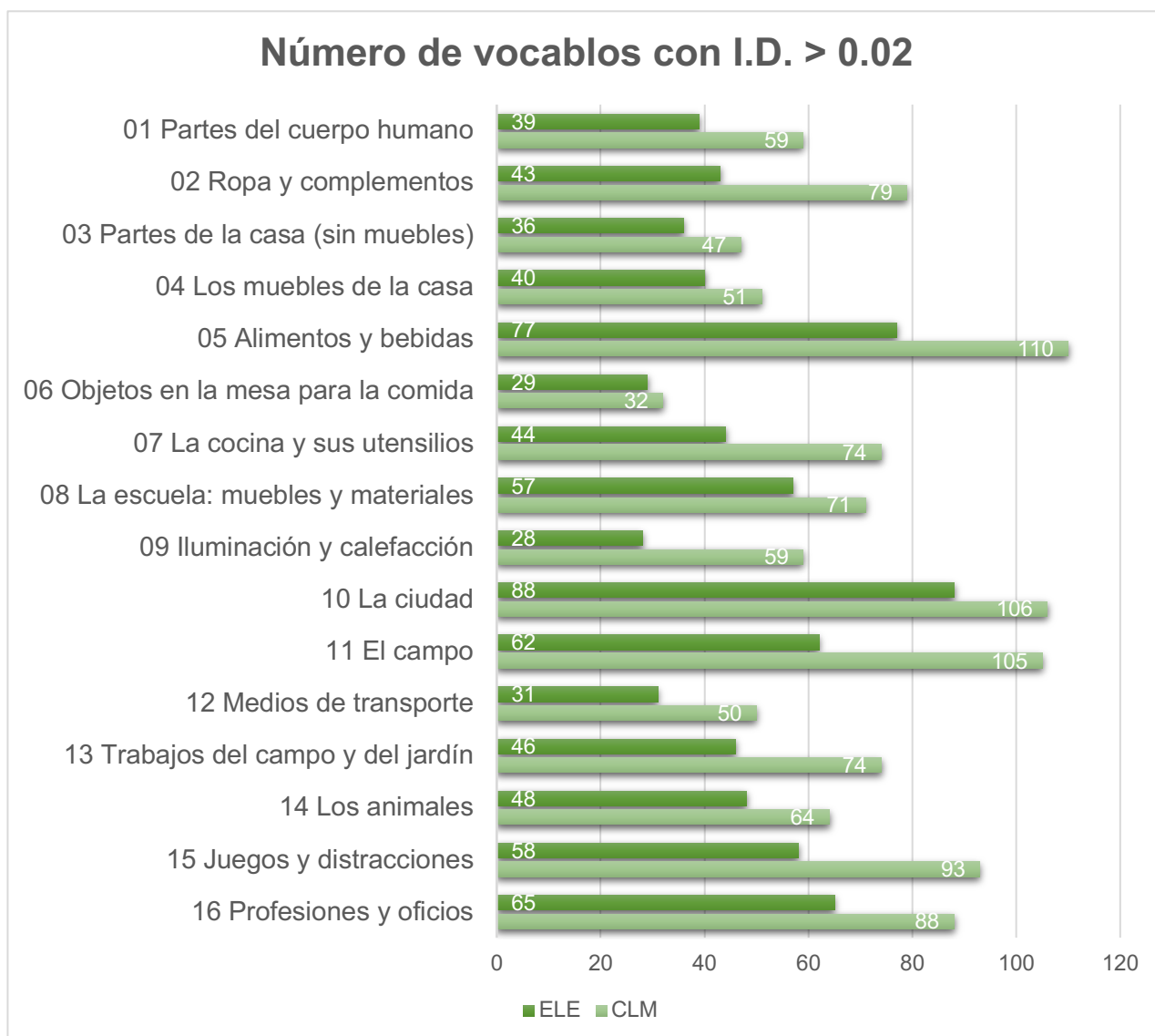


Gráfico 5.14. Número de vocablos con índice de disponibilidad superior a 0.02 por centro de interés en ELE y en CLM

Al igual que en el análisis anterior, hemos marcado en fondo gris, en la Tabla 5.11, aquellos centros de interés cuyo número de vocablos disponibles (hasta el corte en el 0.02 del índice de disponibilidad) supera la media de los dieciséis centros (48.44, en ELE, y 72.63, en CLM). Vemos que, a diferencia de los resultados extraídos a partir de la frecuencia acumulada, los datos de los dos corpus según el índice de disponibilidad no parecen mostrar mucha coincidencia, ni en la distribución de los centros por la media, ni en el rango que alcanzan estos según el número de vocablos disponibles. Ello no resulta del todo ilógico, porque el índice de disponibilidad considera, además de la

frecuencia, la posición de las palabras en las listas. Los estudiantes pueden coincidir fácilmente en los vocablos —lo que determina la frecuencia de un vocablo y el número de vocablos frecuentes—, pero difícilmente coinciden en la posición de los vocablos. Así que el índice de disponibilidad de un vocablo y el número de vocablos disponibles son propensos a mostrar diferencias en las dos lenguas.

De todos modos, como podemos observar en la Tabla 5.11, los campos léxicos que albergan, en referencia a la media, más vocablos disponibles (hasta el 0.02 del índice de disponibilidad) en ELE son, de mayor a menor capacidad, *La ciudad*, *Alimentos y bebidas*, *Profesiones y oficios*, *El campo*, *Juegos y distracciones* y *La escuela: sus muebles y materiales*. Mientras tanto, en CLM, los campos léxicos de esta índole son *Alimentos y bebidas*, *La ciudad*, *El campo*, *Juegos y profesiones*, *Profesiones y oficios*, *Ropa y complementos*, *La cocina y sus utensilios* y *Trabajos del campo y del jardín*. De ellos, *Alimentos y bebidas*, *La ciudad*, *El campo*, *Juegos y distracciones* y *Profesiones y oficios* son los cinco centros en los que coinciden los dos corpus. Por otra parte, a pesar de que los dieciséis centros de interés no presentan afinidad entre las dos lenguas en el rango según la magnitud de vocablos disponibles, es cierto que los primeros dos centros del rango en ambas lenguas son los mismos: *La ciudad* y *Alimentos y bebidas*. Por eso, estos dos son las áreas que dan cabida a una mayor cantidad de vocablos disponibles en ambas lenguas.

En cuanto a las diferencias cuantitativas de cada área temática entre los dos corpus, fijándonos en el Gráfico 5.14, sacamos la conclusión de que, en todos los centros de interés sin excepción, la cantidad de vocablos más disponibles (hasta el 0.02 del índice de disponibilidad) en ELE se encuentra menoscabada con respecto a esa cantidad en CLM. Podemos ver que la mayor divergencia se da en *El campo* (43 vocablos de diferencia en beneficio de la CLM), seguido por, de mayor a menor grado, *Ropa y complementos* (36 de diferencia), *Juegos y distracciones* (35 de diferencia), *Alimentos y bebidas* (33 de diferencia), *Iluminación y calefacción* (31 de diferencia) y *La cocina y sus utensilios* (30 de diferencia). Estos constituyen los centros de interés en los

que los estudiantes dominan muchos menos vocablos disponibles (hasta el 0.02 del índice de disponibilidad) en ELE que en CLM.

De aquí en adelante vamos a analizar por centro de interés en los corpus de ELE y de CLM, por un lado, el número de vocablos altamente disponibles, teniendo en cuenta solamente los vocablos que alcanzan un índice de disponibilidad igual o mayor que 0.1; y, por otro lado, el número de vocablos altamente frecuentes, tomando en consideración aquellos vocablos mencionados por más del 25% de los informantes.

Centro de interés	Número de vocablos con I.D. > 0.1	
	ELE	CLM
01 Partes del cuerpo humano	17	31
02 Ropa y complementos	13	24
03 Partes de la casa (sin muebles)	15	20
04 Los muebles de la casa	12	23
05 Alimentos y bebidas	26	29
06 Objetos colocados en la mesa para la comida	11	14
07 La cocina y sus utensilios	11	23
08 La escuela: sus muebles y materiales	16	27
09 Iluminación y calefacción	7	20
10 La ciudad	15	14
11 El campo	10	19
12 Medios de transporte	11	16
13 Trabajos del campo y del jardín	11	16
14 Los animales	20	24
15 Juegos y distracciones	14	18
16 Profesiones y oficios	14	23

Tabla 5.12. Número de vocablos con índice de disponibilidad superior a 0.1 por centro de interés en ELE y en CLM

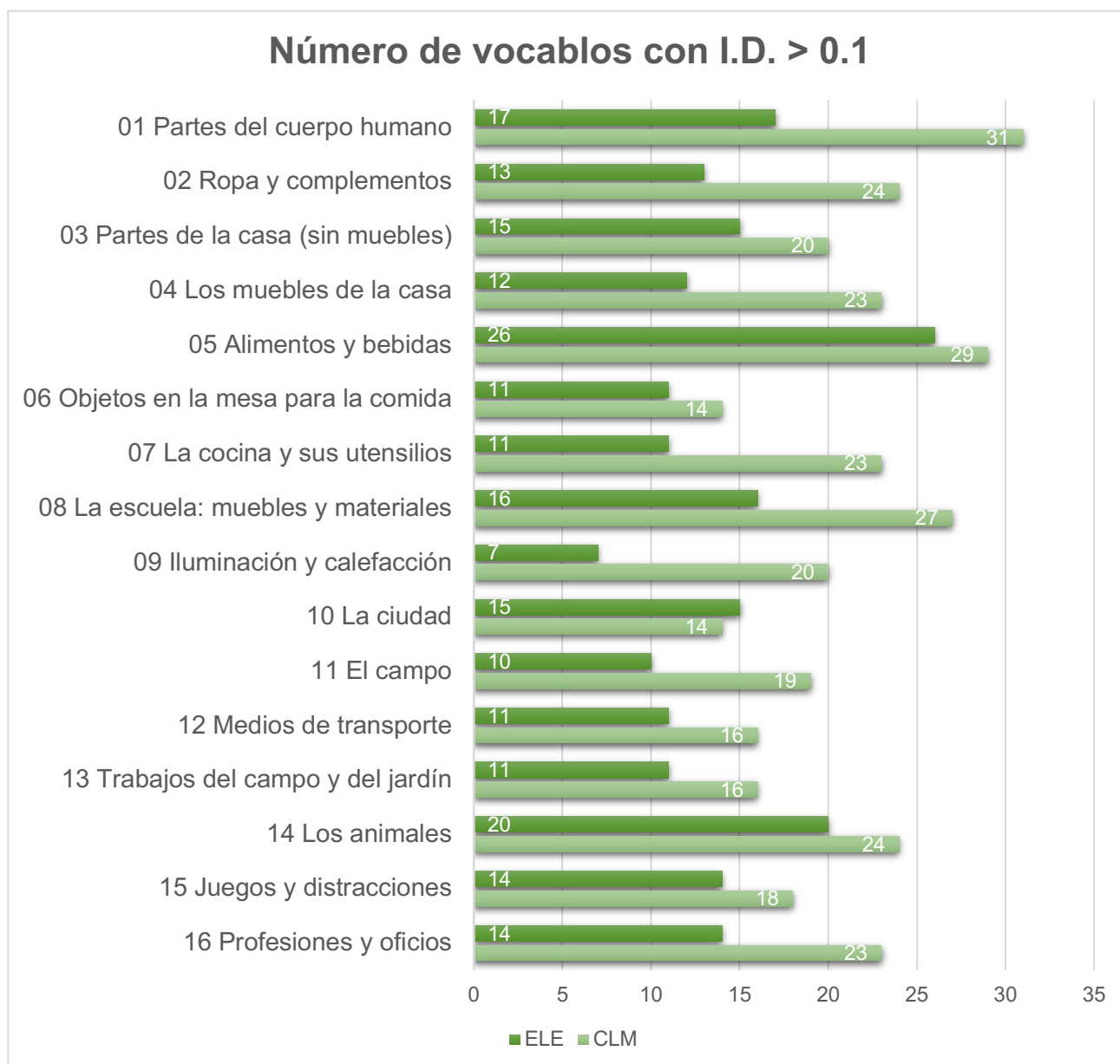


Gráfico 5.15. Número de vocablos con índice de disponibilidad superior a 0.1 por centro de interés en ELE y en CLM

La Tabla 5.12 constata el número de vocablos altamente disponibles (con índice de disponibilidad superior a 0.1) de cada área semántica en las dos lenguas, y el Gráfico 5.15 expone las diferencias entre los centros de cada lengua y entre las dos lenguas en cada centro. Como podemos ver, los campos temáticos —marcados en fondo gris en la Tabla 5.12— que cuenta con más de veinte vocablos altamente disponibles en ELE son solamente dos: *Alimentos y bebidas* (26) y *Los animales* (20). En cambio, en CLM, hay diez centros de interés de esta índole, entre los cuales *Partes del cuerpo humano* (31),

Alimentos y bebidas (29) y *La escuela: sus muebles y materiales* (27) son los que están en las tres primeras posiciones. Cabe mencionar que *Iluminación y calefacción* en ELE (7) es el único campo léxico en el que se registran menos de diez vocablos altamente disponibles en ambos corpus.

Si nos fijamos en las divergencias que se dan entre las dos lenguas, llegamos a ver que, salvo *La ciudad*, en todos los centros de interés el número de vocablos altamente disponibles en ELE es inferior a dicho número en CLM. La mayor divergencia recae en *Partes del cuerpo humano* (14 vocablos de diferencia, en detrimento de ELE), seguido por *Iluminación y calefacción* (13 de diferencia) y *La cocina y sus utensilios* (12 de diferencia). Constituyen los centros en los que los alumnos dominan muchos menos vocablos altamente disponibles en ELE que en CLM. *La ciudad*, aunque señala un resultado mejor en ELE que en CLM, presenta una diferencia muy pequeña —de solo un vocablo— entre las dos lenguas.

Centro de interés	Número de vocablos mencionados por más del 25% de los informantes	
	ELE	CLM
01 Partes del cuerpo humano	16	26
02 Ropa y complementos	9	20
03 Partes de la casa (sin muebles)	11	14
04 Los muebles de la casa	8	18
05 Alimentos y bebidas	18	15
06 Objetos colocados en la mesa para la comida	8	10
07 La cocina y sus utensilios	5	17
08 La escuela: sus muebles y materiales	10	19
09 Iluminación y calefacción	4	12
10 La ciudad	7	5
11 El campo	7	9
12 Medios de transporte	10	14
13 Trabajos del campo y del jardín	4	7
14 Los animales	9	25
15 Juegos y distracciones	7	13
16 Profesiones y oficios	6	18

Tabla 5.13. Número de vocablos mencionados por más del 25% de los informantes por centro de interés en ELE y en CLM

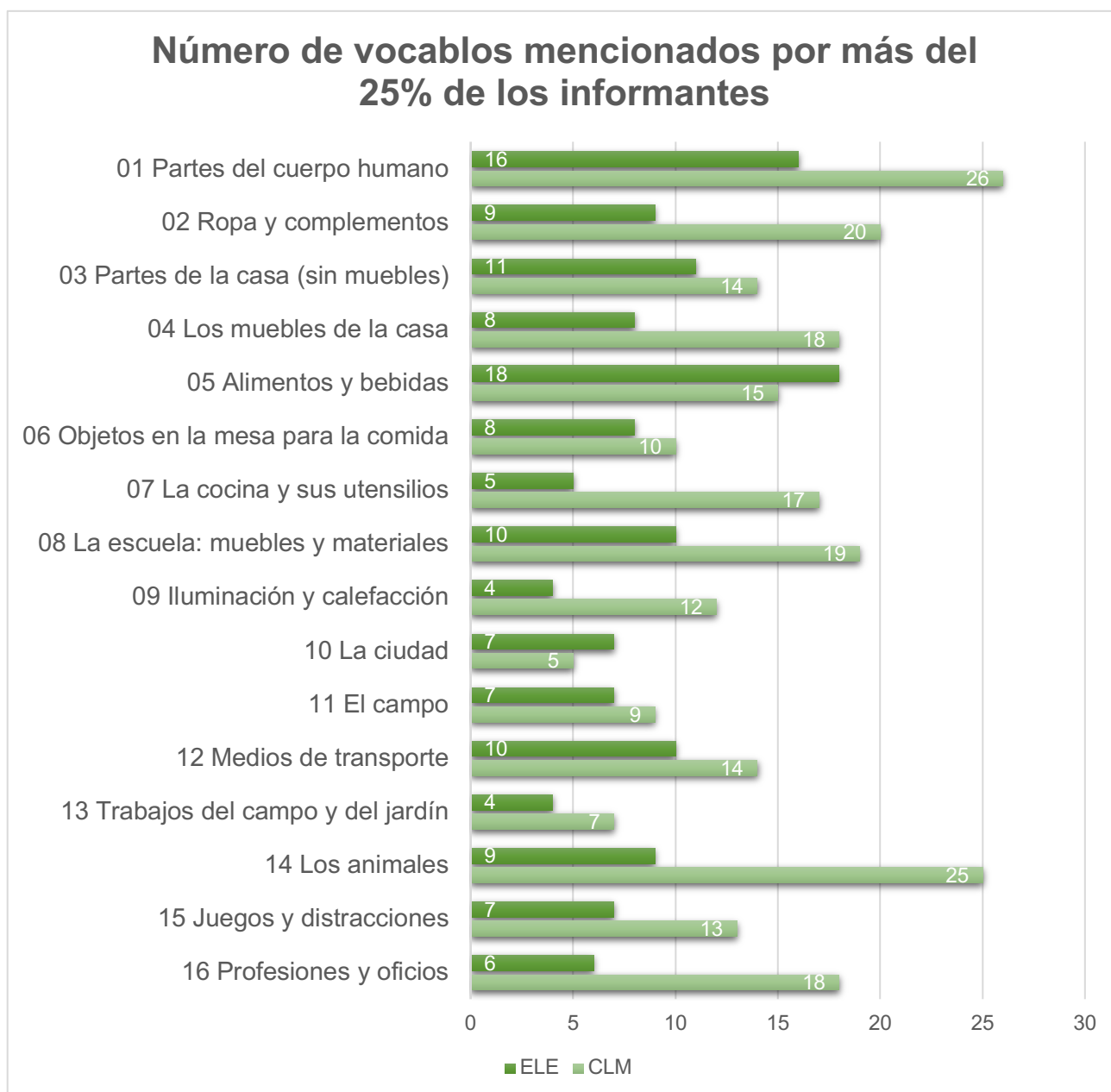


Gráfico 5.16. Número de vocablos mencionados por más del 25% de los informantes por centro de interés en ELE y en CLM

Por último, la Tabla 5.13 muestra el número de vocablos altamente frecuentes (mencionados por más del 25% de los informantes) de cada área semántica en las dos lenguas, y el Gráfico 5.15 presenta las diferencias cuantitativas en dicho número entre los centros de cada lengua y entre las dos lenguas en cada centro. Como podemos ver, los campos temáticos — marcados en fondo gris en la Tabla 5.13— que cuenta con más de quince

vocablos altamente frecuentes en ELE son solamente dos: *Alimentos y bebidas* (18) y *Partes del cuerpo humano* (16). En cambio, en CLM, hay ocho centros de interés de esta índole, entre los cuales *Partes del cuerpo humano* (26), *Los animales* (25) y *Ropa y complementos* (20) son los tres centros que superan veinte vocablos y que están en las primeras tres posiciones en el corpus de CLM. Cabe mencionar que *Iluminación y calefacción* y *Trabajos del campo y del jardín* en ELE (4 vocablos en ambos) son los únicos campos léxicos en los que se registran menos de cinco vocablos altamente frecuentes en ambos corpus.

5.1.2.5. Densidad léxica e índice de cohesión

Hasta aquí, hemos analizado por centro de interés varios índices relacionados o bien con las palabras, o bien con los vocablos. Los hemos estudiado por separado, porque, como se puede comprobar, no se establece una correlación entre las palabras y los vocablos; es decir, no es cierto que cuantas más palabras se produzcan, más vocablos se den, ni viceversa. Así pues, en este apartado vamos a examinar dos índices que relacionan las palabras con los vocablos. Son la densidad léxica y el índice de cohesión, de los que hemos hablado brevemente en el apartado 3.2.2.1.2. Ambos índices sirven para medir la coincidencia en las respuestas dadas por los informantes en un centro de interés determinado.

La densidad léxica es el resultado de dividir el total de palabras de un centro de interés entre el número de vocablos del mismo centro (Alba, 1995). Por otra parte, el índice de cohesión resulta de dividir el promedio de palabras por informante de un centro de interés entre el número de vocablos del mismo centro (Echeverría et al., 1987). A partir de estas dos formas de cálculo, podemos ver que la densidad léxica, al considerar el total de palabras en un campo temático, puede verse afectada por el número de informantes. En cambio, el índice de cohesión tiene en cuenta el promedio de palabras de un área semántica, por lo tanto, es más adecuado para realizar comparaciones entre investigaciones que contemplan números desiguales de informantes.

Además, el índice de cohesión es más utilizado en los estudios de disponibilidad también porque el resultado de su cálculo da un valor siempre entre 0 y 1 —sería 1 si todos los informantes escribieran las mismas palabras en el centro de interés—. De este modo, cuanto más se acerque a 1 el índice de cohesión, mayor homogeneidad existe entre las respuestas de los estudiantes. En cuanto a la densidad léxica, la correlación es la que sigue: cuanto más alto sea el resultado cuantitativo, mayor conformidad habrá entre las respuestas.

Una vez que tenemos una idea del grado de coincidencia entre las respuestas aportadas por los informantes, podemos deducir si un centro de interés es capaz de suscitar relaciones asociativas convergentes. Cuando la convergencia es mayor, al centro de interés se le denomina centro *cerrado* o *compacto*. Por el contrario, si la divergencia es mayor, se trata de un centro *abierto* o *difuso*. Esto es lo que pretendemos averiguar en el presente apartado. Vamos a observar el carácter compacto o difuso de las áreas temáticas en cada uno de los dos corpus de nuestra investigación y comparar las dos lenguas con el fin de buscar coincidencias o divergencias en ese carácter del centro de interés entre ELE y CLM.

A continuación, presentamos en las Tablas 5.14 y 5.15 los resultados, por centro de interés y por lengua de encuesta, la densidad léxica y el índice de cohesión de nuestra investigación y, asimismo, el rango de los centros en cada lengua según cada parámetro. Sin embargo, por las razones arriba mencionadas, nos fijamos en adelante solamente en el índice de cohesión para averiguar el carácter compacto o difuso de las áreas semánticas. El exponer la densidad léxica tiene como objetivo mostrar que el rango que las áreas temáticas alcanzan en una lengua según la densidad léxica es el mismo que alcanzan en esa lengua según el índice de cohesión (compárese el rango de la Tabla 5.14 con el de la 5.15). De este modo, comprobamos que es verdad que los dos parámetros funcionan de manera igual al señalar el carácter cerrado o abierto de los centros de interés.

Centro de interés	Densidad léxica		Rango	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	12.22	13.20	1	1
02 Ropa y complementos	7.76	8.42	11	9
03 Partes de la casa (sin muebles)	10.02	9.48	6	6
04 Los muebles de la casa	11.84	9.59	2	5
05 Alimentos y bebidas	10.91	5.23	3	14
06 Objetos en la mesa para la comida	10.55	10.52	4	3
07 La cocina y sus utensilios	7.88	10.10	10	4
08 La escuela: muebles y materiales	9.34	8.93	7	7
09 Iluminación y calefacción	6.58	8.56	13	8
10 La ciudad	5.63	3.83	14	16
11 El campo	5.10	3.95	15	15
12 Medios de transporte	7.89	7.90	9	10
13 Trabajos del campo y del jardín	4.60	5.35	16	13
14 Los animales	10.47	11.55	5	2
15 Juegos y distracciones	6.77	6.99	12	12
16 Profesiones y oficios	7.90	7.50	8	11

Tabla 5.14. Densidad léxica por centro de interés en ELE y en CLM

Centro de interés	Índice de cohesión		Rango	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	0.0424	0.0768	1	1
02 Ropa y complementos	0.0269	0.0490	11	9
03 Partes de la casa (sin muebles)	0.0348	0.0551	6	6
04 Los muebles de la casa	0.0411	0.0558	2	5
05 Alimentos y bebidas	0.0379	0.0304	3	14
06 Objetos en la mesa para la comida	0.0366	0.0611	4	3
07 La cocina y sus utensilios	0.0274	0.0587	10	4
08 La escuela: muebles y materiales	0.0324	0.0520	7	7
09 Iluminación y calefacción	0.0229	0.0498	13	8
10 La ciudad	0.0195	0.0223	14	16
11 El campo	0.0177	0.0229	15	15
12 Medios de transporte	0.0274	0.0459	9	10
13 Trabajos del campo y del jardín	0.0160	0.0311	16	13
14 Los animales	0.0363	0.0671	5	2
15 Juegos y distracciones	0.0235	0.0406	12	12
16 Profesiones y oficios	0.0274	0.0436	8	11

Tabla 5.15. Índice de cohesión por centros de interés en ELE y en CLM

Como podemos ver en la Tabla 5.15, el centro de interés más compacto en las dos lenguas es el mismo: *Partes del cuerpo humano*. Ello se traduce en que, en comparación con los demás campos léxicos, en *Partes del cuerpo humano* los estudiantes siempre evocan unas palabras universales y extienden, de manera parecida, las relaciones asociativas en torno a estas palabras que constituyen núcleos semánticos de las redes asociativas; por lo cual se da una alta coincidencia entre las respuestas de distintos informantes. En el segundo centro más compacto, las dos lenguas ya no coinciden. Vemos que la segunda posición en ELE —*Los muebles de la casa*— cae a la quita en CLM. De igual modo, el segundo lugar del rango en CLM —*Los animales*— desciende al quinto en ELE.

Por el contrario, observamos que el área temática más difusa en ELE

difiere de aquella en CLM. En ELE, *Trabajos del campo y del jardín* es la más difusa, mientras que en CLM es el área *La ciudad*. De ello, deducimos que en *Trabajos del campo y del jardín* y en *La ciudad*, los estudiantes siempre traen a la mente palabras heterogéneas en ELE y en CLM respectivamente, y emprenden las relaciones asociativas de variadas maneras; por lo cual llega a mostrarse una alta divergencia entre las respuestas dadas por distintos alumnos. En la segunda área temática más difusa, sí coinciden las dos lenguas. Se trata de *El campo*, que ocupa la decimoquinta posición en ambas lenguas.

A continuación, en la Tabla 5.16, se ordenan los centros según el rango que cada uno alcanza en el corpus de ELE y de CLM según el índice de cohesión —o lo que viene a ser lo mismo, según la densidad léxica—.

Rango	ELE	CLM
1	01 Partes del cuerpo humano	01 Partes del cuerpo humano
2	04 Los muebles de la casa	14 Los animales
3	05 Alimentos y bebidas	06 Objetos en la mesa para la comida
4	06 Objetos en la mesa para la comida	07 La cocina y sus utensilios
5	14 Los animales	04 Los muebles de la casa
6	03 Partes de la casa (sin muebles)	03 Partes de la casa (sin muebles)
7	08 La escuela: muebles y materiales	08 La escuela: muebles y materiales
8	16 Profesiones y oficios	09 Iluminación y calefacción
9	12 Medios de transporte	02 Ropa y complementos
10	07 La cocina y sus utensilios	12 Medios de transporte
11	02 Ropa y complementos	16 Profesiones y oficios
12	15 Juegos y distracciones	15 Juegos y distracciones
13	09 Iluminación y calefacción	13 Trabajos del campo y del jardín
14	10 La ciudad	05 Alimentos y bebidas
15	11 El campo	11 El campo
16	13 Trabajos del campo y del jardín	10 La ciudad

Tabla 5.16. Rango de los centros de interés en ELE y en CLM según la densidad léxica o el índice de cohesión

Tal y como muestra la Tabla 5.16, las celdas marcadas con un fondo de gris claro señalan, por un lado, aquellos centros de interés que coinciden en su rango en las dos lenguas (*Partes del cuerpo humano*, *Partes de la casa (sin muebles)*, *La escuelas: sus muebles y materiales*, *Juegos y distracciones* y *El campo*) y, por otro lado, aquellos que presentan una mínima diferencia —solo una posición— en sus rangos en las dos lenguas (*Objetos colocados en la mesa para la comida* y *Medios de transporte*). Por el contrario, las celdas con un fondo de gris oscuro se refieren a los centros cuyo rango en una lengua dista enormemente de su rango en la otra lengua (*Alimentos y bebidas*, *La cocina y sus utensilios* e *Iluminación y calefacción*).

A pesar de que los campos léxicos manifiestan en las dos lenguas distinta escala del grado más compacto al más difuso, podemos ver que, salvo los tres campos de gris oscuro, existe una tendencia más o menos parecida entre las posiciones que los centros alcanzan en ELE y en CLM. Ahora lo explicamos: los centros que llegan a posiciones más altas en el rango de ambas lenguas son *Partes del cuerpo humano*, *Los muebles de la casa*, *Objetos colocados en la mesa para la comida* y *Los animales*. Por eso, son los cuatro centros más compactos tanto en ELE como en CLM. En cambio, en las posiciones más bajas del rango de ambas lenguas se encuentran *Trabajos del campo y del jardín*, *El campo*, *La ciudad* y *Juegos y distracciones*. Por tanto, son los cuatro centros más difusos tanto en ELE como en CLM. El resto de centros —*Partes de la casa (sin muebles)*, *La escuelas: sus muebles y materiales*, *Profesiones y oficios*, *Medios de transporte* y *Ropa y complementos*— se hallan con un grado mediano de compacidad en ambas lenguas. Así, llegamos a la conclusión de que el carácter compacto o difuso es una propiedad natural de los centros de interés y, en la gran mayoría de nuestros casos, dicho carácter no se altera mucho según la lengua de la encuesta.

En cuanto a la causa de que unos centros sean más compactos y otros más difusos en ambas lenguas, creemos que los más compactos, tanto en ELE como en CLM, si exceptuamos *Los animales*, son aquellos que se refieren a lo más cercano: el cuerpo de uno y el lugar en el que vive —que es su segundo

cuerpo—. En cambio, no todas las casas tienen jardín ni todo el mundo tiene contacto habitual con el campo. *La ciudad* ya está en otro nivel de distancia de lo más propio de una persona —su cuerpo y su casa—. En *Juegos y distracciones* se daría otra circunstancia: creemos que se debe a la gran variedad de juegos y distracciones que hay y a que, seguramente, el gusto personal condiciona las asociaciones —puesto que el gusto personal es precisamente eso, personal, individual, no colectivo, no tiene por qué coincidir entre las personas; de ahí la disparidad—.

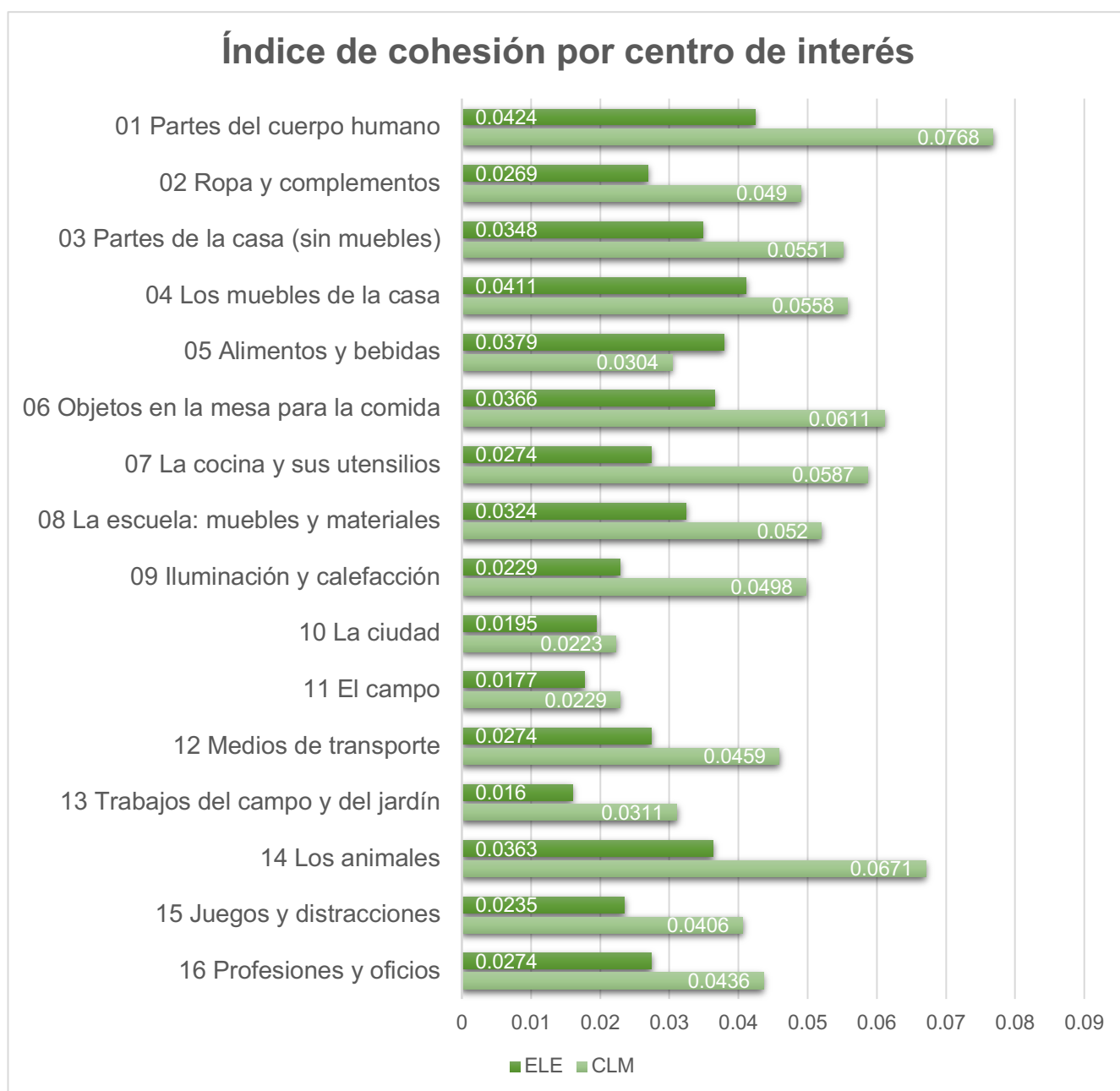


Gráfico 5.17. Índice de cohesión por centros de interés en ELE y en CLM

Por último, en el Gráfico 5.17, ponemos en evidencia, por centro de interés, las diferencias del índice de cohesión entre ELE y CLM. Observamos que en todas las áreas semánticas el índice de cohesión en ELE es menor que el que se da en CLM, con excepción de *Alimentos y bebidas*.

En *Alimentos y bebidas* el índice de cohesión en ELE aventaja al que alcanza este centro en CLM. De ahí se deduce que en este centro los informantes aportan respuestas más homogéneas en ELE que en CLM, o sea, las palabras en ELE sobre *Alimentos y bebidas* presentan menor heterogeneidad que las ofrecidas en CLM. Esto es del todo lógico, puesto que, en su país de origen, los estudiantes tienen contacto diario con una gran variedad de alimentos y bebidas y son capaces de aludirlos todos en su lengua materna. Por el contrario, en lengua extranjera, es posible que el alumnado adquiera de las aulas o de otras fuentes los mismos vocablos, es decir, solamente los referentes a aquellos alimentos y bebidas básicos y habituales de la vida cotidiana o típicos de la cultura que están adquiriendo a través de la lengua extranjera.

En cuanto al resto de centros, en todos ellos las respuestas en ELE muestran una menor homogeneidad, o sea, una mayor heterogeneidad que en CLM. Está claro que ello no es debido a que en su CLM los alumnos anotan respuestas de menor variedad, ya que la cantidad de palabras diferentes en CLM es muy considerable como hemos demostrado en el apartado 5.1.2.3. Suponemos que la mayor heterogeneidad en ELE se debe a la mayor divergencia que se da en la adquisición de las palabras en español. Si un estudiante actualiza unas palabras y otro estudiante aporta otras —según qué palabras se dominan realmente—, naturalmente obtenemos los datos dispersos, lo que contribuye a la mayor heterogeneidad en ELE que en CLM.

5.2. Resultados cuantitativos de las variables extralingüísticas

Como hemos explicado en el capítulo de la Metodología, contemplamos

en nuestras investigaciones de ELE y CLM las mismas variables extralingüísticas con el objetivo de estratificar a los informantes según estas, examinar las diferencias entre un estrato y otro y, así, analizar las incidencias que genera cada variable. En esta sección de la tesis vamos a estudiar el efecto que las cinco variables tenidas en cuenta —el nivel de estudio, el sexo, la procedencia, el nivel de inglés y el tipo de centro— ejercen sobre los resultados del léxico disponible en las dos lenguas entre nuestros informantes. Para cada una de estas variables, vamos a comparar y analizar tres índices de cada estrato según la variable tratada. Los índices para el cotejo son tres de los que hemos presentado en la sección anterior: el promedio de palabras por informante, el número de vocablos y el índice de cohesión. Pero nos vamos a fijar sobre todo en el promedio de palabras por informante, porque, como hemos comentado a lo largo de la sección anterior, resulta ser el índice más fiable para realizar comparaciones entre distintas sub-muestras. En los siguientes apartados, realizaremos comparaciones, sobre la base de estos índices, entre los estratos de cada variante de la variable tratada bajo cada una de las dos lenguas, y entre las dos lenguas bajo cada uno de los estratos obtenidos.

5.2.1. Incidencias cuantitativas de la variable “nivel de estudio”

El nivel de estudio es la primera variable que vamos a analizar en esta sección. Como hemos comentado en el apartado 4.4.1, tenemos bajo ella cuatro variantes que se corresponden con los cuatro cursos universitarios en los que los informantes están estudiando en el momento de rellenar las encuestas. Esta variable difiere mucho del resto de variables y creemos que constituye la más importante entre todas las que contemplamos en la tesis, puesto que es la única variable según la cual los estudiantes están estratificados por naturaleza; es decir, los alumnos de los cuatro niveles de estudio están, de por sí, separados para llevar a cabo distintas tareas académicas en forma de grupos paralelos. Así que los distintos estratos según esta variable son los objetos reales de la enseñanza de ELE en el aula.

5.2.1.1. Para empezar, en la Tabla 5.17 presentamos, por centro de interés, el promedio de palabras por informante del corpus en ELE según los cuatro niveles de estudio en cuestión —el primero, segundo, tercero y cuarto curso de la universidad—. Después de ella, la Tabla 5.18 nos permite ver el rango de los centros de interés en cada uno de los cuatro niveles de estudio según el promedio de palabras por informante en ELE, por lo que se percibe más claramente el mayor o menor grado del valor de las cifras expuestas en la Tabla 5.17.

Centro de interés	Promedio de palabras en ELE			
	Primero	Segundo	Tercero	Cuarto
01 Partes del cuerpo humano	8.74	11.69	15.53	16.75
02 Ropa y complementos	6.26	9.14	10.78	10.56
03 Partes de la casa (sin muebles)	7.06	8.19	10.31	9.58
04 Los muebles de la casa	7.62	9.13	10.04	10.23
05 Alimentos y bebidas	12.83	15.05	17.57	17.67
06 Objetos en la mesa para la comida	4.96	5.00	7.02	9.19
07 La cocina y sus utensilios	5.70	5.98	8.14	8.13
08 La escuela: muebles y materiales	10.34	9.92	13.24	12.71
09 Iluminación y calefacción	3.33	5.24	7.96	8.69
10 La ciudad	10.47	13.45	17.94	16.54
11 El campo	6.09	10.65	14.53	13.10
12 Medios de transporte	6.33	8.63	12.29	12.06
13 Trabajos del campo y del jardín	5.56	6.99	10.24	10.58
14 Los animales	6.57	9.50	14.86	16.46
15 Juegos y distracciones	6.25	9.88	13.53	11.65
16 Profesiones y oficios	8.45	12.04	14.88	14.10
Global	7.29	9.41	12.43	12.38

Tabla 5.17. Promedio de palabras por informante por centro de interés en ELE según la variable nivel de estudio

Centro de interés	Rango según el promedio de palabras por informante en ELE			
	Primero	Segundo	Tercero	Cuarto
01 Partes del cuerpo humano	4	4	3	2
02 Ropa y complementos	10	9	10	11
03 Partes de la casa (sin muebles)	7	12	11	13
04 Los muebles de la casa	6	10	13	12
05 Alimentos y bebidas	1	1	2	1
06 Objetos en la mesa para la comida	15	16	16	14
07 La cocina y sus utensilios	13	14	14	16
08 La escuela: muebles y materiales	3	6	8	7
09 Iluminación y calefacción	16	15	15	15
10 La ciudad	2	2	1	3
11 El campo	12	5	6	6
12 Medios de transporte	9	11	9	8
13 Trabajos del campo y del jardín	14	13	12	10
14 Los animales	8	8	5	4
15 Juegos y distracciones	11	7	7	9
16 Profesiones y oficios	5	3	4	5

Tabla 5.18. Rango de los centros de interés por nivel de estudio según el promedio de palabras por informante en ELE

Si echamos un vistazo a la Tabla 5.17, podemos ver que la mayor producción léxica en ELE de todos los centros de interés y de los cuatro niveles de estudio se encuentra en *La ciudad* del estrato del tercer curso —con un promedio de 17.94—, seguido de *Alimentos y bebidas* del cuarto y del tercer cursos —con 17.67 y 17.57, respectivamente—. Por el contrario, la menor producción léxica en ELE se halla en *Iluminación y calefacción* del primer curso —con una media de solamente 3.33—, seguido de *Objetos colocados en la mesa para la comida* del primero y segundo cursos —con una media de 4.96 y 5.00, respectivamente—.

A continuación, vamos a analizar, por nivel de estudio, el promedio de palabras por informante en ELE. Como podemos comprobar en las Tablas 5.17 y 5.18, la mayor producción léxica de los informantes del primer curso se da en *Alimentos y bebidas*, seguido por *La ciudad*, *La escuela: sus muebles y materiales*, *Partes del cuerpo humano* y *Profesiones y oficios*. Cabe destacar que solo existen tres centros en los que los estudiantes del primer curso tienen un promedio de palabras superior a 10. En cambio, en el estrato del segundo curso, los centros que alcanzan más de 10 palabras de promedio son cinco. La mayor productividad léxica se registra también en *Alimentos y bebidas*, seguido por *La ciudad*, *Profesiones y oficios*, *Partes del cuerpo humano* y *El campo*. Por otra parte, los alumnos del tercer curso tienen la inmensa mayoría de los centros con un promedio superior a 10 palabras. Dentro de ellos, *La ciudad* ostenta la mayor producción léxica, seguido por *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *Los animales*. Por último, arrojan un promedio mayor de 10 palabras también la gran mayoría de los centros en los informantes de cuarto curso. La mayor producción léxica de estos últimos informantes se da en *Alimentos y bebidas*, seguido por *Partes del cuerpo humano*, *La ciudad*, *Los animales* y *Profesiones y oficios*.

De ahí, vemos que los estratos del primero, segundo y cuarto cursos coinciden en el centro de interés más productivo en ELE, que resulta ser *Alimentos y bebidas*. En el estrato del tercer curso, este campo asociativo, aunque no es el más productivo, también ostenta una producción altamente aventajada, que lo sitúa en la segunda posición con una diferencia de solamente 0.37 palabras en comparación con la producción de la primera posición del mismo estrato. Así, llegamos a afirmar que *Alimentos y bebidas* es el área temática más productiva en ELE en los cuatro niveles de estudio. Además, *La ciudad*, *Partes del cuerpo humano* y *Profesiones y oficios* son tres centros que se cuentan entre los campos semánticos más productivos en ELE en todos los niveles de estudio.

En cuanto a las diferencias, sabemos que, aparte de los centros más productivos ya comentados, los alumnos del primer curso disponen de mejor

producción léxica en ELE en *La escuela: sus muebles y materiales*; los del segundo curso, en *El campo*; mientras los del tercer y cuarto curso, en *Los animales*. De ello, por un lado, suponemos que los estudiantes del primer curso, sin haberse aprendido muchos vocablos de español en general por la limitación que se da en su fase de estudio, lógicamente dominan mejor las palabras relacionadas con su entorno habitual de estudio —*La escuela: sus muebles y materiales*—, porque es donde toman las clases de ELE y, por tanto, es la circunstancia comunicativa en la que utilizan con más probabilidad el español. Por otro lado, creemos que, en lo que se refiere a la productividad léxica, *Los animales* es el campo asociativo en ELE que asciende con mayor rapidez a medida que incrementa el nivel de estudio de los estudiantes. Adelantamos que *Los animales* es el centro de interés que alcanza la mayor productividad léxica en CLM en todos los cursos (véase la Tabla 5.20). Por eso, dicho ascenso de *Los animales* es del todo lógico, en el sentido de que, a medida que aumenta el nivel de estudio, el grado de productividad de este centro en ELE se acerca cada vez más al grado de productividad que los alumnos tienen en su lengua materna en este mismo centro.

Ahora volvemos a las Tablas 5.17 y 5.18 para observar las áreas semánticas menos productivas en ELE. Vemos que, en los informantes del primer curso, la menor producción léxica se detecta en *Iluminación y calefacción*, seguido por *Objetos colocados en la mesa para la comida*, *Trabajos del campo y del jardín*, *La cocina y sus utensilios* y *El campo*. En cambio, la menor productividad léxica del estrato del segundo curso cae en *Objetos colocados en la mesa para la comida* y es seguido por *Iluminación y calefacción*, *La cocina y sus utensilios*, *Trabajos del campo y del jardín* y *Partes de la casa (sin muebles)*. Por otra parte, los estudiantes del tercer curso generan la menor cantidad de palabras en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *La cocina y sus utensilios*, *Los muebles de la casa* y *Trabajos del campo y del jardín*. Por último, el estrato del cuarto curso tiene el menor promedio de palabras en *La cocina y sus utensilios*, seguido por *Iluminación y calefacción*, *Objetos colocados en la mesa para la comida*, *Partes de la casa (sin muebles)* y *Los muebles de la*

casa.

Se nota una coincidencia considerable por parte de los cuatro niveles de estudio en cuanto a cuáles constituyen sus áreas semánticas menos productivas en ELE. Por un lado, *Objetos colocados en la mesa para la comida* es el centro de interés con la menor producción léxica en ELE en los estratos del segundo y tercer curso; y en el primero y el cuarto curso, este centro está en la decimoquinta y la decimocuarta posición respectivamente. Por otro lado, *Iluminación y calefacción* es el centro menos productivo en ELE en el estrato del primer curso y el que ocupa la penúltima posición en el segundo, tercero y cuarto cursos. Los dos se consideran los campos léxicos con menor producción en todos los niveles de estudio. Además, *La cocina y sus utensilios* es otro campo léxico que se cuenta entre los centros menos productivos en ELE en todos los niveles de estudio.

En cuanto a las áreas semánticas menos productivas en ELE en algunos de los cuatro estratos, o sea, no comunes en todos los niveles de estudios, destacamos que *Trabajos del campo y del jardín* se queda entre los campos léxicos menos productivos en tres de los cuatro estratos —el primero, segundo y tercer curso—. Mientras tanto, *Partes de la casa (sin muebles)* y *Los muebles de la casa* son comunes en dos de los estratos: en el segundo y cuarto curso y en el tercer y cuarto curso, respectivamente. Como excepción, *El campo* se cuenta entre los centros menos productivos en ELE solamente en el estrato del primer curso, y en el resto de niveles de estudio tiene, por el contrario, buenos resultados.

Promedio de palabras por centro de interés en ELE

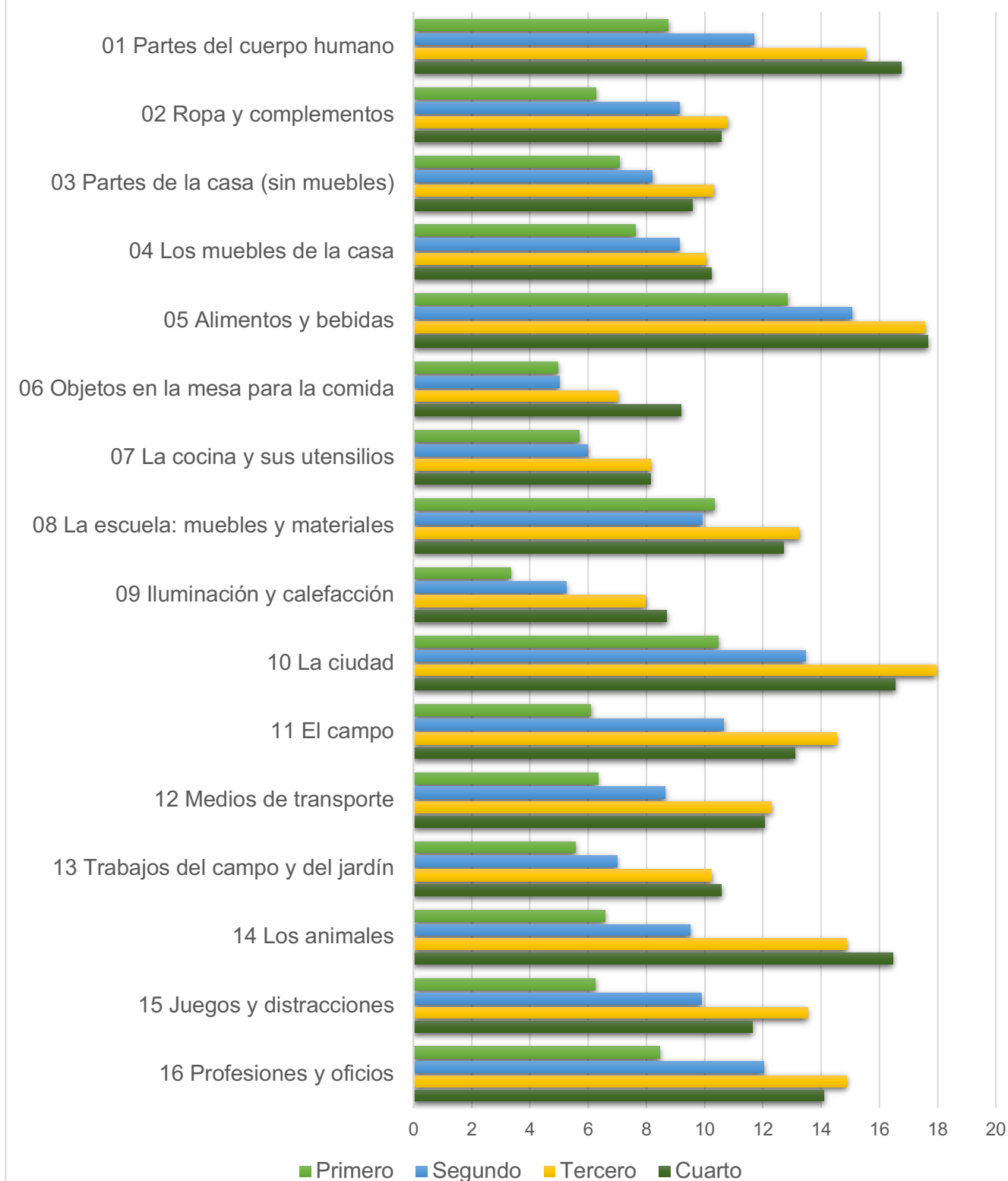


Gráfico 5.18. Promedio de palabras por informante por centro de interés en ELE según la variable nivel de estudio

Si nos fijamos en el gráfico 5.18, que manifiesta las diferencias entre los cuatro niveles de estudio con respecto al promedio de palabras por informante en ELE en cada uno de los centros de interés, llegamos a ver, en todos los campos semánticos, una tendencia de evolución bastante estable de niveles de estudio bajos a niveles altos. Es decir, la producción léxica incrementa progresivamente desde los estudiantes de primer curso hasta los del segundo, y desde los del segundo hasta los del tercero o cuarto curso. Esta tendencia se evidencia en todos los centros de interés con excepción de *La escuela: sus muebles y materiales*, área en la que los alumnos de primer curso no parecen estar demasiado desaventajados. La razón por la que el estrato del primer curso tiene un resultado excepcional en *La escuela: sus muebles y materiales* ya la hemos dado más arriba. A continuación, nos centramos en dicha tendencia de evolución en el corpus de ELE.

Por un lado, dado el aumento del promedio de palabras que entra en relación con el nivel de estudio, llegamos a la conclusión de que, independientemente de cuál sea el centro de interés, la productividad léxica en ELE de los estudiantes crece a medida que incrementa su nivel de estudio —tanto del primero al segundo como del segundo al tercer o cuarto curso—. Por otro lado, vemos que, del tercero al cuarto curso, la productividad léxica no siempre aumenta. Lo que se observa con respecto a estos dos niveles de estudio altos es que, del tercero al cuarto curso, la productividad léxica en ELE 1) incrementa en siete de los dieciséis centros, 2) disminuye en nueve de los centros, y 3) no señala grandes diferencias como las que se dan entre cualesquier otros dos niveles de estudio. De este modo, creemos que la producción léxica en ELE de nuestros informantes se mantiene estable en los últimos dos cursos. Por lo tanto, como complemento de la conclusión anterior, suponemos que el nivel de estudio deja de ejercer influencias en la productividad léxica en ELE de los alumnos una vez que estos entran en el nivel de estudio más alto —el cuarto curso—. Según la experiencia del autor de la tesis como profesor de ELE en Tianjin, pensamos que esta última deducción es del todo explicable, puesto que los estudiantes de esa última fase de estudio tienen mucho menos clases programadas de especialidad en

comparación con las fases anteriores y habitualmente se ocupan de otras tareas ya no académicas, como las prácticas externas y entrevistas de trabajo.

5.2.1.2. De aquí en adelante, presentamos, como referencia de comparación, los resultados sobre el promedio de palabras por informante del corpus en CLM según los cuatro niveles de estudio. Del mismo modo como hemos hecho en los apartados previos sobre ELE, exponemos abajo la Tabla 5.19, en la que se ve dicho índice en CLM en cada centro de interés por nivel de estudio; mientras en la Tabla 5.20 se puede comprobar el rango de los centros también por nivel de estudio según el promedio de palabras por informante en CLM.

Centro de interés	Promedio de palabras en CLM			
	Primero	Segundo	Tercero	Cuarto
01 Partes del cuerpo humano	17.75	20.50	20.90	19.03
02 Ropa y complementos	16.91	17.98	17.30	16.97
03 Partes de la casa (sin muebles)	11.97	12.88	15.35	13.97
04 Los muebles de la casa	14.69	15.42	15.45	16.19
05 Alimentos y bebidas	20.13	21.08	21.00	20.94
06 Objetos en la mesa para la comida	10.84	10.88	12.80	12.41
07 La cocina y sus utensilios	13.12	14.31	17.35	14.75
08 La escuela: muebles y materiales	15.75	16.50	19.05	17.72
09 Iluminación y calefacción	11.47	12.06	13.60	13.31
10 La ciudad	16.71	18.02	18.90	17.91
11 El campo	13.75	18.15	21.25	16.78
12 Medios de transporte	13.10	14.60	16.35	15.47
13 Trabajos del campo y del jardín	10.75	12.31	14.45	14.00
14 Los animales	20.78	22.54	23.50	22.19
15 Juegos y distracciones	14.25	17.35	17.80	16.66
16 Profesiones y oficios	16.91	19.10	20.60	18.13
Global	14.93	16.48	17.85	16.65

Tabla 5.19. Promedio de palabras por informante por centro de interés en CLM según la variable nivel de estudio

Centro de interés	Rango según el promedio de palabras por informante en CLM			
	Primero	Segundo	Tercero	Cuarto
01 Partes del cuerpo humano	3	3	4	3
02 Ropa y complementos	4	7	10	7
03 Partes de la casa (sin muebles)	13	13	13	14
04 Los muebles de la casa	8	10	12	10
05 Alimentos y bebidas	2	2	3	2
06 Objetos en la mesa para la comida	15	16	16	16
07 La cocina y sus utensilios	11	12	9	12
08 La escuela: muebles y materiales	7	9	6	6
09 Iluminación y calefacción	14	15	15	15
10 La ciudad	6	6	7	5
11 El campo	10	5	2	8
12 Medios de transporte	12	11	11	11
13 Trabajos del campo y del jardín	16	14	14	13
14 Los animales	1	1	1	1
15 Juegos y distracciones	9	8	8	9
16 Profesiones y oficios	4	4	5	4

Tabla 5.20. Rango de los centros de interés por nivel de estudio según el promedio de palabras por informante en CLM

Un vistazo a las cifras expuestas en la Tabla 5.19 nos permite saber que la mayor producción léxica en CLM de todos los centros de interés de los cuatro niveles de estudio se encuentra en *Los animales* del estrato del tercer curso — con un promedio de 23.50—. Por el contrario, la menor producción léxica en CLM se halla en *Trabajos del campo y del jardín* del primer curso —con una media de 10.75—. Estos resultados en CLM no contradicen los resultados previamente extraídos en ELE¹³⁴, ya que en ambas lenguas son el tercero y el primer curso los que arrojan, respectivamente, el mayor y el menor promedio de palabras por informante.

¹³⁴ Véase el análisis que sigue las Tablas 5.17 y 5.18 del apartado 5.2.1.1 para consultar los resultados previamente sacados en ELE.

Ahora pasamos al análisis sobre dicho promedio en CLM por nivel de estudio. Como podemos ver en las Tablas 5.19 y 5.20, la mayor producción léxica de los informantes del primer curso se da en *Los animales*, seguido por *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *Ropa y complementos*. Por otra parte, en el estrato del segundo curso, la mayor productividad léxica se detecta en *Los animales* y es seguido por *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *El campo*. Los alumnos del tercer curso, a su vez, tienen el mayor promedio de palabras en *Los animales*, seguido por *El campo*, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios*. Por último, los estudiantes del cuarto curso muestran su mayor productividad léxica también en *Los animales* y es seguido por *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad*.

A partir de lo arriba mencionado, observamos una coincidencia de alto grado entre los cuatro estratos según el nivel de estudio. Concretamente, los estudiantes de los cuatro cursos coinciden totalmente en el centro de interés más productivo en CLM, que resulta ser *Los animales*. Además, llama la atención el hecho de que los estratos del primero, segundo y cuarto curso también coinciden en aquellos campos inmediatamente después del centro más productivo: *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios*; y el tercer curso, aunque no coincide exactamente con los demás cursos en el rango de estos tres centros, muestra una diferencia de solamente una posición en dichos tres campos. Por eso, no hay duda alguna para afirmar que estas tres áreas semánticas se cuentan entre las más productivas en CLM, después de *Los animales*, en todos los niveles de estudio. Comparando estos resultados con los previamente extraídos en ELE según el nivel de estudio, vemos que, en los cuatro niveles de estudio, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son los centros de interés comunes en ELE y en CLM en cuanto a su carácter de alto grado de productividad léxica en el conjunto de los dieciséis centros. Así, deducimos que la producción léxica que los alumnos tienen en ELE viene fuertemente afectada por la que ellos

tienen en CLM, ya que las áreas más productivas en CLM resultan ser las más productivas también en ELE. Suponemos que, independientemente de qué palabras aprendan en el aula de ELE, los estudiantes muestran la inclinación a dominar mejor las palabras españolas de aquellos campos léxicos que ya manejan muy bien en su lengua materna, porque el léxico disponible en CLM les impulsa a aprender el léxico homólogo en ELE.

Con respecto a los centros en los que los cuatro estratos según el nivel de estudio no coinciden en CLM, vemos que, *Ropa y complementos* está entre los campos asociativos más productivos en CLM del primer curso; *El campo*, del segundo y tercer curso; y *La ciudad*, del cuarto curso.

Ahora volvemos a observar las Tablas 5.19 y 5.20 para analizar por nivel de estudio los centros de interés menos productivos en CLM. Vemos que, en los informantes de primer curso, la menor producción léxica se registra en *Trabajos del campo y del jardín*, seguido por *Objetos colocados en la mesa para la comida*, *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *Medios de transporte*. En cambio, la menor productividad léxica del estrato del segundo curso se halla en *Objetos colocados en la mesa para la comida* y es seguido por *Iluminación y calefacción*, *Trabajos del campo y del jardín*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*. Por otra parte, los estudiantes de tercer curso evocan la menor cantidad de palabras en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *Trabajos del campo y del jardín*, *Partes de la casa (sin muebles)* y *Los muebles de la casa*. Por último, el estrato del cuarto curso tiene el menor promedio de palabras en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *Partes de la casa (sin muebles)*, *Trabajos del campo y del jardín* y *La cocina y sus utensilios*.

Los cuatro niveles de estudio coinciden considerablemente en sus centros de interés menos productivos en CLM. El campo semántico que implica la menor productividad léxica es el mismo en tres de los estratos —*Objetos colocados en la mesa para la comida* en el segundo, tercero y cuarto curso—;

y este campo se encuentra en la penúltima posición del primer curso, con una producción léxica también muy limitada. *Iluminación y calefacción* es la segunda área con la menor producción en tres estratos —el segundo, tercero y cuarto curso— y en el primero está en la antepenúltima posición. Algo parecido sucede con *Trabajos del campo y del jardín* y *Partes de la casa (sin muebles)*, ya que los dos se cuentan entre los campos menos productivos de todos los estratos. Comparando estos resultados con los previamente extraídos en ELE, vemos que *Objetos colocados en la mesa para la comida* e *Iluminación y calefacción* son, tanto en ELE como en CLM, los dos centros con una menor productividad léxica en los cuatro niveles de estudio. Esto nos conduce a pensar que, al igual que los centros más productivos en CLM influyen en los centros en ELE, los alumnos tienden a tener menor productividad en ELE en aquellos campos semánticos en los que tienen menos competencia léxica en su lengua materna (CLM). Así que comprobamos — desde la perspectiva de los cuatro niveles de estudio— que la producción léxica que los alumnos tienen en ELE viene fuertemente afectada por la que ellos tienen en CLM, ya que las áreas más productivas o menos productivas en CLM resultan ser las más productivas o menos productivas también en ELE, respectivamente.

En lo que se refiere a las diferencias detectadas sobre los centros menos productivos en CLM, observamos que solo el primer curso muestra debilidad en la producción léxica en *Medios de transporte*; solo el tercer curso, en *Los muebles de la casa*; mientras el segundo y el cuarto curso, en *La cocina y sus utensilios*.

Promedio de palabras por centro de interés en CLM

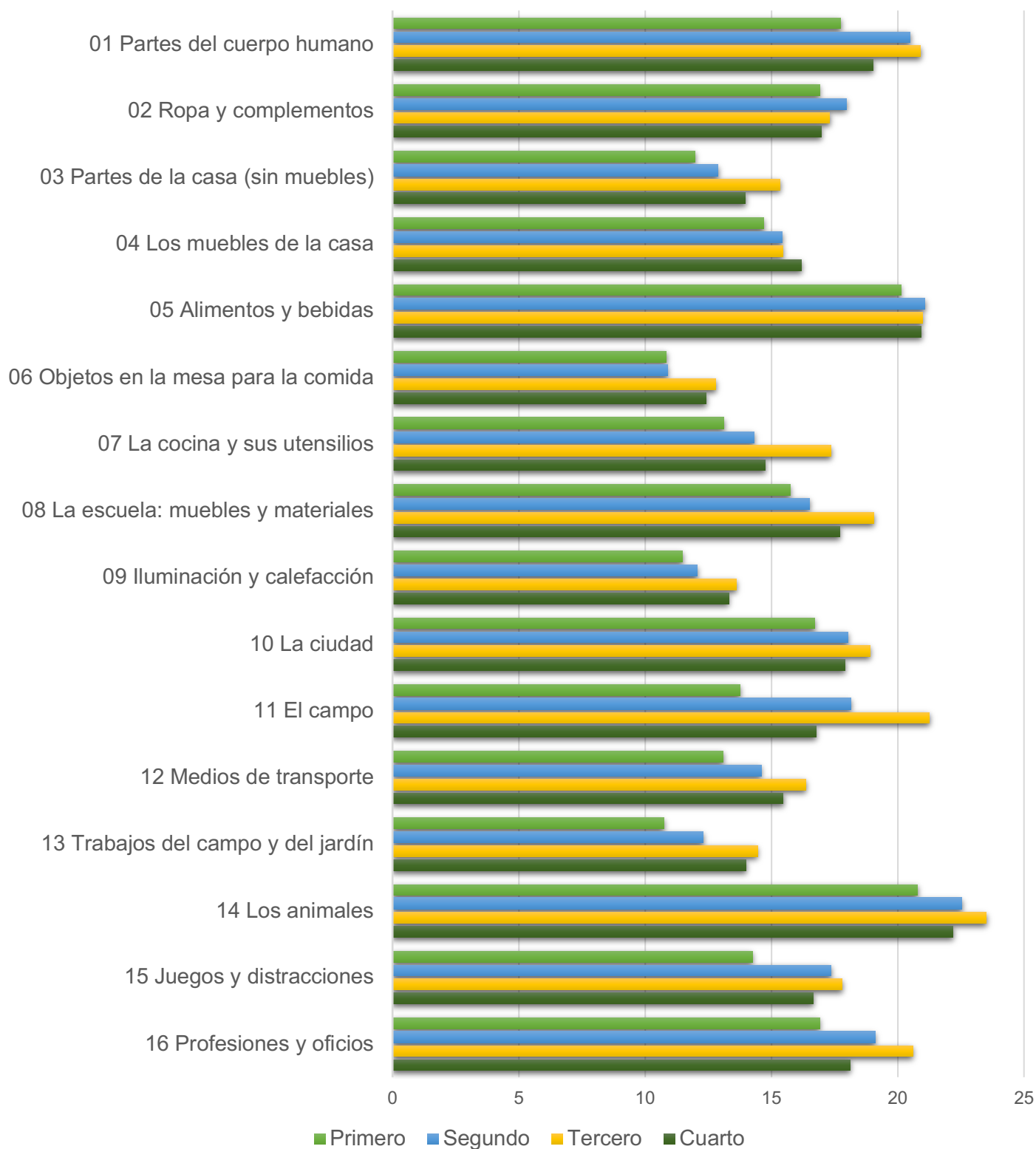


Gráfico 5.19. Promedio de palabras por informante por centro de interés en CLM según la variable nivel de estudio

Si nos fijamos en las diferencias entre los cuatro estratos según el nivel de estudio con respecto a la producción léxica en CLM en cada uno de centros de interés, tal y como evidencia el Gráfico 5.19, podemos darnos cuenta de la misma tendencia de evolución léxica que la observada en el corpus de ELE. Es decir, esta vez en CLM, en todos los centros de interés la producción léxica incrementa de los estudiantes del primer curso a los del segundo, y de los del segundo a los del tercero; con excepción de *Ropa y complementos*, en que el segundo curso supera levemente al tercer curso. Esta superioridad con menos de una palabra —con una diferencia de solamente 0.68— no resulta relevante y, por eso, no impide que saquemos la conclusión sobre la evolución léxica. Así, vemos que la incidencia positiva del nivel de estudio es demostrada tanto en el léxico disponible en ELE como en el actualizado en CLM. En ambas lenguas, cuanto más alto sea el nivel de estudio de los alumnos, mejor producción léxica generan.

Sin embargo, al igual que lo que ya sabemos de la evolución de la productividad léxica en ELE, dicha conclusión no funciona para el estrato del último curso de universidad. Como ya hemos comentado, la producción léxica en ELE de nuestros informantes no evoluciona del tercer curso al cuarto, sino que se mantiene estable. En cambio, lo que se observa en CLM en el último curso es una disminución en quince de los dieciséis centros de interés. Se trata de un retroceso total, ya que la única área en la que el cuarto curso supera al tercero es el área que este curso maneja con debilidad —*Los muebles de la casa*, que se cuenta entre las áreas menos productivas del tercer curso—. La explicación que damos para la excepción del cuarto curso en CLM es la misma a la que hemos recurrido en el apartado sobre los resultados en ELE. Cabe señalar que, en el léxico disponible en ELE los estudiantes no retroceden en el cuarto curso como sucede en el léxico en CLM, quizá porque se mantienen en vilo sabiendo que el español es la especialidad de la que se pueden servir en adelante y, así, no son demasiado flojos para el aprendizaje de esta lengua extranjera.

A continuación, a partir de los datos globales expuestos en las Tablas 5.17

y 5.19, presentamos en los gráficos 5.20 y 5.21 el promedio de palabras global de cada uno de los cuatro niveles de estudio en los corpus de ELE y de CLM.

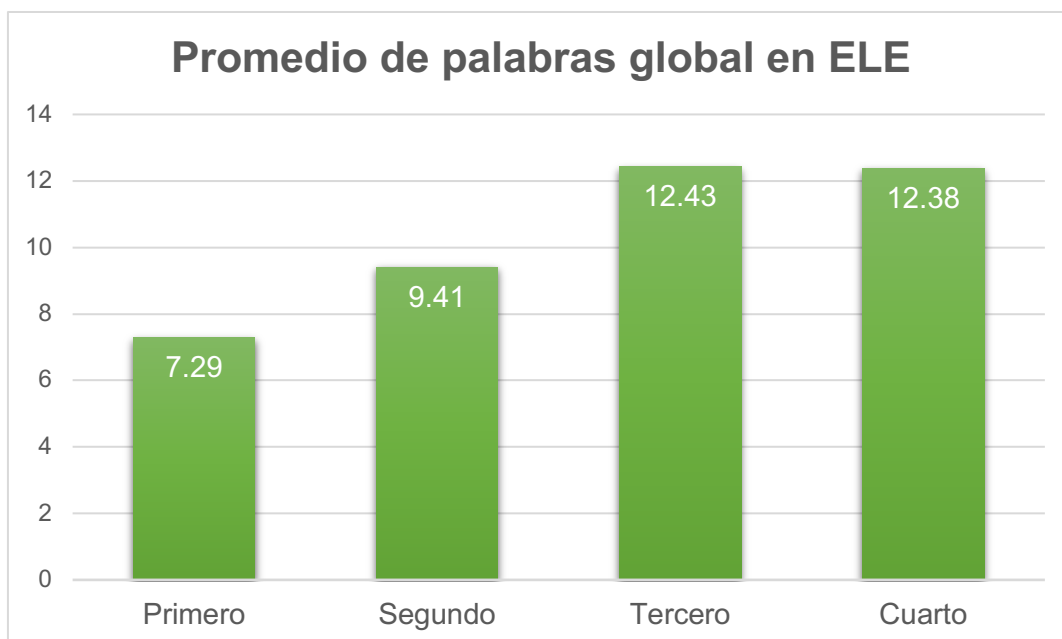


Gráfico 5.20. Promedio global de palabras por informante en ELE según la variable nivel de estudio

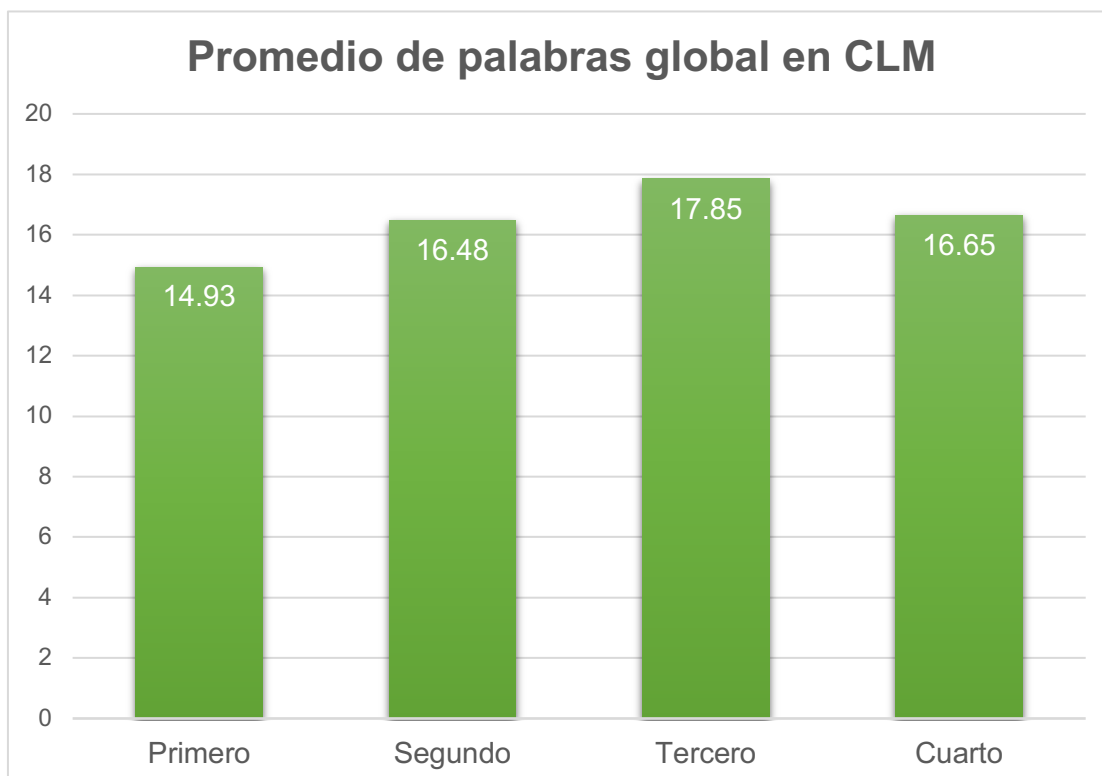


Gráfico 5.21. Promedio global de palabras por informante en CLM según la variable nivel de estudio

Como muestran los dos gráficos de arriba, vemos comprobada la conclusión anteriormente extraída por centro de interés sobre la evolución léxica a lo largo de los años de estudio, pero, esta vez, de manera global e independientemente de las áreas temáticas. Podemos ver en ambos gráficos cómo la producción media de los informantes asciende, en ELE y en CLM, del primer curso al segundo y, de este, al tercero. Por otra parte, en estos gráficos también consta que la productividad léxica de los alumnos en el corpus de ELE se mantiene más o menos igual entre el tercero y el cuarto curso; mientras que en CLM se evidencia un descenso considerable del tercer curso al cuarto.

Otra deducción que podemos sacar comparando los Gráficos 5.20 y 5.21 entre ellos mismos es que, en el estrato del primer curso, la producción léxica en ELE (7.29) supone más o menos la mitad de la que se da en CLM (14.93), pero, a medida que incrementa el nivel de estudio, la producción en ELE no aumenta con la misma velocidad que lo hace en CLM sino con más rapidez. Es decir, a pesar de que el promedio de palabras en CLM crece a lo largo de los años de estudio, el crecimiento gradual y paralelo en ELE es porcentualmente mayor que el de la producción en CLM (9.41 frente a 16.48, en el segundo curso; 12.43 frente a 17.85, en el tercero, y 12.38 frente a 16.65, en el cuarto). La productividad léxica en ELE de nuestros informantes se aproxima cada vez más —a medida que aumenta su nivel de estudio— a la productividad que muestran en CLM. Esta deducción sobre la incidencia del nivel de estudio la vamos a examinar más concretamente en el siguiente apartado, analizando la distancia de producción léxica entre las dos lenguas.

5.2.1.3. La Tabla 5.21 reúne los datos de las Tablas 5.17 y 5.19, presentando, a su vez, los resultados completos sobre el promedio de palabras por informante por nivel de estudio y en las dos lenguas de nuestra investigación. A partir de ella, obtenemos la Tabla 5.22, que muestra, por centro de interés y por nivel de estudio, las diferencias cuantitativas entre el promedio en ELE y el promedio en CLM.

Centro de interés	Promedio de palabras							
	Primero		Segundo		Tercero		Cuarto	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	8.74	17.75	11.69	20.50	15.53	20.90	16.75	19.03
02 Ropa	6.26	16.91	9.14	17.98	10.78	17.30	10.56	16.97
03 Casa	7.06	11.97	8.19	12.88	10.31	15.35	9.58	13.97
04 Muebles	7.62	14.69	9.13	15.42	10.04	15.45	10.23	16.19
05 Alimentos	12.83	20.13	15.05	21.08	17.57	21.00	17.67	20.94
06 Objetos	4.96	10.84	5.00	10.88	7.02	12.80	9.19	12.41
07 Cocina	5.70	13.12	5.98	14.31	8.14	17.35	8.13	14.75
08 Escuela	10.34	15.75	9.92	16.50	13.24	19.05	12.71	17.72
09 Iluminación	3.33	11.47	5.24	12.06	7.96	13.60	8.69	13.31
10 Ciudad	10.47	16.71	13.45	18.02	17.94	18.90	16.54	17.91
11 Campo	6.09	13.75	10.65	18.15	14.53	21.25	13.10	16.78
12 Transporte	6.33	13.10	8.63	14.60	12.29	16.35	12.06	15.47
13 Trabajos	5.56	10.75	6.99	12.31	10.24	14.45	10.58	14.00
14 Animales	6.57	20.78	9.50	22.54	14.86	23.50	16.46	22.19
15 Juegos	6.25	14.25	9.88	17.35	13.53	17.80	11.65	16.66
16 Profesiones	8.45	16.91	12.04	19.10	14.88	20.60	14.10	18.13
Global	7.29	14.93	9.41	16.48	12.43	17.85	12.38	16.65

Tabla 5.21. Promedio de palabras por informante por centro de interés en ELE y en CLM según la variable nivel de estudio

Centro de interés	Diferencia de promedio de palabras entre ELE y CLM			
	Primero	Segundo	Tercero	Cuarto
01 Partes del cuerpo humano	9.01	8.81	5.37	2.28
02 Ropa y complementos	10.65	8.84	6.52	6.41
03 Partes de la casa (sin muebles)	4.91	4.69	5.04	4.39
04 Los muebles de la casa	7.07	6.29	5.41	5.96
05 Alimentos y bebidas	7.30	6.03	3.43	3.27
06 Objetos en la mesa para la comida	5.88	5.88	5.78	3.22
07 La cocina y sus utensilios	7.42	8.33	9.21	6.62
08 La escuela: muebles y materiales	5.41	6.58	5.81	5.01
09 Iluminación y calefacción	8.14	6.82	5.64	4.62
10 La ciudad	6.24	4.57	0.96	1.37
11 El campo	7.66	7.50	6.72	3.68
12 Medios de transporte	6.77	5.97	4.06	3.41
13 Trabajos del campo y del jardín	5.19	5.32	4.21	3.42
14 Los animales	14.21	13.04	8.64	5.73
15 Juegos y distracciones	8.00	7.47	4.27	5.01
16 Profesiones y oficios	8.46	7.06	5.72	4.03
Global	7.64	7.07	5.42	4.27

Tabla 5.22. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable nivel de estudio

Como es comprensible, el promedio de palabras en ELE es inferior al mismo promedio en CLM en todos los centros de interés y en todos los niveles de estudio. Podemos ver en la Tabla 5.22 que la mayor diferencia entre los promedios de ELE y de CLM se halla en *Los animales* en el estrato del primer curso (14.21) en beneficio de la CLM. Cabe destacar que la segunda mayor diferencia también se registra en este centro de interés, pero en los estudiantes del segundo curso (13.04).

Además, las celdas marcadas en fondo gris en la Tabla 5.22 indican aquellas áreas nocionales en las que la diferencia cuantitativa está por encima

del promedio global o, en otras palabras, áreas en las que la producción léxica en ELE de los informantes dista enormemente de la que se da en CLM. Por lo tanto, creemos oportuno que se preste atención a estas áreas semánticas, toda vez que la inferioridad en ELE que se delata en ellas ya no se debe simplemente a que no se les ocurran más palabras a los informantes, sino a que estos no dominan los homólogos en español de las palabras que son capaces de evocar en CLM.

Ropa y complementos y *Los animales* son dos centros en los que los alumnos de todos los niveles de estudio coinciden al respecto. Así es que todos los estudiantes, independientemente del nivel de estudio, muestran considerable debilidad en estos dos centros en ELE, tomando como referencia su producción léxica en CLM. Aparte de estos dos, hay varios campos léxicos que implican deficiencia en ELE con respecto a la producción en CLM en tres de los cuatro estratos. Señalamos que los estratos del segundo, tercero y cuarto cursos muestran deficiencias léxicas en ELE en *La cocina y sus utensilios*. De igual modo, el primero, segundo y tercer curso, en *Iluminación y calefacción*; el primero, segundo y tercer curso, en *El campo*; y el primero, segundo y cuarto, en *Juegos y distracciones*. Como casos únicos, solo el cuarto curso presenta debilidad en *Partes de la casa (sin muebles)* y *Los muebles de la casa*; mientras solo el tercer curso, en *Objetos colocados en la mesa para la comida*.

Así estando las cosas, ahora pasamos a fijarnos en la diferencia global de promedio de palabras entre ELE y CLM por nivel de estudio, la cual se expone en la última fila de la Tabla 5.22 y, de manera ilustrativa, en el Gráfico 5.22. Con dicha diferencia global, retomamos e intentamos comprobar la deducción que hemos hecho al final del apartado 5.2.1.2.

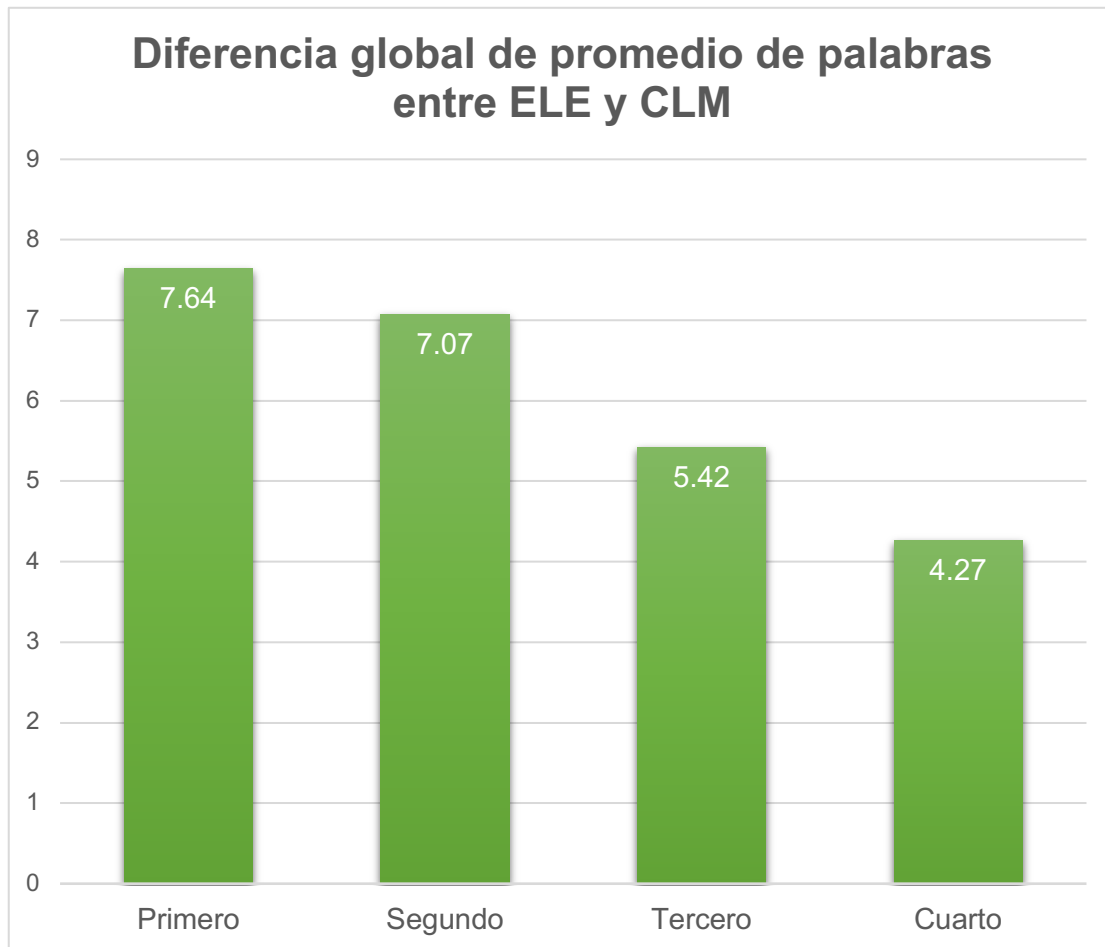


Gráfico 5.22. Diferencia global de promedio de palabras entre ELE y CLM según la variable nivel de estudio

Como podemos observar en el Gráfico 5.22, la diferencia que existe entre los promedios de palabras en ELE y en CLM va disminuyendo a medida que incrementa el nivel de estudio de los informantes. Concretamente, en el primer curso, los alumnos aportan 7.64 palabras menos en ELE que en CLM en un campo semántico medio; pero, mientras van ascendiendo de nivel de estudio, aportan más respuestas en ELE y reducen la diferencia de producción entre las dos lenguas, hasta que, en el cuarto curso, solamente escriben 4.27 palabras menos en ELE que en CLM. Por eso, vemos que es cierto que la productividad léxica en ELE de nuestros informantes se aproxima a la productividad que muestran en CLM a medida que aumenta su nivel de estudio. Ahora, vamos a ver si esta tendencia general también se percibe por centro de interés. Para ello, observaremos el Gráfico 5.23, que resulta ser una visualización de los datos presentados en la Tabla 5.22.

Diferencia de promedio de palabras entre ELE y CLM

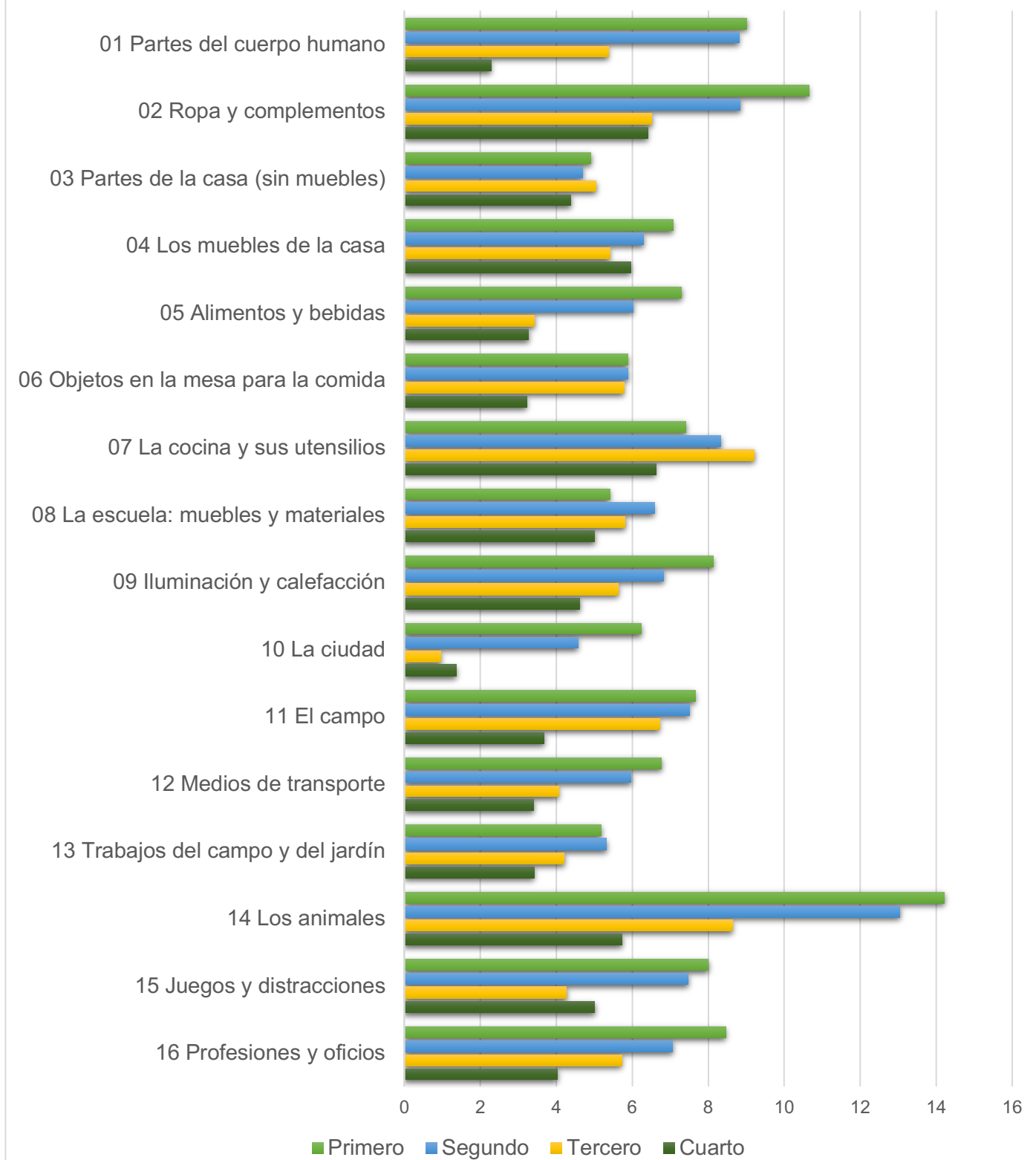


Gráfico 5.23. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable nivel de estudio

En el Gráfico 5.23, se observa que, en casi todas las áreas nocionales —excepto *La cocina y sus utensilios* y *La escuela: sus muebles y materiales*—, las diferencias de producción léxica entre ELE y CLM van reduciéndose desde niveles de estudio bajos hasta los altos. Es decir, en referencia a la productividad léxica que tienen en CLM, los informantes de niveles de estudio altos señalan mejores resultados en ELE que los de niveles bajos. Así que se nos confirma el hecho de que, en la inmensa mayoría de los centros de interés, los estudiantes generan una producción léxica en ELE cada vez más aproximada a la que tienen en CLM, al mismo tiempo que van ascendiendo de nivel de estudio.

En *La escuela: sus muebles y materiales*, dicha tendencia no sirve para el primer curso. Suponemos que es por la causa que hemos dado en un apartado anterior. Repetimos que los estudiantes del primer curso, en general, no tienen un vocabulario abundante por la limitación que supone su fase de estudio y, mientras tanto, sus esfuerzos para aprender nuevos vocablos se centran en su entorno habitual de estudio —*La escuela: sus muebles y materiales*—, porque es la circunstancia comunicativa en la que utilizan con más probabilidad el español. Ello contribuye a la mayor producción léxica de dicho estrato en el centro del que estamos hablando. En cambio, la tendencia casi contraria que se ve en *La cocina y sus utensilios* es realmente curiosa. En esta área léxica, la diferencia de producción léxica entre ELE y CLM es cada vez mayor del primer curso al segundo y, de este, al tercero. Sin embargo, al entrar en el cuarto curso, el cambio parece volver a encajar en algo que ya hemos constatado, puesto que, a pesar del progresivo aumento de palabras en los primeros tres cursos, en el cuarto se ve una disminución drástica que no contradice la deducción hecha.

5.2.1.4. A continuación, vamos a examinar, por nivel de estudio y según la lengua de la encuesta, el total de vocablos que se recaban en cada uno de los centros de interés. Recordamos que este índice no es tan fiable como el promedio de palabras por informante, en que las diferencias por sujeto se ven neutralizadas. Es cierto que el número de vocablos no necesariamente guarda

correlación directa con el número de informantes, como hemos comentado en los análisis hechos sobre los resultados generales; sin embargo, tampoco se sabe en qué sentido la diferencia en el número de informantes puede influir en el número de vocablos. A esto se añade que, en esta sección de la tesis, estamos ante numerosas sub-muestras —por variable extralingüística y por lengua de la encuesta— que no son preestablecidas y que, por tanto, implican entre sí unas diferencias considerables que se dan en el número de informantes. En consecuencia, optamos por analizar, en sentido general, los datos que aporta el número de vocablos para las numerosas sub-muestras; es decir, no realizaremos análisis pormenorizados como hemos hecho con el promedio de palabras por informante.

Centro de interés	Total de vocablos							
	Primero		Segundo		Tercero		Cuarto	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	130	173	118	109	141	130	112	101
02 Ropa	149	228	104	169	158	156	137	144
03 Casa	98	149	95	112	119	99	123	113
04 Muebles	94	178	96	137	117	106	114	124
05 Alimentos	178	419	184	339	215	197	221	278
06 Objetos	66	119	68	96	80	77	106	84
07 Cocina	107	171	115	142	121	120	124	136
08 Escuela	151	211	163	186	195	139	179	163
09 Iluminación	86	157	89	122	125	107	128	120
10 Ciudad	291	460	311	441	339	235	309	279
11 Campo	214	388	249	369	273	229	257	275
12 Transporte	109	177	116	147	164	120	157	132
13 Trabajos	187	243	188	184	216	152	221	181
14 Animales	126	209	100	187	168	143	141	162
15 Juegos	153	231	186	224	212	142	181	194
16 Profesiones	150	255	180	242	232	164	203	201
Global	2289	3768	2362	3206	2875	2316	2713	2687

Tabla 5.23. Total de vocablos por centro de interés en ELE y en CLM según la variable nivel de estudio

En la Tabla 5.23 se expone el número total de vocablos de cada uno de los dieciséis centros de interés según el nivel de estudio y según sea del corpus de ELE o del de CLM. A partir de ahí, podemos darnos cuenta de la riqueza léxica de los informantes de cada curso universitario y en cada una de las dos lenguas contempladas. Ante todo, los resultados globales que se encuentran en la última fila de la tabla nos permiten sacar las siguientes conclusiones. Por un lado, la evolución de la riqueza léxica en ELE de los estudiantes manifiesta, a lo largo de las cuatro fases de estudio, una tendencia igual que la que hemos observado en la productividad léxica en ELE de los mismos. Es decir, va ascendiendo del primer curso (2289) al segundo (2362), y del segundo al

tercero (2875), pero en el cuarto curso (2713), la riqueza léxica tiende a ser estable a la baja, ya que supone una pequeña disminución en la cifra en comparación con el dato del curso anterior. Cuanto más alto sea el nivel de estudio, mayor riqueza léxica alcanzan los estudiantes. En cambio, la evolución que se manifiesta en la riqueza léxica en CLM no parece cumplir nuestras expectativas. Lo que se observa es lo contrario de la tendencia que vemos en ELE: la riqueza léxica en CLM va descendiendo del primer curso (3768) al segundo (3206), y del segundo al tercero (2316), pero en el cuarto curso (2687), la riqueza léxica no continúa disminuyendo, sino que muestra un pequeño repunte, a pesar de que no vuelve al volumen de los primeros dos cursos. Como resultado, nos informa de que, a diferencia de la correlación positiva que se halla entre el aumento de la productividad léxica en ELE y en CLM, la riqueza léxica de los alumnos en las dos lenguas no parece tener las mismas consecuencias a medida que incrementa el nivel de estudio.

Por otro lado, si comparamos, por nivel de estudio, la riqueza léxica en ELE y en CLM, podemos ver que en el primero y segundo curso la riqueza léxica en CLM (3768 y 3206, respectivamente) es abrumadoramente mayor que la que hay en ELE (2289 y 2362, respectivamente); pero, al pasar al tercero y cuarto curso, los estudiantes empiezan a mostrar mejores resultados en ELE (2875 y 2713, respectivamente) que en CLM (2316 y 2687, respectivamente). Ello nos conduce a una deducción similar a la que hemos extraído al analizar la productividad léxica, esto es, a medida que aumenta el nivel de estudio, la diferencia de riqueza léxica entre ELE y CLM se convierte en cada vez menor. Los alumnos de niveles bajos no tienen en ELE una riqueza léxica comparable con la que tienen en CLM; en cambio, los de niveles altos dominan una variedad de vocablos igual o mayor en ELE que en CLM.

Ahora nos centramos en los resultados de las distintas áreas nocionales por nivel de estudio. Como es lógico, el estrato del primer curso manifiesta una variedad léxica menor en ELE que en CLM en todos los centros de interés. Esta tendencia se mantiene en el estrato del segundo curso en general, pero, los alumnos de este curso ya señalan mejores resultados en ELE que en CLM

en *Partes del cuerpo humano* y *Trabajos del campo y del jardín* —aunque muy levemente en esta última área—. Por otra parte, los informantes de tercer curso evidencian una tendencia extremadamente contraria: sus resultados sobre la riqueza léxica en ELE superan en todos los centros de interés a los en CLM. El estrato del cuarto curso es el que tiene más equilibrada la diferencia de riqueza léxica entre ELE y CLM, toda vez que, en más o menos la mitad de los centros —concretamente, en nueve—, los alumnos obtienen mejores datos en ELE que en CLM (*Partes del cuerpo humano*, *Partes de la casa (sin muebles)*, *Objetos colocados en la mesa para la comida*, *La escuela: sus muebles y materiales*, *Iluminación y calefacción*, *La ciudad*, *Medios de transporte*, *Trabajos del campo y del jardín* y *Profesiones y oficios*) y en la otra mitad, al contrario (*Ropa y complementos*, *Los muebles de la casa*, *Alimentos y bebidas*, *La cocina y sus utensilios*, *El campo*, *Los animales* y *Juegos y distracciones*). Vemos, otra vez, cómo los alumnos van alcanzando una mayor riqueza léxica en ELE —con respecto a su competencia léxica en CLM— desde el nivel de estudio más bajo hasta el tercer curso de la universidad, y retroceden ligeramente al llegar al cuarto curso.

Centro de interés	Rango según el total de vocablos							
	Primero		Segundo		Tercero		Cuarto	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	9	12	8	15	11	10	15	15
02 Ropa	8	7	11	9	10	5	10	9
03 Casa	13	15	14	14	14	15	13	14
04 Muebles	14	10	13	12	15	14	14	12
05 Alimentos	4	2	5	3	5	3	3	2
06 Objetos	16	16	16	16	16	16	16	16
07 Cocina	12	13	10	11	13	11	12	10
08 Escuela	6	8	7	7	7	9	7	7
09 Iluminación	15	14	15	13	12	13	11	13
10 Ciudad	1	1	1	1	1	1	1	1
11 Campo	2	3	2	2	2	2	2	3
12 Transporte	11	11	9	10	9	11	8	11
13 Trabajos	3	5	3	8	4	6	3	6
14 Animales	10	9	12	6	8	7	9	8
15 Juegos	5	6	4	5	6	8	6	5
16 Profesiones	7	4	6	4	3	4	5	4

Tabla 5.24. Rango de los centros de interés por nivel de estudio según el total de vocablos en ELE y en CLM

En cuanto al rango de los centros de interés por nivel de estudio según sea su riqueza léxica en las dos lenguas, la Tabla 5.24 nos proporciona informaciones detalladas. Una ojeada a las cifras que aparecen en la tabla nos permite saber que, unánimemente en los cuatro niveles de estudio y en ambas lenguas, *La ciudad* es el área semántica más rica en vocabulario; del mismo modo, *Objetos colocados en la mesa para la comida* es el campo nocional menos rico en vocabulario. Además, *El campo* es el segundo centro de interés más rico en ELE en todos los estratos según el nivel de estudio.

Aparte de dichas áreas semánticas, la Tabla 5.24 informa de que el estrato del primer curso muestra una mayor riqueza léxica en ELE también en *Trabajos*

del campo y del jardín, Alimentos y bebidas y Juegos y distracciones; el segundo curso, en *Trabajos del campo y del jardín, Juegos y distracciones y Alimentos y bebidas*; el tercer curso, en *Profesiones y oficios, Trabajos del campo y del jardín y Alimentos y bebidas*; y el cuarto curso, en *Trabajos del campo y del jardín, Alimentos y bebidas y Profesiones y oficios*. De ahí, vemos que los informantes de los cuatro cursos coinciden en cuatro centros más ricos en vocabulario en ELE: *La ciudad, El campo, Trabajos del campo y del jardín y Alimentos y bebidas*. El primero y el segundo curso también tienen un alto grado de variedad léxica en ELE en *Juegos y distracciones*; mientras el tercero y el cuarto curso, en *Profesiones y oficios*.

Por el contrario, además de *Objetos colocados en la mesa para la comida*, el estrato del primer curso también tiene una menor variedad léxica en ELE en *Iluminación y calefacción, Los muebles de la casa, Partes de la casa (sin muebles) y La cocina y sus utensilios*; el segundo curso, en *Iluminación y calefacción, Partes de la casa (sin muebles), Los muebles de la casa y Los animales*; el tercer curso, en *Los muebles de la casa, Partes de la casa (sin muebles), La cocina y sus utensilios e Iluminación y calefacción*; y finalmente el cuarto curso, en *Partes del cuerpo humano, Los muebles de la casa, Partes de la casa (sin muebles) y La cocina y sus utensilios*. Así, vemos que los informantes de los cuatro cursos coinciden en tres centros menos ricos en vocabulario en ELE: *Objetos colocados en la mesa para la comida, Los muebles de la casa y Partes de la casa (sin muebles)*. Como diferencia, *Iluminación y calefacción* implica una menor riqueza léxica en ELE en tres estratos: el primero, el segundo y el tercer curso; *La cocina y sus utensilios* también en tres estratos: el primero, el segundo y el cuarto curso. Como casos únicos, solo el segundo curso señala una menor riqueza léxica en ELE en *Los animales*, y solo el cuarto curso, en *Partes del cuerpo humano*.

Centro de interés	Índice de cohesión							
	Primero		Segundo		Tercero		Cuarto	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	0.07	0.10	0.10	0.19	0.11	0.16	0.15	0.19
02 Ropa	0.04	0.07	0.09	0.11	0.07	0.11	0.08	0.12
03 Casa	0.07	0.08	0.09	0.12	0.09	0.16	0.08	0.12
04 Muebles	0.08	0.08	0.10	0.11	0.09	0.15	0.09	0.13
05 Alimentos	0.07	0.05	0.08	0.06	0.08	0.11	0.08	0.08
06 Objetos	0.08	0.09	0.07	0.11	0.09	0.17	0.09	0.15
07 Cocina	0.05	0.08	0.05	0.10	0.07	0.14	0.07	0.11
08 Escuela	0.07	0.07	0.06	0.09	0.07	0.14	0.07	0.11
09 Iluminación	0.04	0.07	0.06	0.10	0.06	0.13	0.07	0.11
10 Ciudad	0.04	0.04	0.04	0.04	0.05	0.08	0.05	0.06
11 Campo	0.03	0.04	0.04	0.05	0.05	0.09	0.05	0.06
12 Transporte	0.06	0.07	0.07	0.10	0.07	0.14	0.08	0.12
13 Trabajos	0.03	0.04	0.04	0.07	0.05	0.10	0.05	0.08
14 Animales	0.05	0.10	0.10	0.12	0.09	0.16	0.12	0.14
15 Juegos	0.04	0.06	0.05	0.08	0.06	0.13	0.06	0.09
16 Profesiones	0.06	0.07	0.07	0.08	0.06	0.13	0.07	0.09

Tabla 5.25. Índice de cohesión por centro de interés en ELE y en CLM según la variable nivel de estudio

Por último, presentamos el índice de cohesión, que relaciona los dos índices anteriormente estudiados —el promedio de palabras por informante y el total de vocablos—. La Tabla 5.25 expone el índice de cohesión de cada centro de interés por nivel de estudio y según cuál sea la lengua de la encuesta. De nuevo advertimos de que estamos ante numerosas sub-muestras que implican unas diferencias considerables con respecto al número de informantes. Si nos detenemos a observar las divergencias en dicho índice según el nivel de estudio, o sea, entre los niveles bajos y los altos, podemos ver cómo la cantidad de encuestados influye en los resultados obtenidos en la Tabla 5.25, tanto en el corpus de ELE como en el de CLM. Se observan en general unos índices de cohesión menores en los niveles de estudio bajos que

en los altos. Ello se debe probablemente a que hay menos alumnos encuestados en los niveles altos (véase el apartado 4.4.1). Por tanto, es mayor la posibilidad de que coincidan las respuestas de los informantes de estos niveles altos, lo que contribuye a un mayor valor del índice de cohesión.

Entre las dos lenguas, aunque los valores del índice no son los mismos — lo que quizá sea por causa también de la diferencia en el número de informantes—, sí podemos notar que, en los cuatro estratos de niveles de estudio, generalmente los campos léxicos más compactos en ELE (con valores mayores del índice) también son los más compactos en CLM, y viceversa. Por eso, comprobamos que el carácter compacto o difuso es una cualidad propia de los centros de interés, y no tiene mucho que ver con la lengua en la que los estudiantes realizan las asociaciones semánticas.

5.2.2. Incidencias cuantitativas de la variable “sexo”

El sexo es la segunda variable que vamos a analizar en esta sección de la tesis. Como hemos comentado en el apartado 4.4.2, bajo la variable sexo, vamos a confrontar los resultados obtenidos entre los varones y los obtenidos entre las mujeres. El cotejo entre dichos resultados será interesante, porque, como hemos mencionado a lo largo del capítulo del estado de la cuestión, el sexo es la variable ampliamente analizada en los estudios de disponibilidad léxica, unos afirman que se observa una superioridad en la competencia léxica de las mujeres frente a los varones y otros explican que el sexo no ejerce incidencias significativas en los resultados del léxico disponible. Emprendemos los siguientes apartados con la expectativa de comprobar si la influencia del sexo en nuestras investigaciones coincide con una de dichas conclusiones.

Cabe recordar que, aunque las mujeres suponen una inmensa mayoría en ambos corpus —de ELE y de CLM— de nuestra investigación (véase el apartado 4.4.2), la distribución cuantitativa de los dos sexos representa la situación real de las aulas de ELE en Tianjin, toda vez que es habitual que haya más chicas que chicos que optan por la especialidad universitaria de lengua

extranjera.

5.2.2.1. Para empezar, en la Tabla 5.26, presentamos, por centro de interés, el promedio de palabras por informante del corpus en ELE según sean varones o mujeres. Después de esta tabla, está la Tabla 5.27, que nos permite ver el rango de los centros de interés en cada uno de los dos sexos según el promedio de palabras por informante en ELE, de modo que se perciba más claramente el mayor o menor grado del valor de las cifras expuestas en la Tabla 5.26.

Centro de interés	Promedio de palabras en ELE	
	Varón	Mujer
01 Partes del cuerpo humano	13.00	11.91
02 Ropa y complementos	8.57	8.63
03 Partes de la casa (sin muebles)	8.17	8.44
04 Los muebles de la casa	8.72	8.98
05 Alimentos y bebidas	14.63	15.25
06 Objetos en la mesa para la comida	6.57	5.90
07 La cocina y sus utensilios	6.35	6.69
08 La escuela: muebles y materiales	10.85	11.20
09 Iluminación y calefacción	5.83	5.54
10 La ciudad	13.12	13.82
11 El campo	9.30	10.29
12 Medios de transporte	9.80	8.81
13 Trabajos del campo y del jardín	7.43	7.70
14 Los animales	10.73	10.49
15 Juegos y distracciones	9.15	9.59
16 Profesiones y oficios	10.80	11.78
Global	9.56	9.69

Tabla 5.26. Promedio de palabras por informante por centro de interés en ELE según la variable sexo

Centro de interés	Rango según el promedio de palabras por informante en ELE	
	Varón	Mujer
01 Partes del cuerpo humano	3	3
02 Ropa y complementos	11	11
03 Partes de la casa (sin muebles)	12	12
04 Los muebles de la casa	10	9
05 Alimentos y bebidas	1	1
06 Objetos en la mesa para la comida	14	15
07 La cocina y sus utensilios	15	14
08 La escuela: muebles y materiales	4	5
09 Iluminación y calefacción	16	16
10 La ciudad	2	2
11 El campo	8	7
12 Medios de transporte	7	10
13 Trabajos del campo y del jardín	13	13
14 Los animales	6	6
15 Juegos y distracciones	9	8
16 Profesiones y oficios	5	4

Tabla 5.27. Rango de los centros de interés por sexo según el promedio de palabras por informante en ELE

Si echamos un vistazo a la Tabla 5.26, podemos ver que la mayor producción léxica en ELE de todos los centros de interés y de ambos sexos se detecta en *Alimento y bebidas* del estrato de las chicas, que ostenta una media de 15.25 palabras. Por el contrario, la menor producción léxica en ELE se halla en *Iluminación y calefacción* de las mujeres, con solamente un promedio de 5.54 palabras.

A continuación, vamos a analizar por sexo el promedio de palabras por informante en ELE. Como podemos comprobar en la Tabla 5.27, los chicos y las chicas coinciden totalmente en los primeros cinco campos semánticos con una mayor producción léxica, los cuales son, por orden de más a menos,

Alimento y bebidas, La ciudad, Partes del cuerpo humano, La escuela: sus muebles y materiales y Profesiones y oficios. Con respecto a estos campos más productivos en ELE, la única y ligera diferencia que se observa es que *La escuela: sus muebles y materiales y Profesiones y oficios* alternan sus posiciones en los dos estratos según el sexo: de los varones a las mujeres, *La escuela: sus muebles y materiales* desciende de la cuarta posición a la quinta, mientras *Profesiones y oficios* asciende de la quinta a la cuarta. Por eso, llegamos a ver que el sexo no influye significativamente en cuanto a cuáles constituyen las áreas nocionales más productivas en ELE.

En lo que se refiere a los campos semánticos menos productivos en ELE, observamos que, igualmente, los chicos y las chicas coinciden totalmente en los cinco primeros campos semánticos con una menor producción léxica, los cuales son *Iluminación y calefacción, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Trabajos del campo y del jardín y Partes de la casa (sin muebles)*. En relación con estos campos menos productivos en ELE, la única y ligera diferencia que se percibe es que *Objetos colocados en la mesa para la comida y La cocina y sus utensilios* alternan sus posiciones en los dos estratos según el sexo: de los varones a las mujeres, *Objetos colocados en la mesa para la comida* desciende de la decimocuarta posición a la decimoquinta, mientras *La cocina y sus utensilios* asciende de la decimoquinta a la decimocuarta. Por tanto, deducimos que el sexo tampoco influye significativamente en cuanto a cuáles constituyen las áreas nocionales menos productivas en ELE.

Si nos fijamos en el rango de los demás centros de interés, nos damos cuenta de que, en realidad, los varones y las mujeres coinciden en la mayoría de los centros, con respecto al mayor o menor grado de productividad léxica en ELE. Solo en *Medios de transporte*, se evidencia una mayor diferencia entre los dos sexos: los chicos (la séptima posición) parecen manejar mejor dicha área nocional que las chicas (décima posición), tomando como referencia la producción que tienen en las demás áreas. Así, llegamos a la conclusión de que el sexo no ejerce notoria incidencia en la relación de mayor o menor grado

de productividad léxica en ELE de los centros de interés.

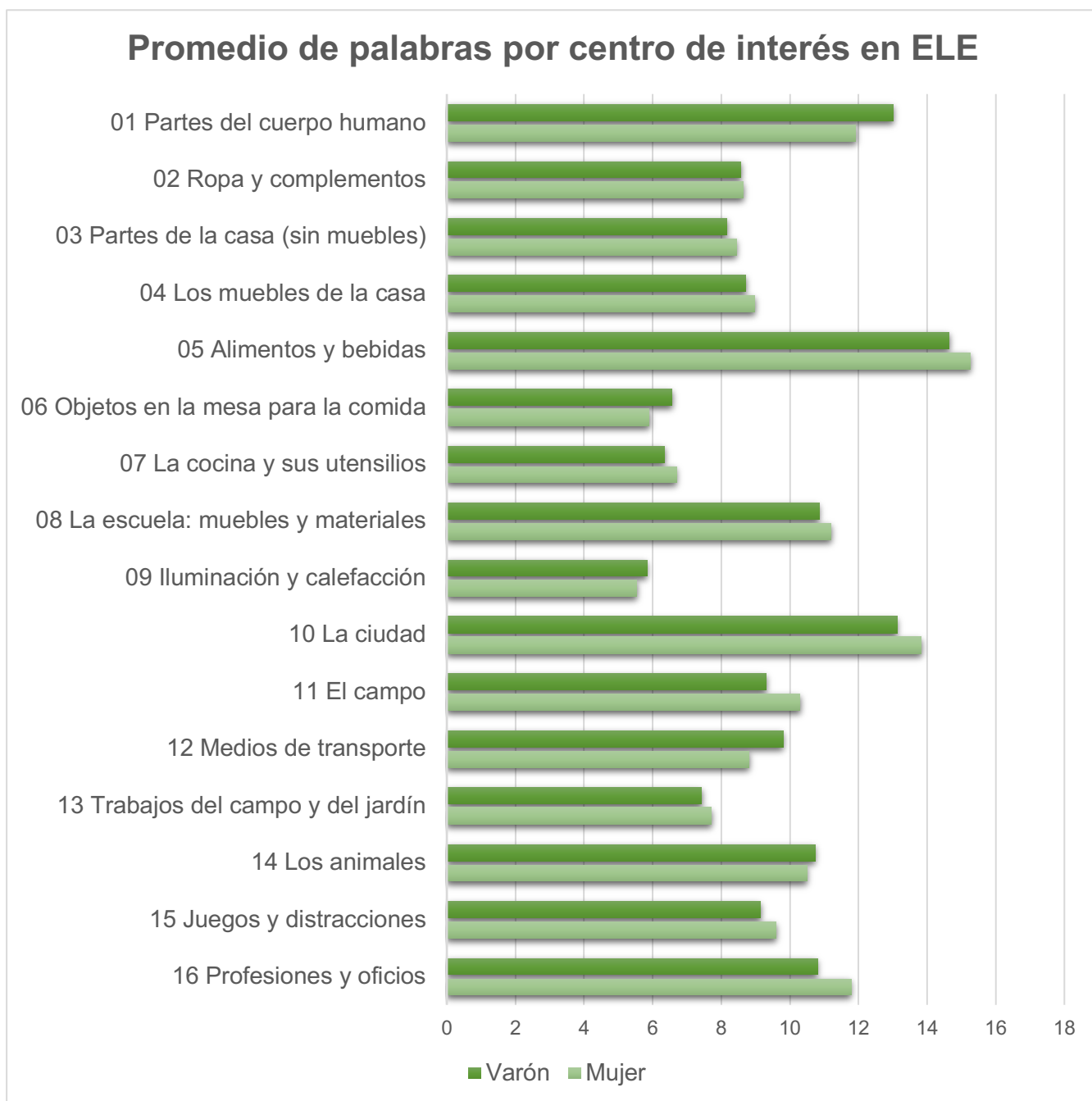


Gráfico 5.24. Promedio de palabras por informante por centro de interés en ELE según la variable sexo

Ahora nos centramos en el Gráfico 5.24, que muestra las diferencias entre los varones y las mujeres con respecto al promedio de palabras por informante en ELE en cada uno de los centros de interés. Como podemos ver, los chicos señalan mejores resultados de producción léxica en ELE que las chicas en cinco campos asociativos: *Partes del cuerpo humano*, *Objetos colocados en la*

mesa para la comida, Iluminación y calefacción, Medios de transporte y Los animales. Por eso, hay once campos semánticos en los que las chicas tienen producción léxica en ELE superior a los chicos: Ropa y complementos, Partes de la casa (sin muebles), Los muebles de la casa, Alimentos y bebidas, La cocina y sus utensilios, La escuela: sus muebles y materiales, La ciudad, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios. Parece que, en la mayoría de las áreas nocionales, las mujeres muestran mayor productividad léxica en ELE que los varones. A continuación, pasamos a analizar si dichas diferencias cuantitativas entre los dos sexos son significativas o, al contrario, muy leves.

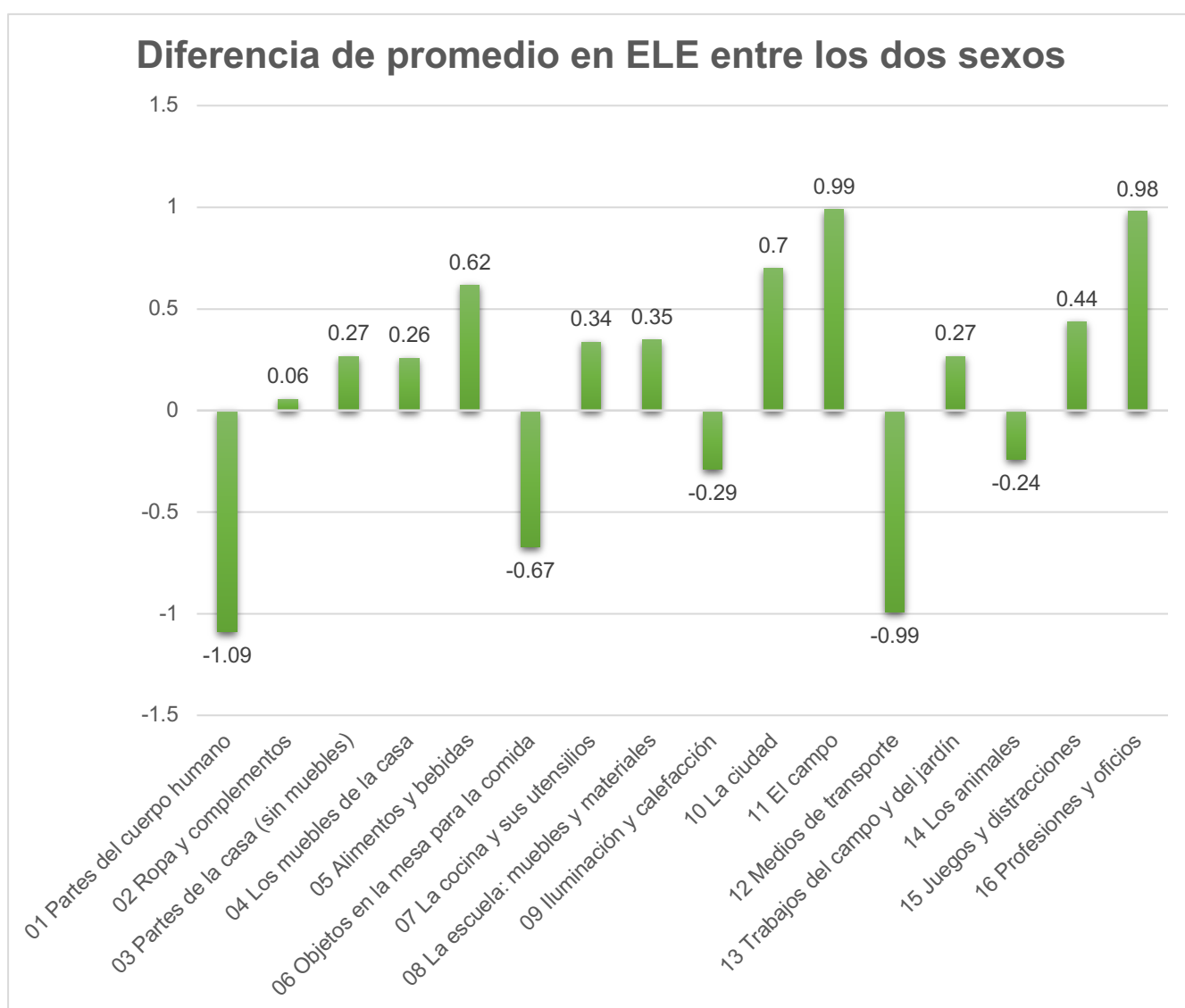


Gráfico 5.25. Diferencia de promedio de palabras en ELE entre los dos sexos por centro de interés

El Gráfico 5.25 exhibe, por centro de interés, las diferencias cuantitativas que se dan entre los varones y las mujeres con respecto al promedio de palabras por informante en ELE. Se elabora a partir de los datos de la Tabla 5.26 y con respecto a los resultados de las mujeres, es decir, las diferencias resultan de restar las cifras de los varones de las de las mujeres. Como podemos ver, la mayor divergencia se halla en *Partes del cuerpo humano* (-1.09), donde los chicos aportan 1.09 palabras más que las chicas en promedio.

Partes del cuerpo humano es la única área nocional en la que la diferencia cuantitativa supera una palabra. En cambio, como demuestra el gráfico, en muchos de los centros de interés, las diferencias cuantitativas que implican la superioridad de un sexo sobre el otro son bastante leves, con una diferencia menor de 0.5 palabras. Los demás centros que señalan una diferencia de entre 0.5 y 1 palabra son los siguientes: *Alimentos y bebidas* (0.62), *La ciudad* (0.7), *El campo* (0.99) y *Profesiones y oficios* (0.98), en los que las mujeres cuentan con mayor producción léxica en ELE que los varones; y *Objetos colocados en la mesa para la comida* (-0.67) y *Medios de transporte* (-0.99), en los que los varones tienen mejores resultados que las mujeres.

De todos modos, no creemos que el sexo tenga una influencia significativa en la productividad léxica de nuestros informantes, dado que, en todos los centros de interés, la divergencia en la producción de respuestas entre los varones y las mujeres apenas supera una palabra.

5.2.2.2. De aquí en adelante, exponemos, como referencia de comparación, los resultados sobre el promedio de palabras por informante del corpus en CLM según sean varones o mujeres. Del mismo modo como hemos hecho en los párrafos previos sobre ELE, presentamos la Tabla 5.28, en la que se ve dicho índice en CLM en cada centro de interés por sexo. Al mismo tiempo, en la Tabla 5.29 se puede comprobar el rango de los centros por sexo según el promedio de palabras por informante en CLM.

Centro de interés	Promedio de palabras en CLM	
	Varón	Mujer
01 Partes del cuerpo humano	21.21	18.86
02 Ropa y complementos	16.54	17.41
03 Partes de la casa (sin muebles)	13.21	12.98
04 Los muebles de la casa	14.21	15.45
05 Alimentos y bebidas	21.58	20.52
06 Objetos en la mesa para la comida	10.63	11.49
07 La cocina y sus utensilios	14.63	14.22
08 La escuela: muebles y materiales	16.58	16.75
09 Iluminación y calefacción	12.42	12.21
10 La ciudad	20.13	17.17
11 El campo	15.38	16.70
12 Medios de transporte	14.29	14.39
13 Trabajos del campo y del jardín	13.33	12.08
14 Los animales	23.04	21.70
15 Juegos y distracciones	16.67	15.95
16 Profesiones y oficios	18.17	18.24
Global	16.38	16.01

Tabla 5.28. Promedio de palabras por informante por centro de interés en CLM según la variable sexo

Centro de interés	Rango según el promedio de palabras por informante en CLM	
	Varón	Mujer
01 Partes del cuerpo humano	3	3
02 Ropa y complementos	8	5
03 Partes de la casa (sin muebles)	14	13
04 Los muebles de la casa	12	10
05 Alimentos y bebidas	2	2
06 Objetos en la mesa para la comida	16	16
07 La cocina y sus utensilios	10	12
08 La escuela: muebles y materiales	7	7
09 Iluminación y calefacción	15	14
10 La ciudad	4	6
11 El campo	9	8
12 Medios de transporte	11	11
13 Trabajos del campo y del jardín	13	15
14 Los animales	1	1
15 Juegos y distracciones	6	9
16 Profesiones y oficios	5	4

Tabla 5.29. Rango de los centros de interés por sexo según el promedio de palabras por informante en CLM

Un vistazo a las cifras expuestas en la Tabla 5.28 nos permite saber que la mayor producción léxica en CLM de todos los centros de interés de los dos sexos se encuentra en *Los animales* del estrato de los varones —con un promedio de 23.04 palabras—. Por el contrario, la menor producción léxica en CLM se halla en *Objetos colocados en la mesa para la comida* de los varones también —con una media de 10.63 palabras—. Estos resultados en CLM no coinciden con los resultados previamente extraídos en ELE¹³⁵, ya que en esta lengua tanto el mayor como el menor promedio de palabras por informante están en el estrato de las mujeres.

¹³⁵ Véase el análisis que sigue las Tablas 5.26 y 5.27 del apartado 5.2.2.1 para consultar los resultados previamente sacados en ELE.

Ahora, vamos a analizar el promedio de palabras por informante en CLM en cada uno de los sexos. Como podemos ver en la Tabla 5.29, los chicos y las chicas coinciden totalmente en los primeros tres campos semánticos con una mayor producción léxica en CLM, los cuales son *Los animales*, *Alimentos y bebidas* y *Partes del cuerpo humano*. *Profesiones y oficios* es otro centro de interés con una mayor producción léxica tanto en los chicos como en las chicas, aunque no coinciden en el rango. Cabe destacar que, comparando estos resultados con los previamente extraídos en ELE según el sexo, vemos que, en ambos sexos, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son los centros de interés comunes en ELE y en CLM en cuanto a su carácter de alto grado de productividad léxica en el conjunto de los dieciséis centros. Comprobamos otra vez —esta vez desde la perspectiva de los dos sexos— cómo las áreas semánticas más productivas en CLM resultan ser las más productivas también en ELE.

En el resto de áreas nocionales más productivas en CLM, se observa diferencia entre cada sexo. *La ciudad* está entre los campos asociativos más productivos en CLM entre los varones; mientras que *Ropa y complementos* lo es en CLM entre las mujeres.

En lo que se refiere a los campos semánticos menos productivos en CLM, observamos que los chicos y las chicas coinciden en el campo semántico menos productivo: *Objetos colocados en la mesa para la comida*. Aparte de este centro de interés, el estrato de los varones tiene los menores promedios de palabras, en este orden, de menos a más, en *Iluminación y calefacción*, *Partes de la casa (sin muebles)*, *Trabajos del campo y del jardín* y *Los muebles de la casa*. Mientras tanto, el estrato de las mujeres tiene una menor producción de palabras en *Trabajos del campo y del jardín*, *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*. Vemos que los varones y las mujeres coinciden, además de en el centro menos productivo, también en tres centros con una menor producción léxica: *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *Trabajos del*

campo y del jardín. Comparando estos resultados con los previamente extraídos en ELE según el sexo, estas cuatro áreas nocionales menos productivas en CLM, comunes para los chicos y las chicas, también se hallan entre las menos productivas en ELE. Ahora comprobamos —desde la perspectiva de la variable del sexo— cómo se relacionan la competencia de los estudiantes en ELE y en CLM.

Por último, señalamos que la diferencia entre los dos sexos, con respecto a los campos asociativos menos productivos en CLM de cada uno, consiste en que los chicos muestran resultados deficientes en *Los muebles de la casa*, mientras que las chicas lo hacen en *La cocina y sus utensilios*.

Promedio de palabras por centro de interés en CLM

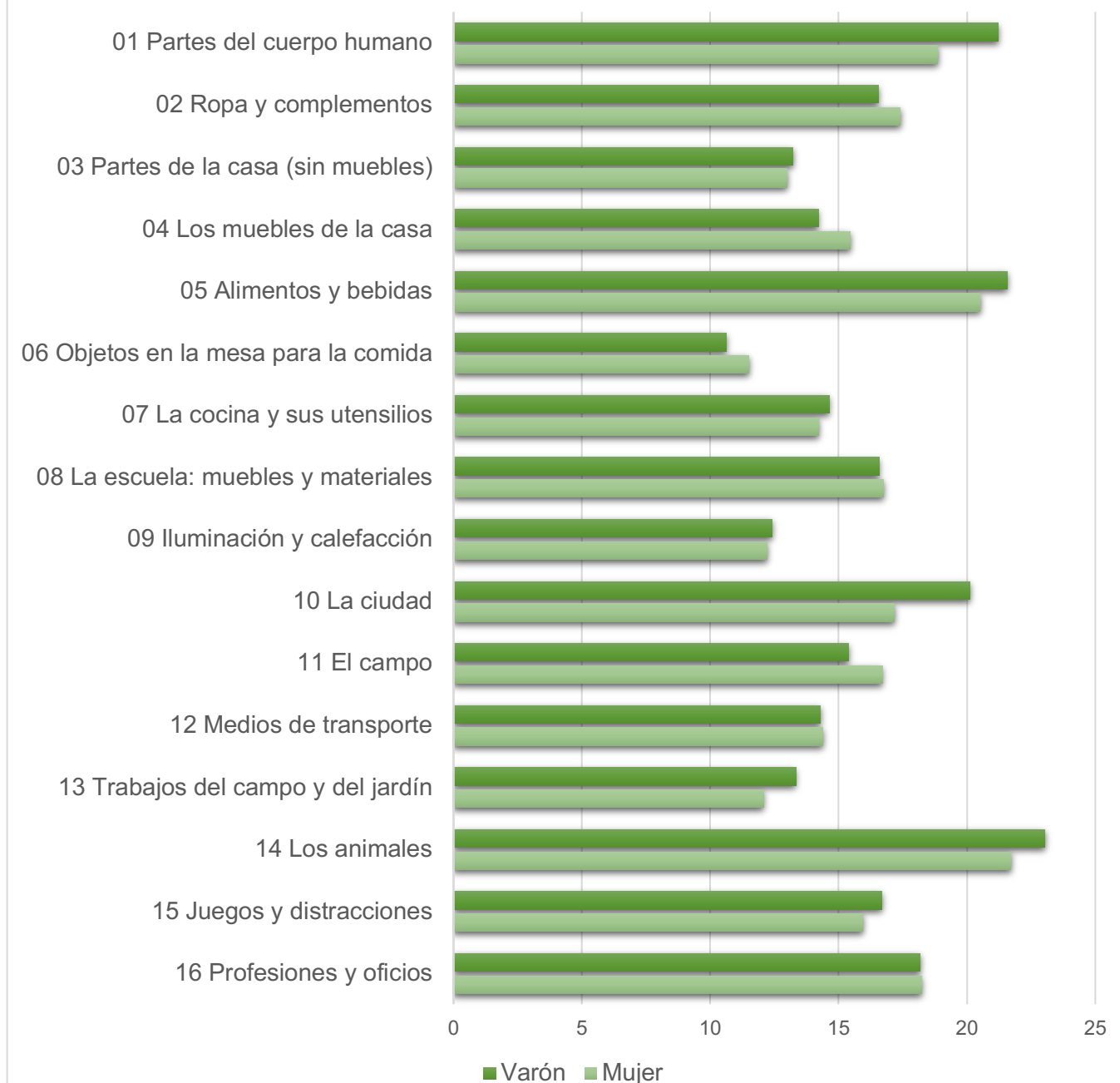


Gráfico 5.26. Promedio de palabras por informante por centro de interés en CLM según la variable sexo

Si nos fijamos en el Gráfico 5.26, que expone las diferencias entre los varones y las mujeres con respecto a la producción léxica en CLM en cada uno de centros de interés, podemos ver que no existe una superioridad general de un sexo sobre el otro. Concretamente, los chicos señalan mejores resultados de producción léxica en CLM que las chicas en nueve campos asociativos:

Partes del cuerpo humano, Partes de la casa (sin muebles), Alimentos y bebidas, La cocina y sus utensilios, Iluminación y calefacción, La ciudad, Trabajos del campo y del jardín, Los animales y Juegos y distracciones. Por tanto, en el resto de campos semánticos —siete— las chicas tienen producción léxica en CLM superior a los chicos: *Ropa y complementos, Los muebles de la casa, Objetos colocados en la mesa para la comida, La escuela: sus muebles y materiales, El campo, Medios de transporte y Profesiones y oficios.* A continuación, pasamos a analizar si dichas diferencias cuantitativas entre los varones y las mujeres son significativas o, al contrario, muy leves.

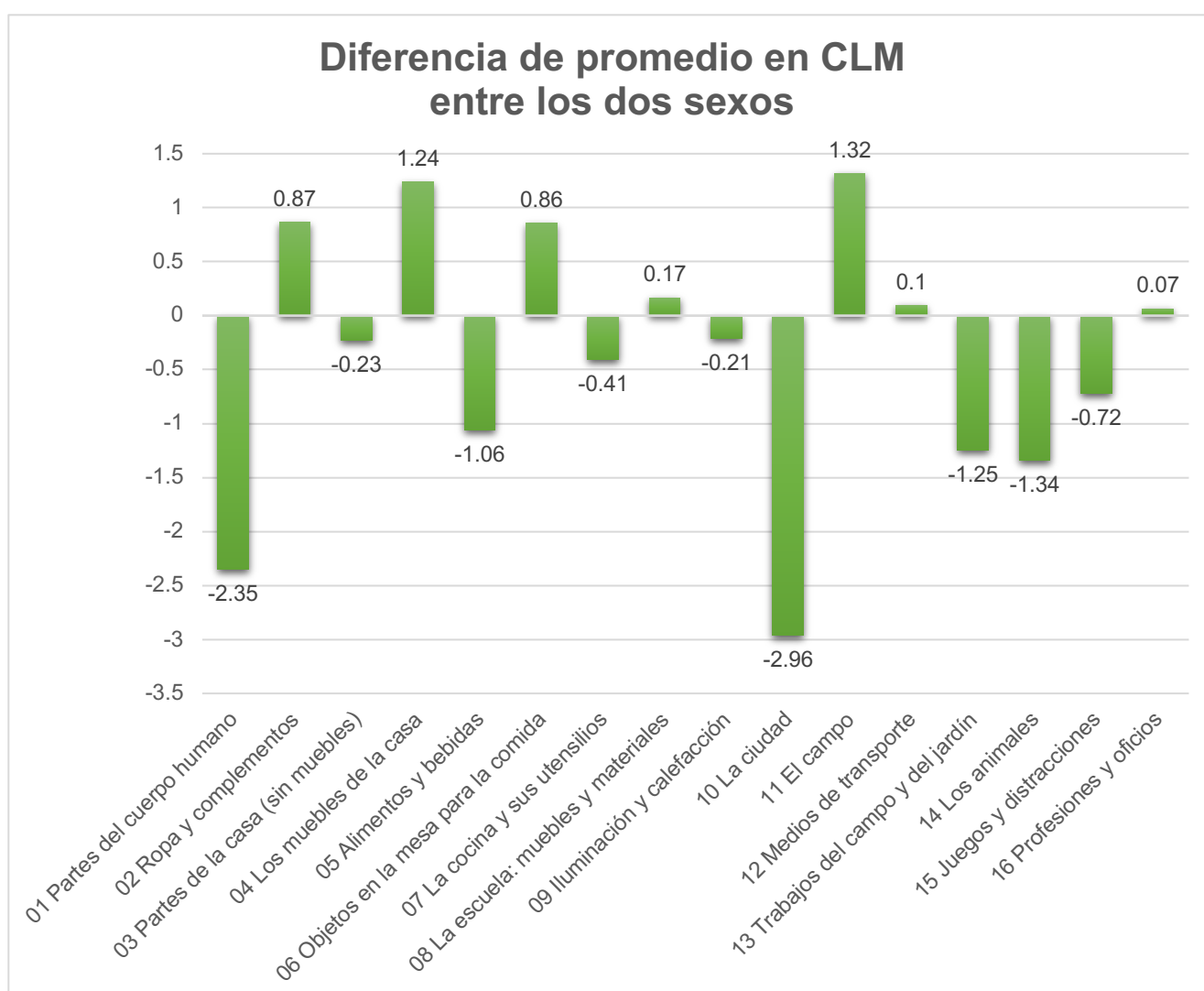


Gráfico 5.27. Diferencia de promedio de palabras en CLM entre los dos sexos por centro de interés

El Gráfico 5.27 exhibe, por centro de interés, las diferencias cuantitativas

que se dan entre los varones y las mujeres con respecto al promedio de palabras por informante en CLM. Está elaborado a partir de los datos de la Tabla 5.28 y con respecto a los resultados de las mujeres, es decir, las diferencias proceden de restar las cifras de los varones de las de las mujeres. Como podemos ver, la mayor divergencia se encuentra en *La ciudad* (-2.96), donde los chicos aportan 2.96 palabras en CLM más que las chicas en promedio.

Hay dos centros de interés en los que la diferencia de promedio supera dos palabras: *La ciudad* (-2.96), que acabamos de ver, y *Partes del cuerpo humano* (-2.35). Las áreas nocionales donde dicha diferencia está entre una y dos palabras son cinco: *Los muebles de la casa* (1.24), *Alimentos y bebidas* (-1.06), *El campo* (1.32), *Trabajos del campo y del jardín* (-1.25) y *Los animales* (-1.34). El resto de centros de interés —nueve— presentan una superioridad bastante leve de un sexo sobre el otro que solamente supone menos de una palabra. Como podemos ver, en los siete campos léxicos donde la divergencia de producción léxica es significativa —por encima de una palabra—, los chicos manifiestan mejores resultados en CLM que las chicas en cinco campos —la gran mayoría—. Por eso, deducimos que los varones disponen, en general, de mayor productividad léxica en CLM que las mujeres, aunque la superioridad no resulta del todo evidente.

A continuación, a partir de los datos globales expuestos en las Tablas 5.26 y 5.28, presentamos en los Gráficos 5.28 y 5.29 el promedio de palabras global de los varones y de las mujeres en los corpus de ELE y de CLM.

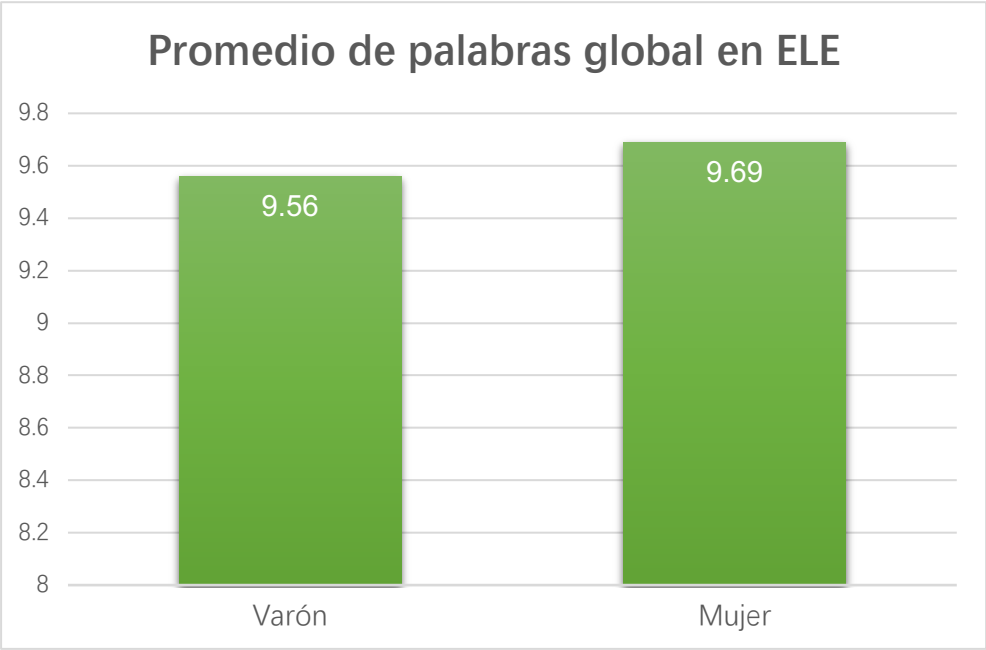


Gráfico 5.28. Promedio global de palabras por informante en ELE según la variable sexo

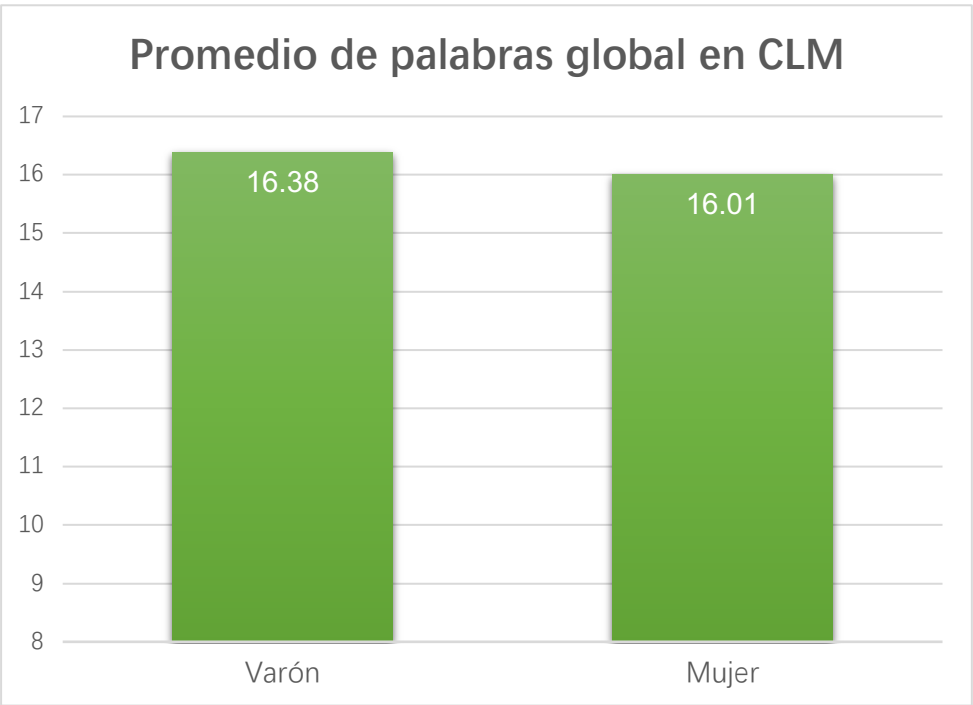


Gráfico 5.29. Promedio global de palabras por informante en CLM según la variable sexo

Como muestran los dos gráficos de arriba, vemos comprobada la conclusión anteriormente extraída por centro de interés sobre la ventaja de

productividad léxica de un sexo sobre el otro, pero, esta vez, de manera global e independientemente de los campos asociativos. En el gráfico 5.28, podemos observar que los varones y las mujeres producen un promedio global de palabras en ELE casi igual (9.56 en los varones frente a 9.69 en las mujeres). La diferencia de la producción léxica global en ELE entre los chicos y las chicas es de solo 0.13. Apenas podemos decir que la variable sexo ejerza alguna incidencia significativa en la productividad léxica de nuestros informantes. En cambio, la diferencia del promedio de palabras global en CLM entre los dos sexos es un poco mayor que la que se da en ELE. Como podemos ver en el gráfico 5.29, dicha diferencia en CLM es de 0.37 palabra (16.38 en los varones y 16.01, en las mujeres). Se trata de una distinción cuantitativa realmente pequeña. Así pues, tal como hemos comentado en los párrafos anteriores, los chicos disponen de mayor productividad léxica en CLM que las mujeres, pero la ventaja resulta ser prácticamente nula.

Otro fenómeno que podemos percibir comparando los gráficos 5.28 y 5.29 entre ellos mismos es que, tanto en el estrato de los varones como en el de las mujeres, la producción léxica en ELE dista enormemente de la que se muestra en CLM (9.56 frente a 16.38 en los varones y 9.69 frente a 16.01 en las mujeres). En el siguiente apartado, vamos a examinar más concretamente la distancia de producción léxica entre las dos lenguas según el sexo del informante.

5.2.2.3. La Tabla 5.30 reúne los datos de las Tablas 5.26 y 5.28, presentando, a su vez, los resultados completos sobre el promedio de palabras por informante por sexo y en las dos lenguas de nuestra investigación. A partir de ella, obtenemos la Tabla 5.31, que muestra, por centro de interés y por nivel de estudio, las diferencias cuantitativas entre el promedio en ELE y el promedio en CLM.

Centro de interés	Promedio de palabras			
	Varón		Mujer	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	13.00	21.21	11.91	18.86
02 Ropa y complementos	8.57	16.54	8.63	17.41
03 Partes de la casa (sin muebles)	8.17	13.21	8.44	12.98
04 Los muebles de la casa	8.72	14.21	8.98	15.45
05 Alimentos y bebidas	14.63	21.58	15.25	20.52
06 Objetos en la mesa para la comida	6.57	10.63	5.90	11.49
07 La cocina y sus utensilios	6.35	14.63	6.69	14.22
08 La escuela: muebles y materiales	10.85	16.58	11.20	16.75
09 Iluminación y calefacción	5.83	12.42	5.54	12.21
10 La ciudad	13.12	20.13	13.82	17.17
11 El campo	9.30	15.38	10.29	16.70
12 Medios de transporte	9.80	14.29	8.81	14.39
13 Trabajos del campo y del jardín	7.43	13.33	7.70	12.08
14 Los animales	10.73	23.04	10.49	21.70
15 Juegos y distracciones	9.15	16.67	9.59	15.95
16 Profesiones y oficios	10.80	18.17	11.78	18.24
Global	9.56	16.38	9.69	16.01

Tabla 5.30. Promedio de palabras por informante por centro de interés en ELE y en CLM según la variable sexo

Centro de interés	Diferencia de promedio de palabras entre ELE y CLM	
	CLM	
	Varón	Mujer
01 Partes del cuerpo humano	8.21	6.95
02 Ropa y complementos	7.97	8.78
03 Partes de la casa (sin muebles)	5.04	4.54
04 Los muebles de la casa	5.49	6.47
05 Alimentos y bebidas	6.95	5.27
06 Objetos en la mesa para la comida	4.06	5.59
07 La cocina y sus utensilios	8.28	7.53
08 La escuela: muebles y materiales	5.73	5.55
09 Iluminación y calefacción	6.59	6.67
10 La ciudad	7.01	3.35
11 El campo	6.08	6.41
12 Medios de transporte	4.49	5.58
13 Trabajos del campo y del jardín	5.90	4.38
14 Los animales	12.31	11.21
15 Juegos y distracciones	7.52	6.36
16 Profesiones y oficios	7.37	6.46
Global	6.82	6.32

Tabla 5.31. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable sexo

Como es esperable, el promedio de palabras en ELE es inferior al mismo promedio en CLM en todos los centros de interés y en ambos sexos. Podemos ver en la Tabla 5.31 que la mayor diferencia entre los promedios de ELE y de CLM se halla en *Los animales* en el estrato de los varones (12.31, en favor de la CLM). Cabe destacar que la segunda mayor diferencia también se registra en este centro de interés, en las mujeres (11.21, en beneficio de la CLM).

Las celdas marcadas en fondo gris en la Tabla 5.22 indican aquellas áreas nocionales en las que la diferencia cuantitativa está por encima del promedio global o, en otras palabras, las áreas en las que la producción léxica en ELE

de los informantes dista enormemente de la que se da en CLM. Hay que prestar atención a estas áreas semánticas, toda vez que la inferioridad en ELE en ellas puede compensarse en el aula con estrategias para que puedan adquirir los homólogos en español de las palabras que los estudiantes son capaces de evocar en CLM.

Existen seis centros de interés en los que tanto los chicos como las chicas señalan gran divergencia de producción léxica entre las dos lenguas: *Partes del cuerpo humano, Ropa y complementos, La cocina y sus utensilios, Los animales, Juegos y distracciones y Profesiones y oficios*. Así es que todos los estudiantes, independientemente del sexo, muestran considerable debilidad en estos seis centros en ELE, tomando como referencia su producción léxica en CLM. A este respecto, también indicamos que solo los varones tienen mucha deficiencia en ELE en *Alimentos y bebidas y La ciudad*. Únicamente en el estrato de las mujeres, son *Los muebles de la casa, Iluminación y calefacción y El campo* en los que la deficiencia de la producción en ELE se ve considerable con respecto a la producción en CLM.

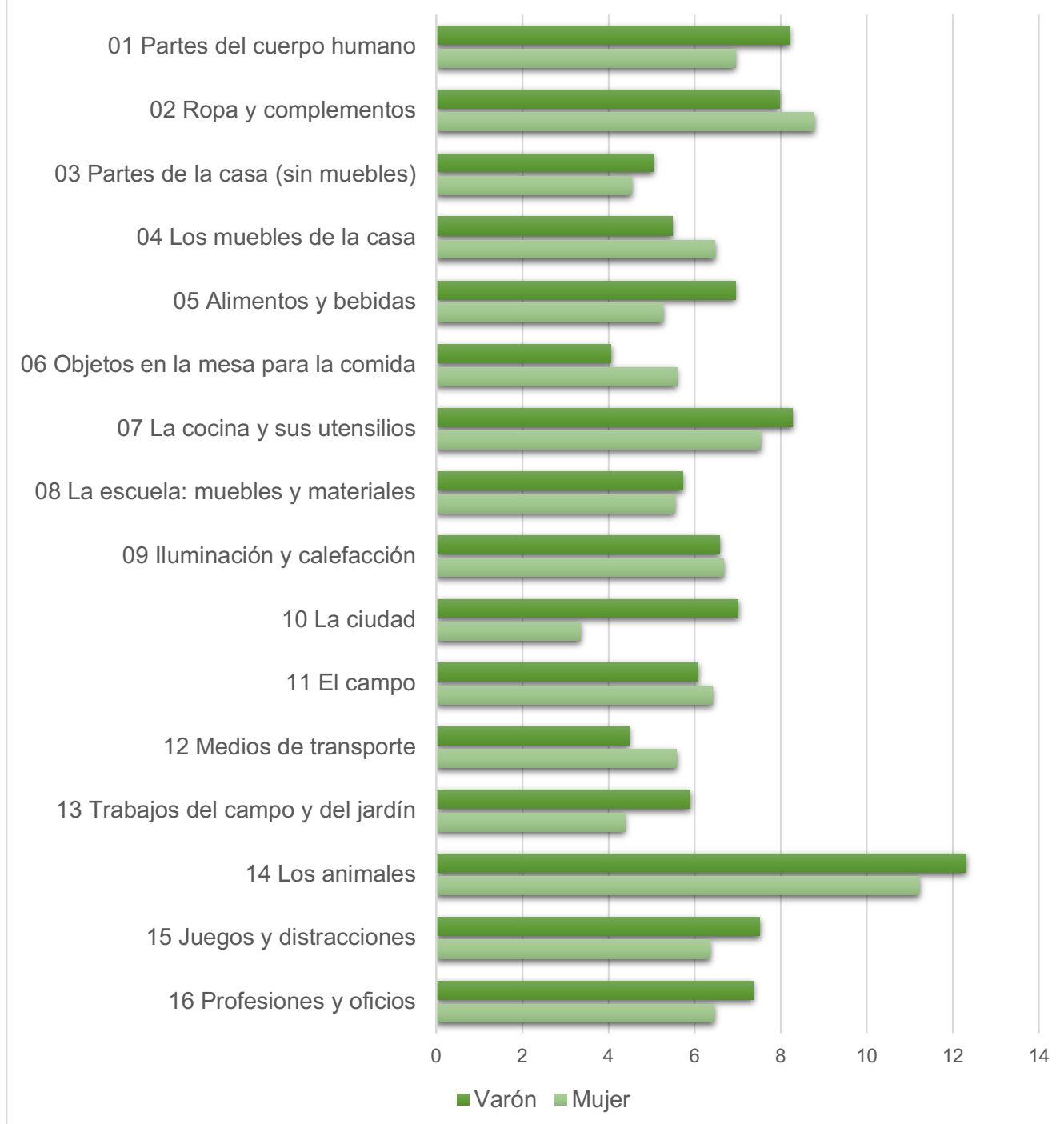


Gráfico 5.30. Diferencia global de promedio de palabras entre ELE y CLM según la variable sexo

Si nos fijamos en la diferencia global de promedio de palabras entre ELE y CLM por sexo —la que se expone en la última fila de la Tabla 5.31 y, de manera ilustrativa, en el Gráfico 5.30—, podemos saber, en ambos sexos, cómo la producción léxica en una lengua dista globalmente de la producción en la otra. Tal y como muestra el Gráfico 5.30, los chicos aportan 6.82 palabras menos en ELE que en CLM en un campo semántico medio; mientras que las chicas escriben 6.32 palabras menos en ELE que en CLM. Notamos una diferencia de 0.5 —poco considerable— entre los dos sexos. De modo que llegamos a ver que, globalmente, el grado de distancia entre la producción léxica en ELE y la propia en CLM es más o menos la misma en los chicos que en las chicas. El sexo tiene una influencia muy pequeña con relación a la diferencia de productividad léxica entre las dos lenguas.

Ahora, vamos a ver si dicho grado de distancia globalmente parecido en los dos sexos también se percibe por centro de interés. Para ello, observaremos el Gráfico 5.31, que resulta ser una visualización de los datos presentados en la Tabla 5.31.

Diferencia de promedio de palabras entre ELE y CLM



*Gráfico 5.31. Diferencia de promedio de palabras entre ELE y CLM
por centro de interés según la variable sexo*

A partir del gráfico 5.31, nos damos cuenta de que los chicos muestran una mayor distancia de producción léxica entre ELE y CLM que las chicas en diez centros de interés: *Partes del cuerpo humano, Partes de la casa (sin*

muebles), Alimentos y bebidas, La cocina y sus utensilios, La escuela: sus muebles y materiales, La ciudad, Trabajos del campo y del jardín, Los animales, Juegos y distracciones y Profesiones y oficios. Por el contrario, las chicas tienen una mayor diferencia de producción léxica entre ELE y CLM que los chicos en seis centros de interés: Ropa y complementos, Los muebles de la casa, Objetos colocados en la mesa para la comida, Iluminación y calefacción, El campo y Medios de transporte.

A continuación, pasamos a analizar si dichas diferencias cuantitativas entre ELE y CLM que arrojan los varones son afines o dispares de las diferencias que generan las mujeres.

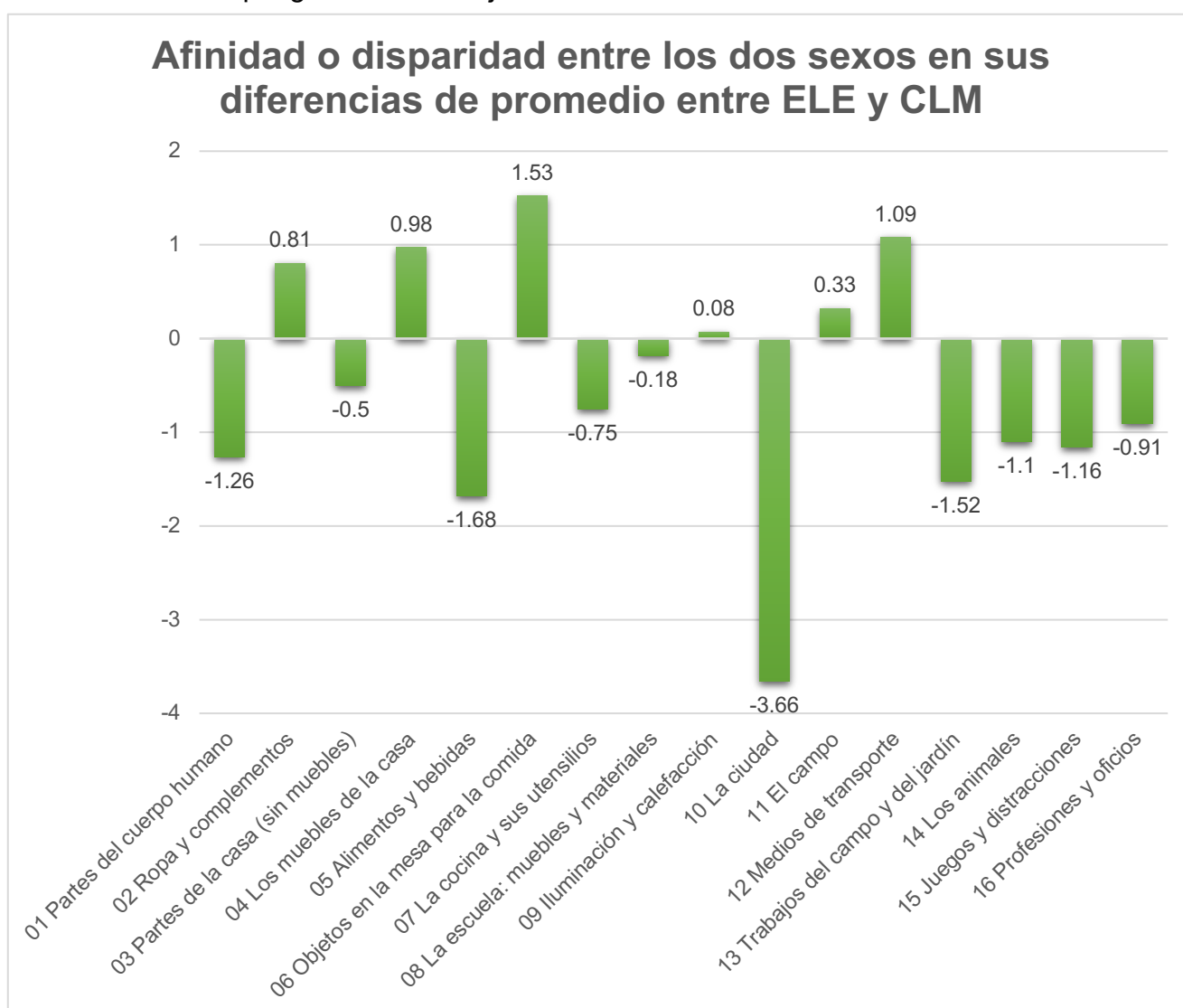


Gráfico 5.32. Afinidad o disparidad entre los dos sexos en sus diferencias de promedio de palabras entre ELE y CLM

El Gráfico 5.32 manifiesta la diferencia entre la barra de los varones y la barra de las mujeres en cada centro de interés del Gráfico 5.31. En otras palabras, los valores en el Gráfico 5.32 resultan ser la resta (con respecto a las cifras de las mujeres) de los datos de cada centro de interés expuestos en la Tabla 5.31. Así pues, nos informa de si son cuantitativamente afines o dispares las diferencias de producción léxica entre ELE y CLM de los varones y las diferencias de las mujeres. Tal como presenta el Gráfico 5.32, el valor 0 del eje de ordenadas indica que dicha diferencia por parte de los chicos en un centro de interés es cuantitativamente lo mismo que la diferencia en ese centro por parte de las chicas, y, por eso, cuanto más se aleje la barra del eje de abscisas —o lo que es lo mismo, cuanto mayor sea el valor absoluto de la barra—, mayor disparidad se da.

Como podemos ver, la mayor afinidad se da en *Iluminación y calefacción* (0.08), donde las mujeres muestran 0.08 palabras más que los varones en cuanto a la diferencia de promedio de palabras entre las dos lenguas. Esto quiere decir que la distancia de producción léxica entre ELE y CLM en dicha área semántica de las chicas supone solamente 0.08 palabra más que la distancia de los chicos. Por lo tanto, los dos estratos del sexo tienen en *Iluminación y calefacción* casi la misma diferencia cuantitativa de productividad léxica entre las dos lenguas; o sea la productividad léxica en ELE de ambos sexos en esa área semántica dista de igual modo de su productividad en CLM. Llegamos a la conclusión de que, en *Iluminación y calefacción*, la variable sexo tiene una incidencia casi nula con relación a la diferencia de productividad léxica entre las dos lenguas.

Siguiendo esa lógica, comprobamos que *Partes de la casa (sin muebles)* (-0.5), *La escuela: sus muebles y materiales* (-0.18) y *El campo* (0.33) son campos asociativos en los que el valor absoluto de la afinidad resulta igual o menor de 0.5, aparte de *Iluminación y calefacción* (0.08). En consecuencia, deducimos que, en esos cuatro campos asociativos, la variable sexo tiene una incidencia poco considerable en la diferencia de productividad léxica entre las

dos lenguas. Por otra parte, *Ropa y complementos* (0.81), *Los muebles de la casa* (0.98), *La cocina y sus utensilios* (-0.75) y *Profesiones y oficios* (-0.91) son áreas nocionales en las que el valor absoluto de la afinidad está entre 0.5 y 1. Por eso, en esas cuatro áreas nocionales, el sexo tiene una incidencia, pero muy pequeña, en la diferencia de productividad léxica entre las dos lenguas.

En el resto de centros de interés, donde dicho valor absoluto de afinidad supera 1, el sexo parece ejercer influencia al respecto, pero todavía pequeña. Podemos ver que, en *Objetos colocados en la mesa para la comida* (1.53) y *Medios de transporte* (1.09), las mujeres señalan más deficiencia que los varones, puesto que muestran, respectivamente, 1.53 y 1.09 palabra más que los varones en cuanto a la diferencia de promedio de palabras entre las dos lenguas. Por el contrario, en *Partes del cuerpo humano* (-1.26), *Alimentos y bebidas* (-1.68), *La ciudad* (-3.66), *Trabajos del campo y del jardín* (-1.52), *Los animales* (-1.1) y *Juegos y distracciones* (-1.16), los varones señalan más deficiencia que las mujeres, ya que, en estos centros de interés, tienen más palabras que las mujeres en cuanto a la diferencia de producción léxica entre ELE y CLM. Cabe mencionar que *La ciudad* es la única área nocional que tiene un valor absoluto que supera 2 y que llega a 3.66. Es donde se halla la mayor disparidad: en *La ciudad*, la productividad léxica en ELE dista de la productividad en CLM en los varones en un grado mayor que la que dista en las mujeres.

5.2.2.4. A continuación, vamos a examinar, por sexo y según la lengua de la encuesta, el total de vocablos que se recaban en cada uno de los centros de interés. Tal y como hemos comentado en el apartado 5.2.1.4, recordamos que, ante sub-muestras que no son preestablecidas y que, por tanto, implican entre sí unas diferencias considerables que se dan en el número de informantes, el total de vocablos no resulta un índice del todo fiable. En este factor social, el sexo, los varones solo ocupan más o menos un 20% de la muestra de ELE y de CLM, frente a la inmensa mayoría —alrededor de un 80%— que suponen las mujeres (véase el apartado 4.4.2). Podemos ver cómo

la diferencia enorme en el número de informantes repercute en los resultados de los vocablos según el sexo.

Centro de interés	Total de vocablos			
	Varón		Mujer	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	129	107	254	221
02 Ropa y complementos	158	139	251	321
03 Partes de la casa (sin muebles)	109	92	209	215
04 Los muebles de la casa	129	90	175	257
05 Alimentos y bebidas	213	244	355	623
06 Objetos en la mesa para la comida	84	61	138	181
07 La cocina y sus utensilios	121	118	217	225
08 La escuela: muebles y materiales	187	135	287	300
09 Iluminación y calefacción	117	95	213	233
10 La ciudad	295	274	632	710
11 El campo	245	216	497	659
12 Medios de transporte	169	111	265	289
13 Trabajos del campo y del jardín	212	155	401	371
14 Los animales	163	136	239	295
15 Juegos y distracciones	178	167	351	357
16 Profesiones y oficios	196	167	357	391
Global	2705	2307	4841	5648

Tabla 5.32. Total de vocablos por centro de interés en ELE y en CLM según la variable sexo

En la Tabla 5.32 se exhibe el número total de vocablos de cada uno de los dieciséis centros de interés según sea el sexo y según pertenezca al corpus de ELE o al de CLM. A partir de ahí, podemos darnos cuenta de la variedad léxica de los chicos y las chicas en cada una de las dos lenguas contempladas. En primer lugar, los resultados globales que se encuentran en la última fila de la tabla ya evidencian cómo el mayor número de informantes en el estrato de las mujeres se relaciona con una mayor riqueza léxica de las mismas, tanto en

ELE (2705, en los varones, frente a 4841, en las mujeres) como en CLM (2307, en los varones, frente a 5648, en las mujeres). La ventaja de las chicas sobre los chicos en ambas lenguas es completamente comprensible en el sentido de que, cuanto más informantes haya, más probablemente encontraremos a informantes atípicos que contribuyan al aumento de la diversificación del vocabulario.

Por otra parte, si comparamos, en cada sexo, la riqueza léxica global en ELE y en CLM, podemos ver que, en los varones, la riqueza léxica en ELE (2705) es un poco mayor que la que hay en CLM (2307); pero, en las mujeres, sucede lo contrario, es decir, la riqueza léxica en ELE (4841) es menor que la que hay en CLM (5648). Así pues, tomando como referencia la riqueza léxica en CLM, suponemos que los chicos tienen mayor riqueza léxica global en ELE que las chicas.

Si pasamos a mirar los resultados de las distintas áreas nocionales, vemos que no contradicen las observaciones desde la perspectiva de los datos globales. Por un lado, en todos los centros de interés, las mujeres aventajan a los chicos en riqueza léxica tanto en ELE como en CLM, lo cual se debe principalmente a su mayor número de integrantes. Por otro lado, al comparar las dos lenguas, vemos que los varones disponen de una mayor riqueza léxica en ELE que en CLM, en todos los centros de interés con excepción de *Alimentos y bebidas*. En cambio, las mujeres exponen una menor riqueza léxica en ELE que en CLM, en todos los centros de interés con excepción de *Partes del cuerpo humano y Trabajos del campo y del jardín*. Nos damos cuenta de que, tomando como referencia la riqueza léxica en CLM, los chicos también obtienen, por centro de interés, mejores resultados de riqueza léxica en ELE que las chicas.

Centro de interés	Rango según el total de vocablos			
	Varón		Mujer	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	11	12	9	14
02 Ropa y complementos	10	7	10	7
03 Partes de la casa (sin muebles)	15	14	14	15
04 Los muebles de la casa	11	15	15	11
05 Alimentos y bebidas	3	2	5	3
06 Objetos en la mesa para la comida	16	16	16	16
07 La cocina y sus utensilios	13	10	12	13
08 La escuela: muebles y materiales	6	9	7	8
09 Iluminación y calefacción	14	13	13	12
10 La ciudad	1	1	1	1
11 El campo	2	3	2	2
12 Medios de transporte	8	11	8	10
13 Trabajos del campo y del jardín	4	6	3	5
14 Los animales	9	8	11	9
15 Juegos y distracciones	7	4	6	6
16 Profesiones y oficios	5	4	4	4

Tabla 5.33. Rango de los centros de interés por sexo según el total de vocablos en ELE y en CLM

La Tabla 5.33 nos informa del rango de los centros de interés por sexo según sea su riqueza léxica en las dos lenguas respectivas. Un vistazo a las cifras en la tabla nos informa de que, unánimemente en ambos sexos y en ambas lenguas, *La ciudad* es el área semántica más rica en vocabulario; del mismo modo, *Objetos colocados en la mesa para la comida* es el campo nocional menos rico en vocabulario. Además, *El campo* es el segundo centro de interés más rico en ELE tanto en los varones como en las mujeres, aunque en los varones convive con *Alimentos y bebidas*.

Aparte de dichas áreas semánticas, la Tabla 5.33 indica que los chicos muestran una mayor riqueza léxica en ELE también en *Trabajos del campo* y

del jardín y Profesiones y oficios; y las chicas, en estos mismos centros de interés, pero con posiciones diferentes en el rango. Así vemos que los chicos y las chicas coinciden en los cinco centros más ricos en vocabulario en ELE: *La ciudad, El campo, Alimentos y bebidas, Trabajos del campo y del jardín y Profesiones y oficios*.

Por el contrario, además de *Objetos colocados en la mesa para la comida*, los chicos también tienen una menor variedad léxica en ELE en *Partes de la casa (sin muebles), Iluminación y calefacción, La cocina y sus utensilios, Partes del cuerpo humano y Los muebles de la casa*; mientras las chicas, en estos mismos centros de interés menos *Partes del cuerpo humano*. De modo que los informantes de ambos sexos coinciden en cinco centros menos ricos en vocabulario en ELE: *Objetos colocados en la mesa para la comida, Partes de la casa (sin muebles), Iluminación y calefacción, La cocina y sus utensilios y Los muebles de la casa*. Dada las coincidencias observadas, deducimos que la variable sexo tiene poca influencia con respecto a cuáles son las áreas nocionales más ricas o menos ricas en vocabulario en ELE.

Centro de interés	Índice de cohesión			
	Varón		Mujer	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	0.10	0.20	0.05	0.09
02 Ropa y complementos	0.05	0.12	0.03	0.05
03 Partes de la casa (sin muebles)	0.07	0.14	0.04	0.06
04 Los muebles de la casa	0.07	0.16	0.05	0.06
05 Alimentos y bebidas	0.07	0.09	0.04	0.03
06 Objetos en la mesa para la comida	0.08	0.17	0.04	0.06
07 La cocina y sus utensilios	0.05	0.12	0.03	0.06
08 La escuela: muebles y materiales	0.06	0.12	0.04	0.06
09 Iluminación y calefacción	0.05	0.13	0.03	0.05
10 La ciudad	0.04	0.07	0.02	0.02
11 El campo	0.04	0.07	0.02	0.03
12 Medios de transporte	0.06	0.13	0.03	0.05
13 Trabajos del campo y del jardín	0.04	0.09	0.02	0.03
14 Los animales	0.07	0.17	0.04	0.07
15 Juegos y distracciones	0.05	0.10	0.03	0.04
16 Profesiones y oficios	0.06	0.11	0.03	0.05

Tabla 5.34. Índice de cohesión por centro de interés en ELE y en CLM según la variable sexo

Para finalizar los análisis según la variable sexo, presentamos el índice de cohesión, que relaciona los dos índices que acabamos de observar —el promedio de palabras por informante y el total de vocablos—. La Tabla 5.34 expone el índice de cohesión de cada centro de interés por sexo y según cuál sea la lengua de la encuesta. Volvemos a ver cómo la diferencia en el número de informantes afecta los resultados de dicho índice en cada sub-muestra.

Por un lado, según cuál sea el sexo de los alumnos, en ambos corpus de ELE y de CLM, se observan unos índices de cohesión mayores en los varones que en las mujeres. Esto se debe a que, en el conjunto de cada corpus por separado, hay menos chicos que chicas (véase el apartado 4.4.2) y es más

probable que los chicos coincidan en sus respuestas, lo que conduce al aumento del valor del índice de cohesión. Por otro lado, según cuál sea la lengua de la encuesta, podemos ver que, tanto en la sub-muestra de los varones como en la de las mujeres, el índice de cohesión de cualquier área nacional en ELE es menor que el de la misma área en CLM. Creemos que, de igual modo, es por causa de que en las sub-muestras en CLM se halla un menor número de informantes. De todas maneras, sí comprobamos que los campos léxicos más compactos en una sub-muestra (con valores mayores del índice) también son los más compactos en la otra, y viceversa.

5.2.3. Incidencias cuantitativas de la variable “procedencia”

La variable procedencia contempla el lugar de origen de los estudiantes según de cuál de las cuatro zonas socioeconómicas del país procedan, tal y como hemos comentado en el apartado 4.4.3. Bajo ella, tenemos cuatro variantes correspondientes a dichas cuatro zonas: la zona del este, la del nordeste, la del centro y la del oeste. Con la procedencia, pretendemos averiguar si la disponibilidad léxica de los alumnos tienen que ver con el grado de desarrollo de su lugar de procedencia. Para ello, recordamos que el orden de las zonas de la más desarrollada económicamente a la menos desarrollada es el siguiente: este, nordeste, centro y oeste.

Por otra parte, debido a que las sub-muestras según la procedencia no están preestablecidas, hemos obtenido una sub-muestra del nordeste de tamaño muy reducido en los corpus tanto de ELE como de CLM. Concretamente, los informantes del nordeste solo suponen un 3% —ocho estudiantes— del total de encuestados en ELE y un 4% —siete estudiantes—, en CLM (véase el apartado 4.4.3). Sin embargo, cabe mencionar que los alumnos del nordeste constituyen, de por sí, un grupo minoritario en el universo estadístico de nuestra investigación. Por eso, no creemos que haga falta reunir a más informantes que correspondan a ese estrato.

5.2.3.1. Para empezar, en la Tabla 5.35, presentamos, por centro de

interés, el promedio de palabras por informante del corpus en ELE según las cuatro zonas de procedencia. Después de esta tabla, figura la Tabla 5.36, que nos permite ver el rango de los centros de interés en cada una de las cuatro zonas según el promedio de palabras por informante en ELE, de modo que se perciba más claramente el mayor o menor grado del valor de las cifras expuestas en la Tabla 5.35.

Centro de interés	Promedio de palabras en ELE			
	Este	Nordeste	Centro	Oeste
01 Partes del cuerpo humano	11.36	16.00	12.86	12.39
02 Ropa y complementos	7.93	11.50	9.05	9.09
03 Partes de la casa (sin muebles)	7.80	10.75	9.27	8.28
04 Los muebles de la casa	8.30	11.38	9.55	9.15
05 Alimentos y bebidas	14.88	18.25	14.55	15.78
06 Objetos en la mesa para la comida	6.00	7.00	6.23	5.82
07 La cocina y sus utensilios	6.25	7.38	7.09	6.72
08 La escuela: muebles y materiales	10.95	12.88	11.34	11.03
09 Iluminación y calefacción	5.33	7.00	6.32	5.22
10 La ciudad	12.79	20.00	14.49	13.76
11 El campo	9.14	12.63	10.91	10.68
12 Medios de transporte	8.90	11.38	9.43	8.54
13 Trabajos del campo y del jardín	7.77	12.63	7.57	6.96
14 Los animales	9.87	15.13	10.88	10.91
15 Juegos y distracciones	8.71	14.63	10.41	9.43
16 Profesiones y oficios	10.28	15.00	12.45	12.65
Global	9.14	12.72	10.15	9.78

Tabla 5.35. Promedio de palabras por informante por centro de interés en ELE según la variable procedencia

Centro de interés	Rango según el promedio de palabras por informante en ELE			
	Este	Nordeste	Centro	Oeste
01 Partes del cuerpo humano	3	3	3	4
02 Ropa y complementos	11	10	12	10
03 Partes de la casa (sin muebles)	12	13	11	12
04 Los muebles de la casa	10	11	9	9
05 Alimentos y bebidas	1	2	1	1
06 Objetos en la mesa para la comida	15	15	16	15
07 La cocina y sus utensilios	14	14	14	14
08 La escuela: muebles y materiales	4	7	5	5
09 Iluminación y calefacción	16	15	15	16
10 La ciudad	2	1	2	2
11 El campo	7	8	6	7
12 Medios de transporte	8	11	10	11
13 Trabajos del campo y del jardín	13	8	13	13
14 Los animales	6	4	7	6
15 Juegos y distracciones	9	6	8	8
16 Profesiones y oficios	5	5	4	3

Tabla 5.36. Rango de los centros de interés por procedencia según el promedio de palabras por informante en ELE

Si echamos un vistazo a la Tabla 5.35, podemos ver que la mayor producción léxica en ELE de todos los centros de interés y de las cuatro zonas de procedencia se halla en *La ciudad* del estrato del nordeste, con un promedio de 20 palabras por informante. Por el contrario, la menor producción léxica en ELE se halla en *Iluminación y calefacción* del estrato del oeste, con solamente una media de 5.22 palabras.

A continuación, vamos a analizar, por zona de procedencia, el promedio de palabras por informante en ELE. Como podemos comprobar en las Tablas 5.35 y 5.36, la mayor producción léxica de los informantes del este se da en *Alimentos y bebidas*, seguido por *La ciudad*, *Partes del cuerpo humano*, *La*

escuela: sus muebles y materiales y *Profesiones y oficios*. En cambio, en el estrato del nordeste, la mayor productividad léxica se registra en *La ciudad* y es seguido por *Alimentos y bebidas*, *Partes del cuerpo humano*, *Los animales* y *Profesiones y oficios*. Por otra parte, en los alumnos del centro, *Alimentos y bebidas* vuelve a ostentar la mayor producción léxica y es seguido por *La ciudad*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La escuela: sus muebles y materiales*. Por último, el mayor promedio de palabras de los estudiantes del oeste se encuentra en *Alimentos y bebidas*, seguido por *La ciudad*, *Profesiones y oficios*, *Partes del cuerpo humano* y *La escuela: sus muebles y materiales*.

A partir de ahí, observamos una coincidencia considerable por parte de las cuatro zonas, en cuanto a cuáles constituyen los centros más productivos en ELE. Los estratos del este, centro y oeste coinciden en el centro de interés más productivo en ELE, que resulta ser *Alimentos y bebidas*. En el estrato del nordeste, esta área semántica, aunque no es la más productiva, también ostenta una producción altamente aventajada, que la sitúa en la segunda posición. Así pues, llegamos a confirmar que *Alimentos y bebidas* es el área temática más productiva en ELE en las cuatro zonas de procedencia. Además, *La ciudad*, *Partes del cuerpo humano* y *Profesiones y oficios* son tres centros que se cuentan entre los campos semánticos más productivos en ELE en todas las zonas.

En lo referente a las diferencias, nos damos cuenta de que, aparte de los centros más productivos ya comentados, los alumnos del este, centro y oeste también disponen de mejor producción léxica en ELE en *La escuela: sus muebles y materiales*; mientras que los informantes del nordeste, como caso único, muestra mejor producción léxica en *Los animales*.

Ahora volvemos a las tablas de arriba para fijarnos en las áreas semánticas menos productivas en ELE. Como podemos ver, en los informantes del este, la menor producción léxica se detecta en *Iluminación y calefacción*, seguido por *Objetos colocados en la mesa para la comida*, *La cocina y sus*

utensilios, Trabajos del campo y del jardín y Partes de la casa (sin muebles). En cambio, la menor productividad léxica del estrato del nordeste se da en *Objetos colocados en la mesa para la comida e Iluminación y calefacción* (los dos con promedios de palabras iguales), seguidos por *La cocina y sus utensilios, Partes de la casa (sin muebles), Los muebles de la casa y Medios de transporte.* Por otra parte, los estudiantes del centro generan la menor cantidad de palabras en *Objetos colocados en la mesa para la comida*, que es seguido por *Iluminación y calefacción, La cocina y sus utensilios, Trabajos del campo y del jardín y Ropa y complementos.* Por último, los estudiantes del oeste tienen el menor promedio de palabras en *Iluminación y calefacción*, seguido por *Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Trabajos del campo y del jardín y Partes de la casa (sin muebles).*

Es notoria la coincidencia de las cuatro zonas de procedencia en sus áreas semánticas menos productivas en ELE. Por un lado, *Iluminación y calefacción* es el centro de interés menos productivo en ELE en los estratos del este, nordeste y oeste, y ocupa la penúltima posición en la zona del centro. Por otro lado, *Objetos colocados en la mesa para la comida* es el campo asociativo con la menor producción léxica en ELE en los estratos del nordeste y centro y está en la penúltima posición en el este y oeste. Así pues, los dos centros son los menos productivos en ELE en las cuatro zonas. Además, *La cocina y sus utensilios* es otro campo léxico que se cuenta entre los centros menos productivos en ELE de todas las zonas de procedencia.

Con respecto a las diferencias, o sea, los centros menos productivos no comunes en todas las zonas, indicamos que *Trabajos del campo y del jardín* se queda en los campos léxicos menos productivos en tres de los estratos — el este, centro y oeste—; *Partes de la casa (sin muebles)*, también en tres estratos: el este, nordeste y oeste. Como excepción, *Ropa y complementos* se cuenta entre los centros menos productivos en ELE solamente en el estrato del centro; mientras *Los muebles de la casa y Medios de transporte* solamente en los estudiantes del nordeste.

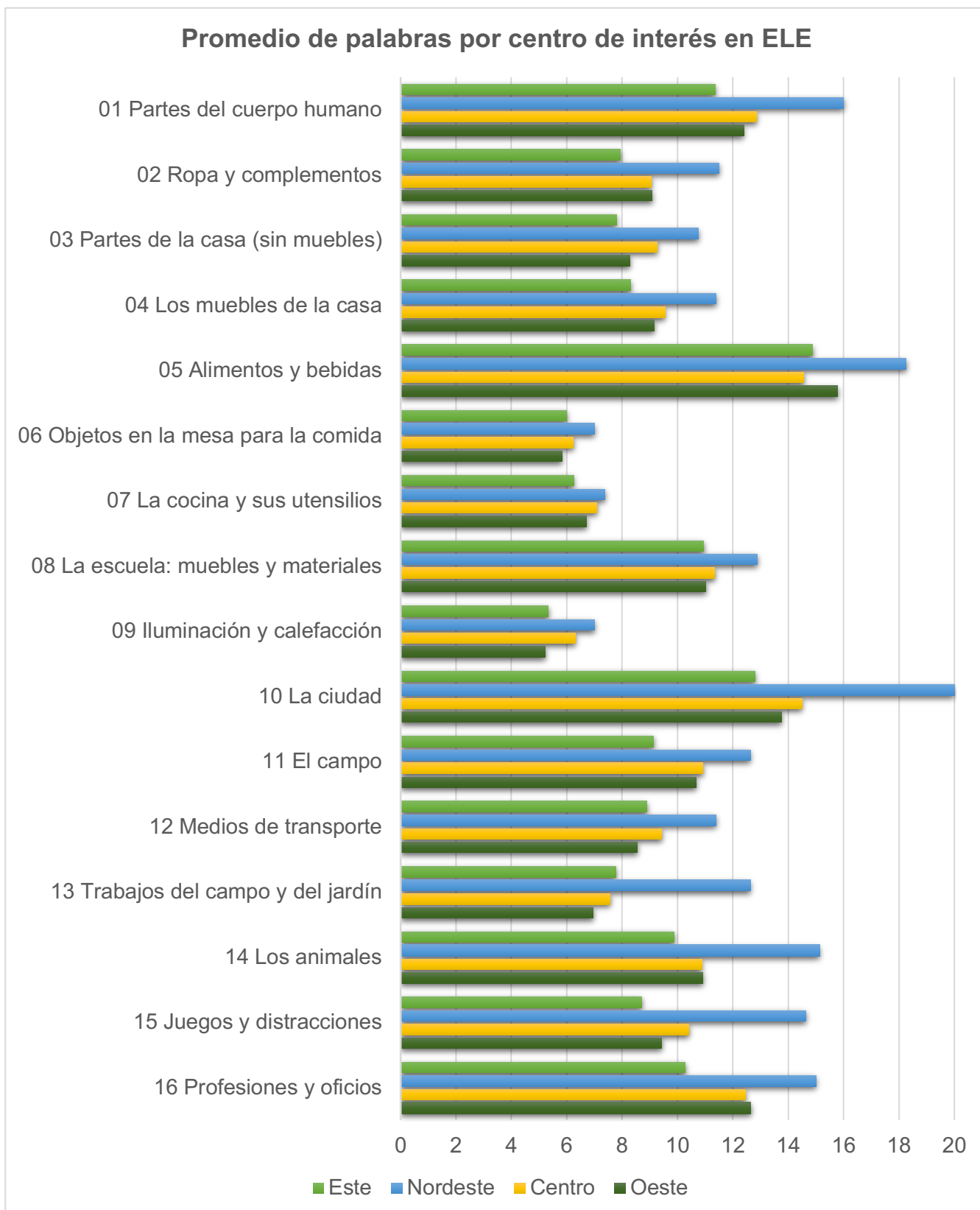


Gráfico 5.33. Promedio de palabras por informante por centro de interés en ELE según la variable procedencia

Ahora nos fijamos en el Gráfico 5.33, que manifiesta las diferencias entre

las cuatro zonas de procedencia con respecto al promedio de palabras por informante en ELE en cada uno de los centros de interés. Podemos ver que la productividad léxica en ELE de los informantes de la zona del nordeste ostenta una superioridad evidente en todos los centros de interés, en comparación con los datos de las demás zonas. Como hemos comentado al inicio de este apartado, los resultados del estrato del nordeste se obtienen de ocho informantes. Sabemos que los estudiantes del nordeste son minoritarios en el universo estadístico y, así, que los alumnos de esta zona que no participan en nuestra investigación también son pocos. Pero, puede que, por casualidad, los ocho integrantes que constituyen dicho estrato sean alumnos sobresalientes y los demás del nordeste —aunque pocos— que no se incluyen en la muestra sean alumnos mediocres. Es decir, no estamos seguros de que el promedio del estrato del nordeste sea un promedio adecuado. Así pues, para conseguir valorar adecuadamente las disparidades entre las distintas zonas con grados de desarrollo diferentes, creemos oportuno analizar solamente los datos de las otras tres zonas, aunque presentemos también los datos del nordeste en el Gráfico 5.33.

Así estando las cosas, nos centramos en los resultados en ELE arrojados por los informantes de las zonas del este, centro y oeste. Destacamos que, en nueve de las dieciséis áreas temáticas —*Partes del cuerpo humano, Partes de la casa (sin muebles), Los muebles de la casa, La cocina y sus utensilios, La escuela: sus muebles y materiales, La ciudad, El campo, Los animales y Juegos y distracciones*—, las diferencias cuantitativas en ELE entre las tres zonas evidencian la misma escala de mayor a menor producción léxica: del centro al oeste y del oeste al este. Es decir, en nueve centros de interés, el estrato del centro muestra la mayor productividad léxica en ELE, seguido por el estrato del oeste, y el del este tiene la menor productividad.

En otras cinco áreas nocionales —*Ropa y complementos, Objetos colocados en la mesa para la comida, Iluminación y calefacción, Medio de transporte y Profesiones y oficios*—, aunque no se manifiesta exactamente dicha escala, vemos que la desviación de la tendencia detectada —del centro

al oeste y del oeste al este— carece de importancia. Por ejemplo, en *Ropa y complementos*, solo se registra 0.04 palabras de diferencia entre las cifras del centro (9.05) y el oeste (9.09), mientras la cifra del este (7.93) es evidentemente más baja. Creemos que se trata de una leve variación de la tendencia referida, puesto que en *Ropa y complementos* los alumnos del centro y del oeste tienen una productividad léxica parecida y superior a la de los estudiantes del este. Al igual que en *Ropa y complementos*, en los demás centros mencionados podemos comprobar también una desviación poco importante o una leve variación de la escala, de mayor a menor producción léxica en ELE, del centro al oeste y del oeste al este.

Solamente en *Alimentos y bebidas y Trabajos del campo y del jardín* se observa una escala que no es la comentada. Suponemos que ello se debe a lo siguiente: además de que la producción léxica de los tres estratos varía según la zona socioeconómica, el vocabulario de esas dos áreas nocionales también viene fuertemente afectado por la zona geográfica. Es decir, generalmente la geografía puede determinar, por un lado, qué tipo de comidas la gente consume por hábito o por gusto y, por otro lado, qué tipo de vegetales es capaz de crecer en cada territorio y en qué modalidades la gente trabaja con lo que planta. A diferencia de *Alimentos y bebidas y Trabajos del campo y del jardín*, el vocabulario de los demás centros de interés se mantiene estable sea cual sea la zona geográfica. Por eso, en los demás campos léxicos, observamos más claramente la incidencia de la zona de procedencia en la producción léxica de los estudiantes, pero, en *Alimentos y bebidas y Trabajos del campo y del jardín* se registra mejor la incidencia de la zona de procedencia en el vocabulario usual de los mismos centros de interés.

Dado todo lo arriba indicado, deducimos que la producción léxica en ELE tiende a presentar una escala, de mayor a menor grado, del centro al oeste y del oeste al este. Al exceptuar los resultados de la zona del nordeste, los estratos del este, centro y oeste representan respectivamente zonas de grado de desarrollo alto, medio y bajo. Por lo tanto, vemos que la productividad léxica en ELE aventaja en la zona de grado de desarrollo medio, seguida por la zona

de grado de desarrollo bajo, mientras la menor productividad léxica en ELE se da en la zona de grado de desarrollo alto.

5.2.3.2. De aquí en adelante, presentamos, como referencia de comparación, los resultados sobre el promedio de palabras por informante del corpus en CLM según las cuatro zonas de procedencia. Del mismo modo como hemos hecho en los apartados previos sobre ELE, exponemos abajo la Tabla 5.37, en la que se encuentra dicho índice en CLM en cada centro de interés por zona de procedencia; mientras en la Tabla 5.38 se puede comprobar el rango de los centros también por zona de procedencia según el promedio de palabras por informante en CLM.

Centro de interés	Promedio de palabras en CLM			
	Este	Nordeste	Centro	Oeste
01 Partes del cuerpo humano	18.18	17.71	19.27	20.61
02 Ropa y complementos	16.29	17.14	17.85	18.08
03 Partes de la casa (sin muebles)	12.15	13.29	13.02	14.08
04 Los muebles de la casa	14.53	15.14	15.48	16.08
05 Alimentos y bebidas	20.21	21.71	20.33	21.43
06 Objetos en la mesa para la comida	10.91	11.86	11.60	11.69
07 La cocina y sus utensilios	13.95	13.43	14.42	14.67
08 La escuela: muebles y materiales	16.00	15.43	16.81	17.76
09 Iluminación y calefacción	11.45	11.57	12.92	12.71
10 La ciudad	17.50	16.29	16.50	18.88
11 El campo	15.06	13.57	16.83	18.51
12 Medios de transporte	14.00	14.86	14.71	14.47
13 Trabajos del campo y del jardín	12.21	12.71	11.54	12.92
14 Los animales	21.56	22.57	21.73	22.37
15 Juegos y distracciones	14.95	18.14	16.08	17.14
16 Profesiones y oficios	16.70	18.43	18.42	20.00
Global	15.35	15.87	16.09	16.96

Tabla 5.37. Promedio de palabras por informante por centro de interés en CLM según la variable procedencia

Centro de interés	Rango según el promedio de palabras por informante en CLM			
	Este	Nordeste	Centro	Oeste
01 Partes del cuerpo humano	3	5	3	3
02 Ropa y complementos	6	6	5	7
03 Partes de la casa (sin muebles)	14	13	13	13
04 Los muebles de la casa	10	9	10	10
05 Alimentos y bebidas	2	2	2	2
06 Objetos en la mesa para la comida	16	15	15	16
07 La cocina y sus utensilios	12	12	12	11
08 La escuela: muebles y materiales	7	8	7	8
09 Iluminación y calefacción	15	16	14	15
10 La ciudad	4	7	8	5
11 El campo	8	11	6	6
12 Medios de transporte	11	10	11	12
13 Trabajos del campo y del jardín	13	14	16	14
14 Los animales	1	1	1	1
15 Juegos y distracciones	9	4	9	9
16 Profesiones y oficios	5	3	4	4

Tabla 5.38. Rango de los centros de interés por procedencia según el promedio de palabras por informante en CLM

Los datos exhibidos en la Tabla 5.37 nos señalan que la mayor producción léxica en CLM de todos los centros de interés de las cuatro zonas de procedencia se halla en *Los animales* del estrato del nordeste —con un promedio de 22.57 palabras por informante—. Por el contrario, la menor producción léxica en CLM se encuentra en *Objetos colocados en la mesa para la comida* de la zona del este —con una media de 10.91 palabras—. Estos resultados en CLM coinciden parcialmente con los resultados previamente extraídos en ELE según la zona de procedencia: en ambas lenguas, es el estrato del nordeste el que tiene la mayor producción léxica; sin embargo, mientras el menor promedio en CLM se da en la zona del este, en ELE, se

halla en la zona del oeste.

Ahora pasamos al análisis sobre dicho promedio en CLM por zona de procedencia. Como podemos ver en la Tabla 5.38, las cuatro zonas coinciden totalmente en los primeros dos campos semánticos con una mayor producción léxica en CLM, los cuales son *Los animales* y *Alimentos y bebidas*. Aparte de estos dos centros de interés, la zona del este tiene un mayor promedio de palabras en CLM también, en orden de mayor a menor, en *Partes del cuerpo humano*, *La ciudad* y *Profesiones y oficios*; la zona del nordeste, en *Profesiones y oficios*, *Juegos y distracciones* y *Partes del cuerpo humano*; la zona del centro, en *Partes del cuerpo humano*, *Profesiones y oficios* y *Ropa y complementos*; y, por último, la zona del oeste, en *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad*.

Así pues, observamos que *Los animales*, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son cuatro áreas temáticas muy productivas en CLM, comunes en todas las zonas de procedencia. Recordemos que, comparando estos resultados con los previamente extraídos en ELE según la procedencia, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son los centros de interés comunes en ELE y en CLM en cuanto a tener un alto grado de productividad léxica en el conjunto de los dieciséis centros. Comprobamos otra vez —ahora desde la perspectiva de las cuatro zonas de procedencia— cómo las áreas semánticas más productivas en CLM resultan ser las más productivas también en ELE.

En el resto de áreas nocionales más productivas en CLM, se observan diferencias entre las zonas. El estrato del este y el del oeste cuentan con *La ciudad* entre sus áreas con una mayor producción léxica en CLM. Solamente la zona del nordeste incluye entre sus campos más productivos *Juegos y distracciones*, y únicamente la zona del centro, *Ropa y complementos*.

En lo referente a los centros de interés menos productivos en CLM, la Tabla 5.38 informa de lo siguiente. En los alumnos del este, la menor

producción léxica se detecta en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *Partes de la casa (sin muebles)*, *Trabajos del campo y del jardín* y *La cocina y sus utensilios*. En cambio, la menor productividad léxica del estrato del nordeste se halla en *Iluminación y calefacción*, seguido por *Objetos colocados en la mesa para la comida*, *Trabajos del campo y del jardín*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*. Por otra parte, los estudiantes del centro evocan la menor cantidad de palabras en *Trabajos del campo y del jardín*, seguido por *Objetos colocados en la mesa para la comida*, *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*. Por último, en la zona del oeste tienen el menor promedio de palabras en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *Trabajos del campo y del jardín*, *Partes de la casa (sin muebles)* y *Medios de transporte*.

Vemos que es destacada la coincidencia de las cuatro zonas en los campos léxicos menos productivos en CLM. *Objetos colocados en la mesa para la comida*, *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *Trabajos del campo y del jardín* son comunes en todas las zonas de procedencia como centros con un menor promedio de palabras por informante en CLM. Comparado estos resultados con los previamente extraídos en ELE según la procedencia, nos damos cuenta de que *Objetos colocados en la mesa para la comida* e *Iluminación y calefacción* son dos centros con una menor productividad léxica en todas las cuatro zonas socioeconómicas tanto en ELE como en CLM.

En lo que se refiere a las diferencias detectadas sobre los centros menos productivos en CLM, *La cocina y sus utensilios* se halla en los estratos del este, nordeste y centro, siendo otro campo léxico menos productivo. Como caso único, solo la zona del oeste tiene *Medios de transporte* entre sus áreas nocionales con una menor producción léxica en CLM.

Promedio de palabras por centro de interés en CLM

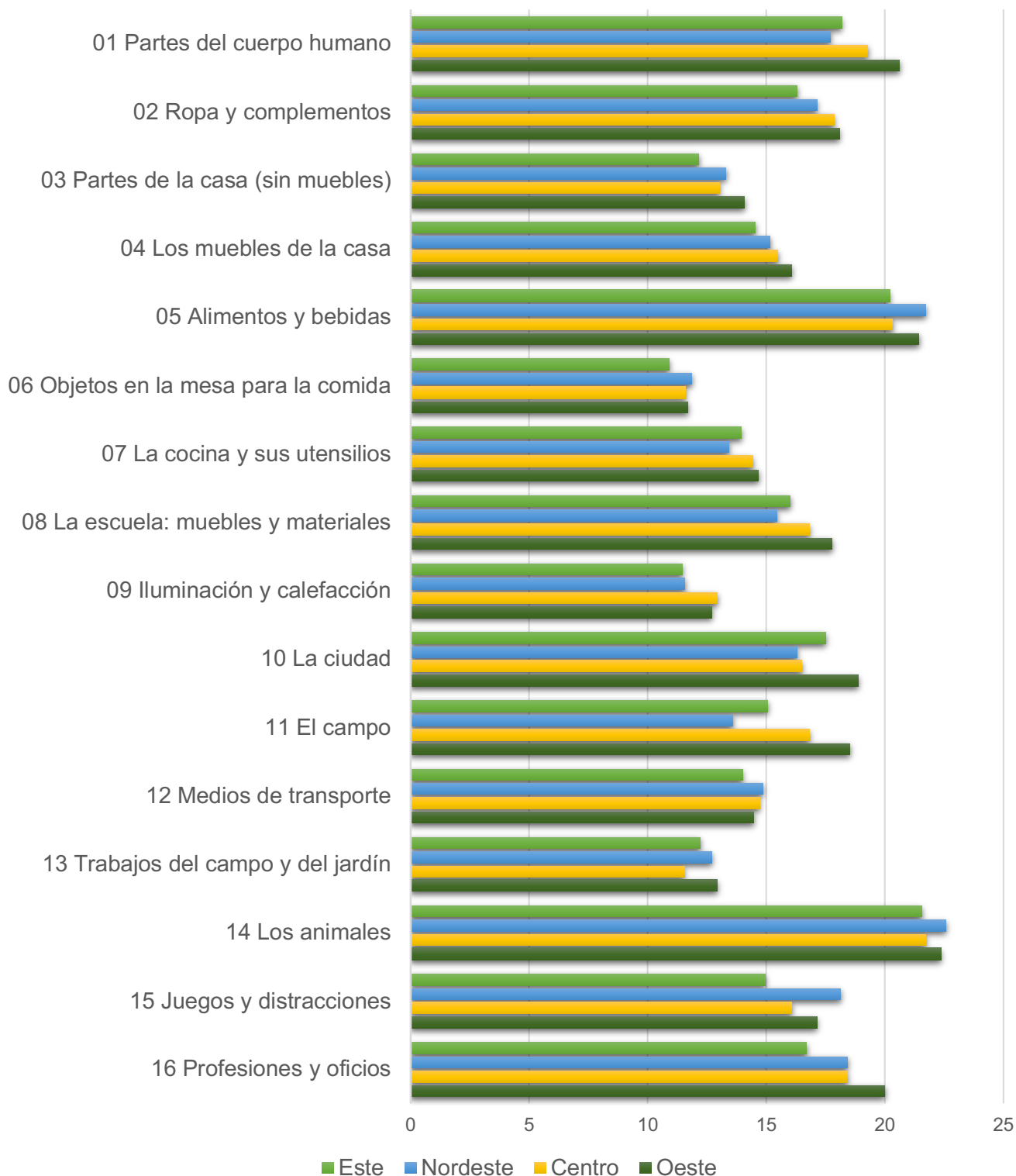


Gráfico 5.34. Promedio de palabras por informante por centro de interés en CLM según la variable procedencia

Ahora nos fijamos en el Gráfico 5.34, que expone las diferencias entre las

cuatro zonas con respecto a la producción léxica en CLM en cada uno de centros de interés. Echando un vistazo al gráfico, no se observa que la zona del nordeste ostente una superioridad evidente en CLM en todos los centros de interés, tal y como se comporta en ELE en el Gráfico 5.33. Pero, dadas las mismas consideraciones que hemos tenido en cuenta al analizar las diferencias de promedio en ELE entre las zonas de procedencia, optamos por pasar por alto los datos de la zona del nordeste en el Gráfico 5.34.

A continuación, nos centramos en los resultados en CLM arrojados por los informantes de las zonas del este, centro y oeste. Llamamos la atención sobre el hecho de que, en doce de las dieciséis áreas temáticas —*Partes del cuerpo humano, Ropa y complementos, Partes de la casa (sin muebles), Los muebles de la casa, Alimentos y comidas, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela: sus muebles y materiales, El campo, Los animales, Juegos y distracciones y Profesiones y oficios*—, las diferencias cuantitativas en CLM entre las tres zonas evidencian el mismo patrón de mayor a menor producción léxica: del oeste al centro y del centro al este. Es decir, en doce centros de interés, el estrato del oeste muestra la mayor productividad léxica en CLM, seguido por el estrato del centro, y el del este tiene la menor productividad.

En otras dos áreas nocionales —*Iluminación y calefacción y Medios de transporte*—, aunque no se manifiesta exactamente dicho patrón, vemos que la desviación de la tendencia detectada —del oeste al centro y del centro al este— carece de importancia. Por un lado, en *Iluminación y calefacción*, solo se registra 0.21 palabras de diferencia entre las cifras del oeste (12.71) y el centro (12.92), mientras que la cifra del este (11.45) es claramente más baja. Por otro lado, en *Medios de transporte*, solo se registra 0.24 palabras de diferencia entre las cifras del oeste (14.47) y el centro (14.71), mientras que la cifra del este (14) es más baja. Así pues, creemos que se trata de una leve variación de la tendencia referida, puesto que en *Iluminación y calefacción y Medios de transporte* los alumnos del oeste y del centro tienen una productividad léxica parecida y superior a la de los estudiantes del este.

Solamente en *La ciudad* y *Trabajos del campo y del jardín* se observa un patrón que se aleja del comentado. Hemos explicado en un apartado anterior la posible razón por la que *Trabajos del campo y del jardín* difiere de los demás centros de interés. Suponemos que aquí la geografía sigue siendo una variable que incide en ese campo léxico y, por tanto, en él obtenemos resultados distintos. Por otra parte, *La ciudad* es un área nocional que desobedece la tendencia detectada en el resto de centros de interés, principalmente porque el estrato del centro presenta una deficiencia de producción léxica en el tema tratado.

Dado todo lo arriba mencionado, deducimos que la producción léxica en CLM tiende a presentar un patrón descendente, de mayor a menor grado, del oeste al centro y del centro al este. Puesto que los estratos del este, centro y oeste representan, respectivamente, zonas de grado de desarrollo alto, medio y bajo, vemos que la productividad léxica en CLM de la zona de grado de desarrollo bajo aventaja la zona de grado de desarrollo medio, mientras la menor productividad léxica en CLM se da en la zona de grado de desarrollo alto. Esta deducción coincide parcialmente con la que hemos sacado en ELE: en las dos lenguas, los estudiantes de la zona de grado de desarrollo alto son los que tienen la menor productividad léxica. Pero, los estratos de las zonas de grado de desarrollo medio y bajo alternan sus posiciones en las dos lenguas: en ELE la zona de grado de desarrollo medio ostenta la mayor productividad léxica, mientras que en CLM la mayor productividad la tiene la zona de grado de desarrollo bajo.

A continuación, a partir de los datos globales expuestos en las Tablas 5.35 y 5.37, presentamos en los Gráficos 5.35 y 5.36 el promedio de palabras global de cada una de las cuatro zonas de procedencia en los corpus de ELE y de CLM.

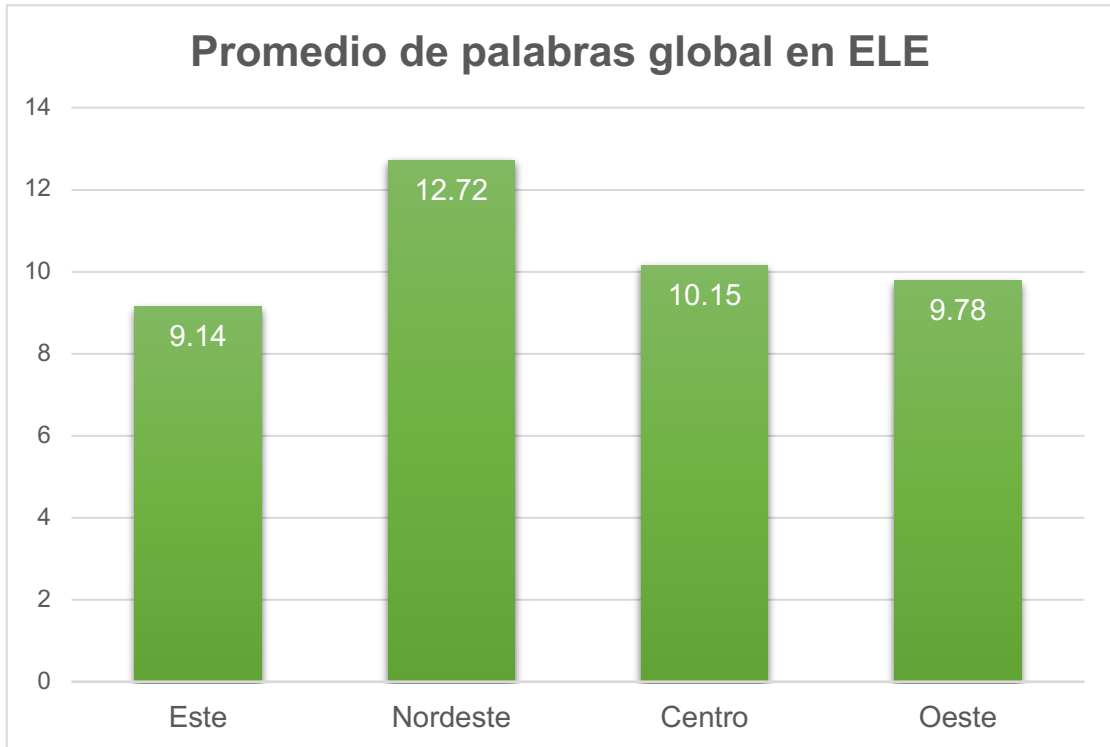


Gráfico 5.35. Promedio global de palabras por informante en ELE según la variable procedencia

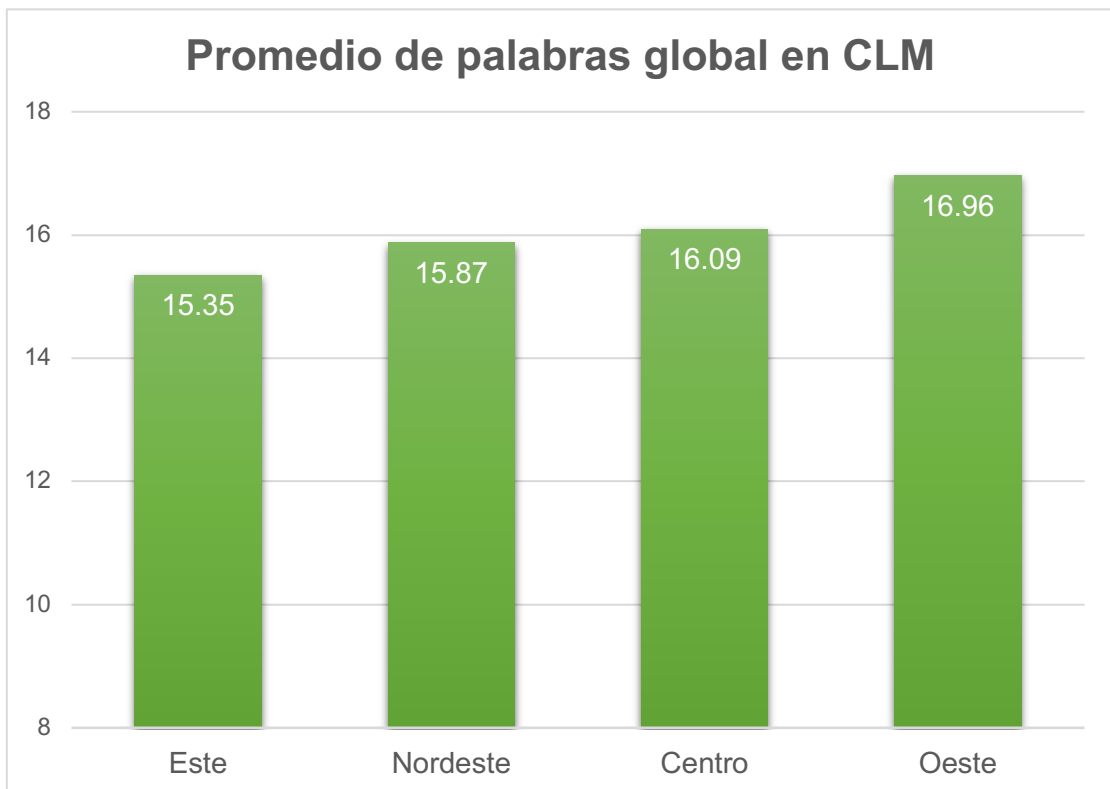


Gráfico 5.36. Promedio global de palabras por informante en CLM según la variable procedencia

Como muestran los dos gráficos de arriba, vemos comprobada las deducciones anteriormente extraídas por centro de interés. De manera global e independientemente de las áreas temáticas, al exceptuar los datos de la zona del nordeste, podemos ver 1) en el Gráfico 5.35, cómo la producción media de los informantes va descendiendo, en ELE, del centro al oeste y del oeste al este; y 2) en el Gráfico 5.36, cómo dicho índice va descendiendo, en CLM, del oeste al centro y del centro al este. Según el promedio global, la zona de grado de desarrollo medio dispone de mejor resultado en ELE, la zona de grado de desarrollo bajo tiene mejor resultado en CLM, y tanto en ELE como CLM la menor productividad léxica se halla en la zona de grado de desarrollo alto.

Por otra parte, comparando los Gráficos 5.35 y 5.36 entre ellos mismos, parece que los estudiantes procedentes de las cuatro zonas socioeconómicas no revelan una misma divergencia de producción léxica entre ELE y CLM. De modo que vamos a examinar en el siguiente apartado, por zona de procedencia, la distancia de producción léxica entre las dos lenguas.

5.2.3.3. La Tabla 5.39 reúne los datos de las Tablas 5.35 y 5.37, presentando, a su vez, los resultados completos sobre el promedio de palabras por informante por zona de procedencia y en las dos lenguas de nuestra investigación. A partir de ella, obtenemos la Tabla 5.40, que muestra, por centro de interés y por zona de procedencia, las diferencias cuantitativas entre el promedio en ELE y el promedio en CLM.

Centro de interés	Promedio de palabras							
	Este		Nordeste		Centro		Oeste	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	11.36	18.18	16.00	17.71	12.86	19.27	12.39	20.61
02 Ropa	7.93	16.29	11.50	17.14	9.05	17.85	9.09	18.08
03 Casa	7.80	12.15	10.75	13.29	9.27	13.02	8.28	14.08
04 Muebles	8.30	14.53	11.38	15.14	9.55	15.48	9.15	16.08
05 Alimentos	14.88	20.21	18.25	21.71	14.55	20.33	15.78	21.43
06 Objetos	6.00	10.91	7.00	11.86	6.23	11.60	5.82	11.69
07 Cocina	6.25	13.95	7.38	13.43	7.09	14.42	6.72	14.67
08 Escuela	10.95	16.00	12.88	15.43	11.34	16.81	11.03	17.76
09 Iluminación	5.33	11.45	7.00	11.57	6.32	12.92	5.22	12.71
10 Ciudad	12.79	17.50	20.00	16.29	14.49	16.50	13.76	18.88
11 Campo	9.14	15.06	12.63	13.57	10.91	16.83	10.68	18.51
12 Transporte	8.90	14.00	11.38	14.86	9.43	14.71	8.54	14.47
13 Trabajos	7.77	12.21	12.63	12.71	7.57	11.54	6.96	12.92
14 Animales	9.87	21.56	15.13	22.57	10.88	21.73	10.91	22.37
15 Juegos	8.71	14.95	14.63	18.14	10.41	16.08	9.43	17.14
16 Profesiones	10.28	16.70	15.00	18.43	12.45	18.42	12.65	20.00
Global	9.14	15.35	12.72	15.87	10.15	16.09	9.78	16.96

Tabla 5.39. Promedio de palabras por informante por centro de interés en ELE y en CLM según la variable procedencia

Centro de interés	Diferencia de promedio de palabras entre ELE y CLM			
	Este	Nordeste	Centro	Oeste
01 Partes del cuerpo humano	6.82	1.71	6.41	8.22
02 Ropa y complementos	8.36	5.64	8.80	8.99
03 Partes de la casa (sin muebles)	4.35	2.54	3.75	5.80
04 Los muebles de la casa	6.23	3.76	5.93	6.93
05 Alimentos y bebidas	5.33	3.46	5.78	5.65
06 Objetos en la mesa para la comida	4.91	4.86	5.37	5.87
07 La cocina y sus utensilios	7.70	6.05	7.33	7.95
08 La escuela: muebles y materiales	5.05	2.55	5.47	6.73
09 Iluminación y calefacción	6.12	4.57	6.60	7.49
10 La ciudad	4.71	-3.71	2.01	5.12
11 El campo	5.92	0.94	5.92	7.83
12 Medios de transporte	5.10	3.48	5.28	5.93
13 Trabajos del campo y del jardín	4.44	0.08	3.97	5.96
14 Los animales	11.69	7.44	10.85	11.46
15 Juegos y distracciones	6.24	3.51	5.67	7.71
16 Profesiones y oficios	6.42	3.43	5.97	7.35
Global	6.21	3.15	5.94	7.18

Tabla 5.40. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable procedencia

Como podemos ver en la Tabla 5.40, las cifras revelan las distancias de producción léxica entre las dos lenguas, con respecto al promedio de palabras en CLM. La producción léxica en ELE es inferior a la que hay en CLM en todos los centros de interés y en todas las zonas socioeconómicas, con excepción de *La ciudad* del estrato del nordeste (-3.71), donde el promedio en ELE supera dicho índice en CLM con 3.71 palabras más. Hemos visto en los apartados anteriores que los informantes del nordeste muestran resultados sobresalientes en ELE. Ahora observamos que esos informantes tienen una productividad léxica aún más relevante en ELE en *La ciudad*, puesto que es la única área nocional en la que un estrato presenta un promedio de palabras en

ELE por encima del promedio en CLM.

Además de ello, la tabla nos informa de que la mayor diferencia entre los promedios de ELE y de CLM se halla en *Los animales* en el estrato del este (11.69, en beneficio de CLM). Cabe destacar que la segunda y tercera mayor diferencia también se registra en este centro de interés en los estratos del oeste (11.46) y del centro (10.85), respectivamente.

Las celdas marcadas en fondo gris en la Tabla 5.40 indican aquellas áreas nocionales en las que la diferencia cuantitativa está por encima del promedio global o, en otras palabras, áreas en las que la producción léxica en ELE de los informantes dista enormemente de la que se da en CLM. Como hemos recordado en los apartados anteriores, sugerimos que se preste atención a estas áreas semánticas en la enseñanza de la lengua extranjera, ya que la inferioridad en ELE en ellas puede compensarse a través de adquirir los homólogos en español de las palabras que los estudiantes son capaces de evocar en CLM.

Hay cuatro centros de interés en los que los estudiantes de todas las zonas señalan gran divergencia de producción léxica entre las dos lenguas: *Ropa y complementos*, *La cocina y sus utensilios*, *Los animales* y *Profesiones y oficios*. Esto quiere decir que todos los estudiantes, independientemente de la procedencia, muestran considerable debilidad en estos cuatro centros en ELE, tomando como referencia su producción léxica en CLM. Aparte de estos cuatro centros de interés, hay varios campos léxicos que implican deficiencia en ELE con respecto a la producción en CLM en tres de los estratos. Señalamos que los estratos del este, centro y oeste tienen deficiencia en ELE en *Partes del cuerpo humano*. De igual modo, las zonas del nordeste, centro y oeste, en *Iluminación y calefacción*; y las zonas del este, nordeste y oeste, en *Juegos y distracciones*. Como casos únicos, solo los alumnos del nordeste presentan debilidad en *Alimentos y bebidas*, *Objetos colocados en la mesa para la comida* y *Medios de transporte*; mientras solo la zona del oeste, en *El campo*.

Ahora pasamos a fijarnos en la diferencia global de promedio de palabras entre ELE y CLM por zona de procedencia, la cual se expone en la última fila de la Tabla 5.40 y, de manera ilustrativa, en el Gráfico 5.37.

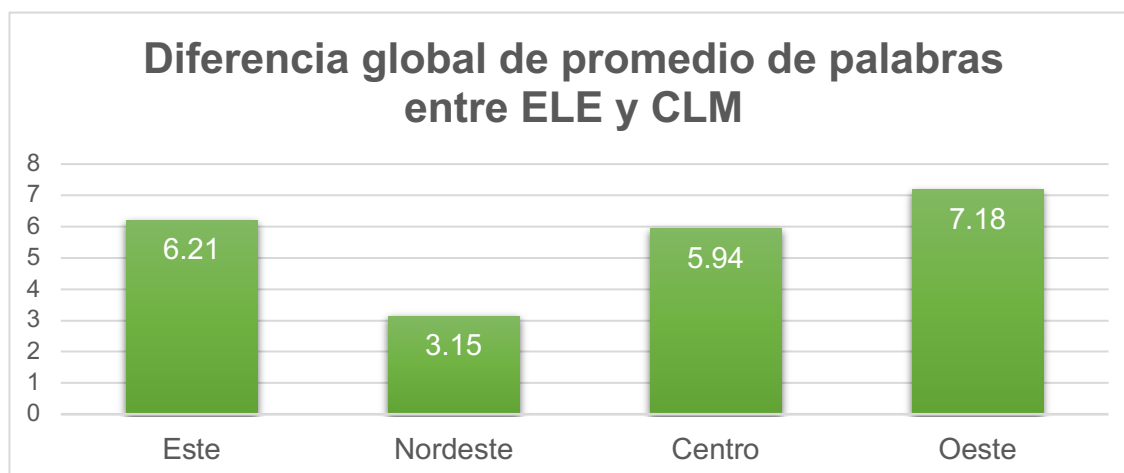


Gráfico 5.37. Diferencia global de promedio de palabras entre ELE y CLM según la variable procedencia

Al igual que en los análisis anteriores, hemos de pasar por alto el dato de la zona del nordeste por considerarlo inadecuado al contar con pocos informantes. Como podemos ver en el Gráfico 5.37, la diferencia que existe entre los promedios de palabras en ELE y en CLM parece dar ventaja al estrato de la zona del centro —sub-muestra, como ya sabemos, que obtiene la mayor productividad léxica en ELE—. En otras palabras, los informantes del centro solamente aportan, en promedio, 5.94 palabras menos en ELE que en CLM —la menor distancia entre las dos lenguas—. Por el contrario, el estrato del oeste muestra la mayor distancia de producción léxica entre las dos lenguas (7.18). Cabe destacar que, a pesar de que los alumnos del este señalan la menor productividad léxica tanto en ELE como en CLM, su divergencia de promedio entre ELE y CLM se ve menor que la del estrato del oeste. De ahí, encontramos un patrón lineal, de menor a mayor distancia entre las dos lenguas, del centro al este y del este al oeste. A continuación, vamos a ver si esta tendencia también se percibe por centro de interés. Para ello, observaremos el Gráfico 5.38, que resulta ser una visualización de los datos presentados en la Tabla 5.40.

Diferencia de promedio de palabras entre ELE y CLM

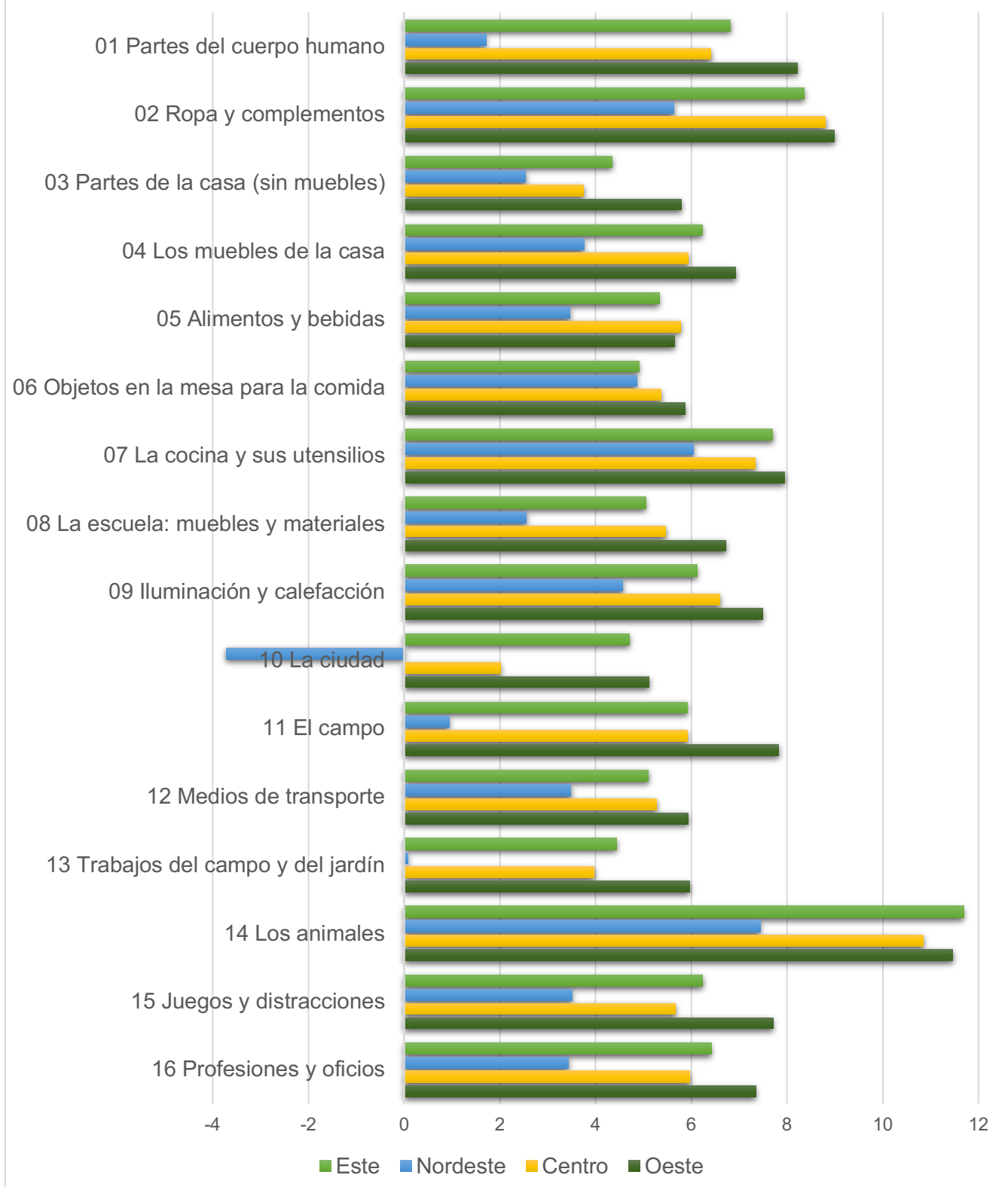


Gráfico 5.38. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable procedencia

Mediante el Gráfico 5.38, tratamos de comprobar la tendencia que las tres zonas de procedencia (sin contar con la zona del nordeste) manifiestan en los resultados globales, con respecto al patrón lineal —de menor a mayor distancia entre las dos lenguas— del centro al este y del este al oeste. Así pues, el gráfico nos informa de que, en nueve de los dieciséis centros de interés — *Partes del cuerpo humano, Partes de la casa (sin muebles), Los muebles de la casa, La cocina y sus utensilios, La ciudad, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios*—, el patrón detectado de menor a mayor divergencia de producción léxica entre ELE y CLM sí existe. De modo que comprobamos que, en la mayoría de las áreas temáticas, el estrato de la zona del centro revela la menor diferencia entre sus producciones de cada lengua, o sea, los informantes del centro son los estudiantes cuya productividad léxica en ELE se aproxima más a su productividad en CLM. A este respecto, el estrato del este se queda inmediatamente después del estrato del centro, mientras que los alumnos del oeste presentan la mayor distancia de la productividad léxica en ELE en comparación con su productividad en CLM.

Ahora pasamos a examinar qué sucede en el resto de campos léxicos donde no se evidencia la tendencia referida. Vemos que, en seis centros de interés —*Ropa y complementos, Alimentos y bebidas, Objetos colocados en la mesa para la comida, La escuela: sus muebles y materiales, Iluminación y calefacción y Medios de transporte*—, se mantiene la superioridad de la zona del este sobre la del oeste; es decir, tal como hemos deducido en el párrafo anterior, el estrato del este presenta una menor distancia entre las dos lenguas que el estrato del oeste. Por lo tanto, deducimos que lo que cambia la tendencia detectada —en los resultados globales y en los nueve centros de interés— es la deficiencia de producción léxica en ELE del estrato del centro o, por el contrario, la excelencia de su producción léxica en CLM en los seis campos semánticos que acabamos de comentar, puesto que en estos campos el estrato del centro pierde su ventaja al respecto. Por último, únicamente en *Los animales* observamos que el cambio de la tendencia detectada —en los resultados globales y en los nueve centros de interés— se debe a que el

estrato del este muestra un resultado distinto: se convierte en el estrato con mayor divergencia de producción léxica en dicho centro de interés.

Como conclusión de esta parte, aseguramos encontrar, en las tres zonas socioeconómicas, una escala general de menor a mayor distancia de producción léxica entre ELE y CLM: de la zona del centro, a la del este y, finalmente, a la del oeste. Como podemos ver, con respecto a la diferencia de productividad léxica entre las dos lenguas, la zona de grado de desarrollo medio tiene la mayor ventaja. Mientras tanto, la zona de grado de desarrollo alto aventaja a la zona de grado de desarrollo bajo.

5.2.3.4. A continuación, vamos a examinar, por zona de procedencia y según la lengua de la encuesta, el total de vocablos que se recaban en cada uno de los centros de interés.

Centro de interés	Total de vocablos							
	Este		Nordeste		Centro		Oeste	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	206	174	56	51	127	131	117	130
02 Ropa	205	228	54	63	158	192	129	198
03 Casa	163	134	38	43	120	138	105	131
04 Muebles	150	166	43	52	124	157	103	152
05 Alimentos	297	418	81	94	228	333	214	353
06 Objetos	109	123	29	36	96	99	71	104
07 Cocina	157	165	38	58	142	153	137	148
08 Escuela	235	201	61	63	192	194	168	212
09 Iluminación	152	157	36	45	146	140	95	137
10 Ciudad	450	488	113	81	374	362	325	393
11 Campo	355	405	69	79	308	359	244	372
12 Transporte	218	196	45	49	176	163	101	143
13 Trabajos	303	243	75	72	239	201	190	235
14 Animales	195	209	60	65	164	199	123	196
15 Juegos	273	254	68	77	234	210	167	223
16 Profesiones	278	255	66	79	220	237	205	266
Global	3746	3816	932	1007	3048	3268	2494	3393

Tabla 5.41. Total de vocablos por centro de interés en ELE y en CLM según la variable procedencia

En la Tabla 5.32 se expone el número total de vocablos de cada uno de los dieciséis centros de interés según sea la zona y la lengua. A partir de ahí, podemos darnos cuenta de la variedad léxica de los estudiantes de cada zona socioeconómica en cada una de las dos lenguas contempladas.

Recordemos que, ante sub-muestras que implican entre sí unas diferencias considerables motivadas por el número de informantes, el total de vocablos no resulta un índice del todo fiable. Como hemos llamado la atención a lo largo de los anteriores apartados, en el caso de las zonas de procedencia, los estudiantes del nordeste solo suponen un 3% —ocho estudiantes— del

total de encuestados en ELE y un 4% —siete estudiantes—, en CLM. Podemos ver cómo la diferencia enorme en el número de informantes repercute en los resultados de los vocablos según la procedencia. Es evidente que el menor número de informantes en el estrato del nordeste se relaciona con una menor riqueza léxica, tanto en ELE como en CLM y tanto en los resultados globales como en los resultados por centro de interés.

Los resultados globales que se encuentran en la última fila de la tabla nos informan de que el estrato del este cuenta con la mayor riqueza léxica tanto en ELE (3746) como en CLM (3816). Esto puede deberse igualmente a su superioridad en el número de informantes, ya que el estrato del este supone un 46% y un 38% del total de informantes en el corpus de ELE y de CLM, respectivamente (véase el apartado 4.4.3). El estrato del centro y del oeste, que sí tienen números de informantes parecidos, parecen disponer de una riqueza léxica poco distante en CLM (3268 y 3393, respectivamente); pero, la zona del centro aventaja a la del oeste en la variedad léxica en ELE (3048 frente a 2494).

Dada la gran desigualdad del número de informantes en las distintas submuestras de las cuatro zonas, no nos adentramos a analizar el número de vocablos por centro de interés. En cambio, pasamos a observar el rango de los centros según los vocablos, desde la perspectiva de las cuatro zonas socioeconómicas.

Centro de interés	Rango según el total de vocablos							
	Este		Nordeste		Centro		Oeste	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	9	11	9	12	13	15	11	15
02 Ropa	10	7	10	8	10	9	9	8
03 Casa	12	15	13	15	15	14	12	14
04 Muebles	15	12	12	11	14	11	13	10
05 Alimentos	4	2	2	1	5	3	3	3
06 Objetos	16	16	16	16	16	16	16	16
07 Cocina	13	13	13	10	12	12	8	11
08 Escuela	7	9	7	8	7	8	6	7
09 Iluminación	14	14	15	14	11	13	15	13
10 Ciudad	1	1	1	2	1	1	1	1
11 Campo	2	3	4	3	2	2	2	2
12 Transporte	8	10	11	13	8	10	14	12
13 Trabajos	3	6	3	6	3	6	5	5
14 Animales	11	8	8	7	9	7	10	9
15 Juegos	6	5	5	5	4	5	7	6
16 Profesiones	5	4	6	3	6	4	4	4

Tabla 5.42. Rango de los centros de interés por procedencia según el total de vocablos en ELE y en CLM

La Tabla 5.33 nos informa del rango de los centros de interés por zona de procedencia según cómo sea la riqueza léxica en las dos lenguas respectivas. De un vistazo podemos ver que *La ciudad* es el campo léxico más rico en vocabulario en ELE unánimemente en las cuatro zonas. *Objetos colocados en la mesa para la comida* es el área nocional menos rico en vocabulario en las dos lenguas en las cuatro zonas.

Aparte de dichas áreas semánticas, vemos que el estrato del este muestra una mayor riqueza léxica en ELE también en *El campo*, *Trabajos del campo y del jardín*, *Alimentos y bebidas* y *Profesiones y oficios*; el estrato del nordeste, en *Alimentos y bebidas*, *Trabajos del campo y del jardín*, *El campo* y *Juegos y*

distracciones; la zona del centro, en *El campo*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Alimentos y bebidas*; y la zona del oeste, en *El campo*, *Alimentos y bebidas*, *Profesiones y oficios* y *Trabajos del campo y del jardín*. De ahí, vemos que los informantes de las cuatro zonas coinciden en cuatro centros más ricos en vocabulario en ELE: *La ciudad*, *El campo*, *Trabajos del campo y del jardín* y *Alimentos y bebidas*. Las zonas del este y oeste también tienen un alto grado de variedad léxica en ELE en *Profesiones y oficios*; mientras las zonas del nordeste y centro lo tienen en *Juegos y distracciones*.

Por otra parte, además de *Objetos colocados en la mesa para la comida*, el estrato del este también tiene una menor variedad léxica en ELE en *Los muebles de la casa*, *Iluminación y calefacción*, *La cocina y sus utensilios* y *Las partes de la casa (sin muebles)*; la zona del nordeste, en *Iluminación y calefacción*, *La cocina y sus utensilios*, *Partes de la casa (sin muebles)* y *Los muebles de la casa*; la zona del centro, en *Las partes de la casa (sin muebles)*, *Los muebles de la casa*, *Partes del cuerpo humano* y *La cocina y sus utensilios*; y el estrato del oeste, en *Iluminación y calefacción*, *Medios de transporte*, *Los muebles de la casa* y *Partes de la casa (sin muebles)*. Así es que los informantes de las cuatro zonas coinciden en tres centros menos ricos en vocabulario en ELE: *Objetos colocados en la mesa para la comida*, *Los muebles de la casa* y *Partes de la casa (sin muebles)*. En cuanto a las diferencias, *La cocina y sus utensilios* implica una menor riqueza léxica en ELE en tres estratos: este, nordeste y centro; *Iluminación y calefacción*, también en tres estratos: este, nordeste y oeste. Como casos únicos, solo la zona del centro señala una menor riqueza léxica en ELE en *Partes del cuerpo humano*, y solo el estrato del oeste, en *Medios de transporte*.

Centro de interés	Índice de cohesión							
	Este		Nordeste		Centro		Oeste	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	0.06	0.10	0.29	0.35	0.10	0.15	0.11	0.16
02 Ropa	0.04	0.07	0.21	0.27	0.06	0.09	0.07	0.09
03 Casa	0.05	0.09	0.28	0.31	0.08	0.09	0.08	0.11
04 Muebles	0.06	0.09	0.26	0.29	0.08	0.10	0.09	0.11
05 Alimentos	0.05	0.05	0.23	0.23	0.06	0.06	0.07	0.06
06 Objetos	0.06	0.09	0.24	0.33	0.06	0.12	0.08	0.11
07 Cocina	0.04	0.08	0.19	0.23	0.05	0.09	0.05	0.10
08 Escuela	0.05	0.08	0.21	0.24	0.06	0.09	0.07	0.08
09 Iluminación	0.04	0.07	0.19	0.26	0.04	0.09	0.05	0.09
10 Ciudad	0.03	0.04	0.18	0.20	0.04	0.05	0.04	0.05
11 Campo	0.03	0.04	0.18	0.17	0.04	0.05	0.04	0.05
12 Transporte	0.04	0.07	0.25	0.30	0.05	0.09	0.08	0.10
13 Trabajos	0.03	0.05	0.17	0.18	0.03	0.06	0.04	0.05
14 Animales	0.05	0.10	0.25	0.35	0.07	0.11	0.09	0.11
15 Juegos	0.03	0.06	0.22	0.24	0.04	0.08	0.06	0.08
16 Profesiones	0.04	0.07	0.23	0.23	0.06	0.08	0.06	0.08

Tabla 5.43. Índice de cohesión por centro de interés en ELE y en CLM según la variable procedencia

Por último, presentamos el índice de cohesión, que relaciona los dos índices anteriormente estudiados —el promedio de palabras por informante y el total de vocablos—. La Tabla 5.43 exhibe el índice de cohesión de cada centro de interés por zona de procedencia y según cuál sea la lengua de la encuesta. Si nos detenemos a observar las divergencias de dicho índice según la zona, podemos ver cómo la cantidad de encuestados influye en los resultados obtenidos. Como hemos advertido en otros apartados de la tesis, cuantos menos informantes haya en una sub-muestra, mayor índice de cohesión puede darse en la misma.

Dado lo anterior, entre las cuatro zonas socioeconómicas, observamos

que el estrato del nordeste tiene el mayor índice de cohesión en ambas lenguas y en todos los centros de interés; mientras que el estrato del este, en general, presenta el menor índice. Por otro lado, entre ELE y CLM, podemos ver que, en las sub-muestras de las dos lenguas bajo cualquier zona, el índice de cohesión de cualquier área nocional en ELE es igual o menor que el de la misma área en CLM (solo con excepción de *Alimentos y bebidas* de la zona del oeste). De igual modo, tal cosa es debida a que en las sub-muestras en CLM se halla un menor número de informantes. De todos modos, volvemos a comprobar que los campos léxicos más compactos en una sub-muestra (con valores mayores del índice) también son los más compactos en la otra, y viceversa.

5.2.4. Incidencias cuantitativas de la variable “nivel de inglés”

El nivel de inglés es otra variable que vamos a analizar en esta sección de la tesis. Como hemos comentado en el apartado 4.4.4, bajo esta variable, vamos a examinar los resultados obtenidos en informantes con cuatro niveles de inglés distintos: el nivel básico, el intermedio, el avanzado y el superior. El cotejo entre dichos resultados será interesante, porque se supone que la competencia lingüística de los estudiantes de español tiene mucho que ver con su conocimiento de otras lenguas extranjeras. En nuestro caso, el inglés es la lengua extranjera que todos los alumnos aprenden además del español. Por eso, en el presente apartado de la tesis, pretendemos averiguar en qué consiste la relación de la disponibilidad léxica en español con el nivel de inglés.

Destacamos que, al igual que sucede con el estrato de la zona del nordeste en la variable procedencia, hemos tenido la sub-muestra del nivel de inglés superior con pocos informantes. En concreto, los integrantes de dicha sub-muestra solo suponen un 3% —ocho estudiantes— del total de encuestados en ELE y un 1% —dos estudiantes—, en CLM (véase el apartado 4.4.4). Dado que estos porcentajes reflejan la composición real de la población estadística (el universo), no vemos necesario encuestar a más informantes con el nivel superior.

5.2.4.1. Para empezar, en la Tabla 5.44, presentamos, por centro de interés, el promedio de palabras por informante del corpus en ELE según los cuatro niveles de inglés —el básico, intermedio, avanzado y superior—. Después de esta tabla, está la Tabla 5.45 que nos permite ver el rango de los centros de interés en cada uno de los niveles de inglés según el promedio de palabras por informante en ELE, de modo que se perciba más claramente el mayor o menor grado del valor de las cifras expuestas en la Tabla 5.44.

Centro de interés	Promedio de palabras en ELE			
	Básico	Intermedio	Avanzado	Superior
01 Partes del cuerpo humano	9.40	11.12	15.90	16.38
02 Ropa y complementos	6.94	8.22	10.57	12.38
03 Partes de la casa (sin muebles)	7.07	8.47	9.65	9.63
04 Los muebles de la casa	7.76	9.15	9.95	9.25
05 Alimentos y bebidas	12.59	15.20	17.41	20.00
06 Objetos en la mesa para la comida	4.93	5.36	7.83	8.00
07 La cocina y sus utensilios	5.82	6.14	7.77	9.38
08 La escuela: muebles y materiales	9.64	11.20	12.49	13.38
09 Iluminación y calefacción	3.62	5.31	8.01	6.75
10 La ciudad	11.01	13.16	16.78	18.13
11 El campo	6.71	10.26	13.53	11.38
12 Medios de transporte	6.93	8.82	11.32	11.50
13 Trabajos del campo y del jardín	5.61	7.68	9.67	9.75
14 Los animales	6.95	9.51	15.05	17.13
15 Juegos y distracciones	6.99	9.49	12.07	12.00
16 Profesiones y oficios	9.30	11.16	14.55	11.75
Global	7.58	9.39	12.03	12.30

Tabla 5.44. Promedio de palabras por informante por centro de interés en ELE según la variable nivel de inglés

Centro de interés	Rango según el promedio de palabras por informante en ELE
-------------------	---

	Básico	Intermedio	Avanzado	Superior
01 Partes del cuerpo humano	4	5	3	4
02 Ropa y complementos	10	12	10	6
03 Partes de la casa (sin muebles)	7	11	13	12
04 Los muebles de la casa	6	9	11	14
05 Alimentos y bebidas	1	1	1	1
06 Objetos en la mesa para la comida	15	15	15	15
07 La cocina y sus utensilios	13	14	16	13
08 La escuela: muebles y materiales	3	3	7	5
09 Iluminación y calefacción	16	16	14	16
10 La ciudad	2	2	2	2
11 El campo	12	6	6	10
12 Medios de transporte	11	10	9	9
13 Trabajos del campo y del jardín	14	13	12	11
14 Los animales	9	7	4	3
15 Juegos y distracciones	8	8	8	7
16 Profesiones y oficios	5	4	5	8

Tabla 5.45. Rango de los centros de interés por nivel de inglés según el promedio de palabras por informante en ELE

Echando una ojeada a la Tabla 5.44, podemos ver que la mayor producción léxica en ELE de todos los centros de interés y de los cuatro niveles de inglés se detecta en *Alimentos y bebidas* del estrato del nivel superior — con un promedio de 20 palabras—. Por el contrario, la menor producción léxica en ELE se halla en *Iluminación y calefacción* del estrato del nivel básico —con solamente una media de 3.62 palabras—.

A continuación, analizaremos, por nivel de inglés, el promedio de palabras por informante en ELE. Como podemos ver en la Tabla 5.45, *Alimentos y bebidas* es el área temática más productiva en ELE en todos los estratos según el nivel de inglés. Después de ella, *La ciudad* es la segunda área temática más productiva en ELE también en todos los estratos según el nivel de inglés.

Aparte de estos dos centros de interés, la Tabla 5.45 nos ilustra sobre la mayor producción léxica en ELE de los informantes del nivel básico de inglés en *La escuela: sus muebles y materiales*, *Partes del cuerpo humano* y *Profesiones y oficios*. Cabe destacar que, según la Tabla 5.44, solo existen dos centros de interés en los que los estudiantes del nivel básico tienen un promedio de palabras superior a 10. En cambio, en el estrato del nivel intermedio de inglés, los centros que alcanzan más de 10 palabras de promedio son seis. Una mayor productividad léxica se registra en este estrato también en *La escuela: sus muebles y materiales*, *Profesiones y oficios* y *Partes del cuerpo humano*. Por otra parte, los alumnos de nivel avanzado de inglés tienen la mayoría de los centros con un promedio superior a 10 palabras. *Partes del cuerpo humano*, *Los animales* y *Profesiones y oficios* son otros campos temáticos donde estos alumnos tienen una mayor productividad léxica. Por último, la mayoría de los centros en los informantes del nivel superior de inglés también arroja un promedio mayor de 10 palabras. Una mayor producción léxica de estos últimos informantes se da igualmente en *Los animales*, *Partes del cuerpo humano* y *La escuela: sus muebles y materiales*.

Como resultado, en cuanto a cuáles constituyen los centros más productivos en ELE, los cuatro estratos según el nivel de inglés coinciden en tres áreas nocionales: *Alimentos y bebidas*, *La ciudad* y *Partes del cuerpo humano*. En lo referente a los centros no comunes en todos los estratos, señalamos que *La escuela: sus muebles y materiales* se cuenta entre los campos semánticos más productivos en ELE en los estratos del nivel básico, intermedio y superior; *Profesiones y oficios*, en los del nivel básico, intermedio y avanzado; y *Los animales*, en los estratos del nivel avanzado y superior.

En lo que se refiere a los campos semánticos menos productivos en ELE, observamos que la menor producción léxica en ELE en los estudiantes de nivel de inglés básico se da en *Iluminación y calefacción*, seguido por *Objetos colocados en la mesa para la comida*, *Trabajos del campo y del jardín*, *La cocina y sus utensilios* y *El campo*. En el estrato del nivel intermedio de inglés, la menor productividad léxica cae en *Iluminación y calefacción*, seguido por

Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Trabajos del campo y del jardín y Ropa y complementos. En cambio, los estudiantes del nivel intermedio de inglés generan la menor cantidad de palabras en *La cocina y sus utensilios*, seguido por *Objetos colocados en la mesa para la comida, Iluminación y calefacción, Partes de la casa (sin muebles)* y *Trabajos del campo y del jardín*. Por último, el estrato del nivel superior tiene el menor promedio de palabras en *Iluminación y calefacción*, seguido por *Objetos colocados en la mesa para la comida, Los muebles de la casa, La cocina y sus utensilios* y *Partes de la casa (sin muebles)*.

Como resultado, en cuanto a cuáles constituyen los centros menos productivos en ELE, los cuatro estratos según el nivel de inglés coinciden en tres áreas nocionales: *Iluminación y calefacción, Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios*. En lo que se refiere a las diferencias, *Trabajos del campo y del jardín* se cuenta entre los campos semánticos menos productivos en ELE en los estratos del nivel de inglés básico, intermedio y avanzado; *Partes de la casa (sin muebles)*, en los estratos del nivel avanzado y superior; *El campo*, únicamente en el estrato del nivel básico de inglés; *Ropa y complementos*, únicamente en los informantes del nivel intermedio; y *Los muebles de la casa*, solamente en los alumnos del nivel de inglés superior.

Promedio de palabras por centro de interés en ELE

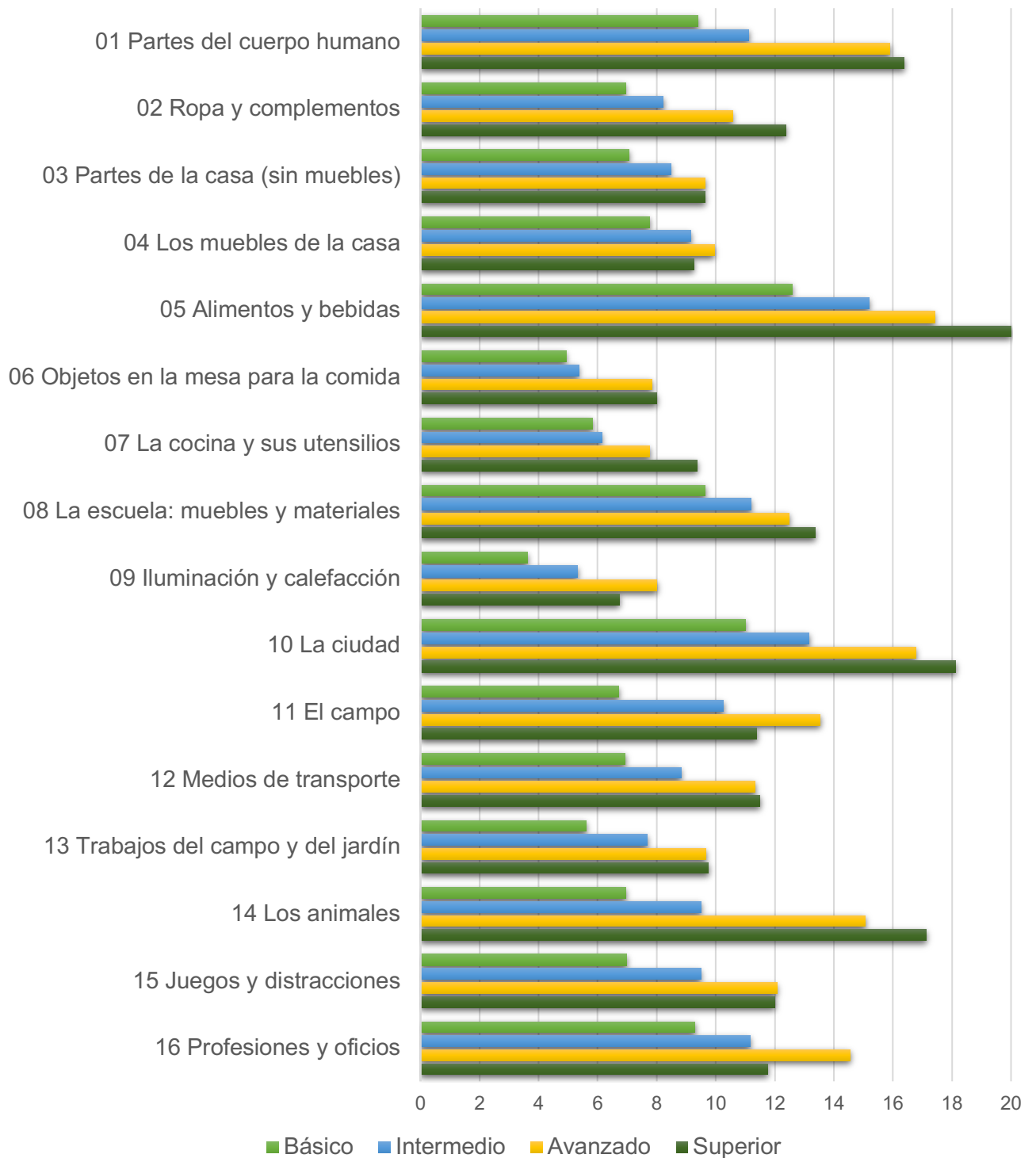


Gráfico 5.39. Promedio de palabras por informante por centro de interés en ELE según la variable nivel de inglés

Ahora nos fijamos en el Gráfico 5.39, que muestra las diferencias entre los sujetos de los cuatro niveles de inglés con respecto al promedio de palabras

por informante en ELE en cada uno de los centros de interés. Si nos centramos en los resultados de los niveles básico, intermedio y avanzado, llegamos a ver, en todos los campos semánticos y sin excepciones, una tendencia de evolución bastante estable de nivel de inglés bajo a nivel de inglés alto. Es decir, la producción léxica en ELE incrementa de los estudiantes de nivel de inglés básico a los de nivel intermedio, y de los de nivel intermedio a los de nivel avanzado. Por otra parte, los resultados del estrato de nivel de inglés superior no siempre manifiestan ventaja sobre el estrato de nivel avanzado, pero sí superan los promedios de palabras de los alumnos de nivel de inglés intermedio y, por consiguiente, básico. De este modo, se supone que, independientemente de cuál sea el centro de interés, la productividad léxica en ELE de los estudiantes crece a medida que incrementa su nivel de inglés.

La razón por la que el estrato de nivel de inglés superior no siempre manifiesta ventaja sobre el estrato de nivel avanzado puede darse en el reducido número de integrantes de este último estrato, tal y como hemos comentado al inicio de este apartado, toda vez que, cuantos menos informantes tenga una sub-muestra, menos fiable puede ser su resultado. De todas maneras, el Gráfico 5.39 nos da una idea de la evolución léxica aun tratándose del estrato del nivel superior de inglés: aunque no siempre, en diez de los dieciséis centros de interés sí se observa una ventaja de este estrato sobre el de nivel avanzado de inglés. Dado todo lo arriba indicado, llegamos a la conclusión de que, en términos generales, la productividad léxica en ELE de los estudiantes incrementa a medida que aumenta su nivel de inglés.

5.2.4.2. De aquí en adelante, presentamos, como referencia de comparación, los resultados sobre el promedio de palabras por informante del corpus en CLM según los cuatro niveles de inglés. Del mismo modo como hemos hecho en los apartados previos sobre ELE, exponemos abajo la Tabla 5.46, en la que se ve dicho índice en CLM en cada centro de interés por nivel de inglés; mientras que en la Tabla 5.47 se puede comprobar el rango de los centros también por nivel de inglés según el promedio de palabras por informante en CLM.

Centro de interés	Promedio de palabras en CLM			
	Básico	Intermedio	Avanzado	Superior
01 Partes del cuerpo humano	17.59	19.98	20.35	17.50
02 Ropa y complementos	16.45	17.63	17.98	16.00
03 Partes de la casa (sin muebles)	11.73	13.24	14.23	13.50
04 Los muebles de la casa	14.30	15.98	15.84	13.50
05 Alimentos y bebidas	19.84	21.22	21.19	18.50
06 Objetos en la mesa para la comida	10.45	11.80	11.96	13.50
07 La cocina y sus utensilios	12.67	15.43	15.11	13.50
08 La escuela: muebles y materiales	15.81	17.20	17.40	15.00
09 Iluminación y calefacción	11.33	12.49	12.96	14.50
10 La ciudad	16.33	18.76	17.86	21.00
11 El campo	13.84	17.67	18.44	19.00
12 Medios de transporte	12.88	15.06	15.40	16.00
13 Trabajos del campo y del jardín	10.38	13.24	13.37	16.50
14 Los animales	20.44	22.59	22.88	23.00
15 Juegos y distracciones	13.67	17.29	17.56	18.50
16 Profesiones y oficios	16.55	18.94	19.46	19.50
Global	14.64	16.78	17.00	16.81

Tabla 5.46. Promedio de palabras por informante por centro de interés en CLM según la variable nivel de inglés

Centro de interés	Rango según el promedio de palabras por informante en CLM			
	Básico	Intermedio	Avanzado	Superior
01 Partes del cuerpo humano	3	3	3	7
02 Ropa y complementos	5	7	6	9
03 Partes de la casa (sin muebles)	13	13	13	13
04 Los muebles de la casa	8	10	10	13
05 Alimentos y bebidas	2	2	2	5
06 Objetos en la mesa para la comida	15	16	16	13
07 La cocina y sus utensilios	12	11	12	13
08 La escuela: muebles y materiales	7	9	9	11
09 Iluminación y calefacción	14	15	15	12
10 La ciudad	6	5	7	2
11 El campo	9	6	5	4
12 Medios de transporte	11	12	11	9
13 Trabajos del campo y del jardín	16	13	14	8
14 Los animales	1	1	1	1
15 Juegos y distracciones	10	8	8	5
16 Profesiones y oficios	4	4	4	3

Tabla 5.47. Rango de los centros de interés por nivel de inglés según el promedio de palabras por informante en CLM

Un vistazo a las cifras expuestas en la Tabla 5.46 nos permite saber que la mayor producción léxica en CLM de todos los centros de interés de los cuatro niveles de inglés se encuentra en *Los animales* del estrato del nivel superior —con un promedio de 23 palabras—. Por el contrario, la menor producción léxica en CLM se halla en *Trabajos del campo y del jardín* del nivel básico —con una media de 10.38 palabras—. Estos resultados en CLM no contradicen los resultados previamente extraídos en ELE según el nivel de inglés, ya que en ambas lenguas son los estratos del nivel de inglés superior y básico los que arrojan, respectivamente, el mayor y el menor promedio de palabras por informante.

Ahora pasamos al análisis sobre dicho índice en CLM por nivel de inglés. Como podemos ver en la Tabla 5.47, las cuatro sub-muestras con niveles de inglés distintos coinciden totalmente en el campo semántico con la mayor producción léxica en CLM: *Los animales*. Aparte de este centro de interés, los alumnos de nivel básico tienen un mayor promedio de palabras en CLM también en *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *Ropa y complementos*; los informantes de nivel intermedio, en *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad*; los estudiantes de nivel avanzado de inglés, en *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *El campo*; los de nivel superior, en *La ciudad*, *Profesiones y oficios*, *El campo* y *Alimentos y bebidas*.

Nos damos cuenta de que *Los animales*, *Alimentos y bebidas* y *Profesiones y oficios* son tres áreas temáticas más productivas en CLM, comunes en estudiantes de todos los niveles de inglés. Comparando estos resultados con los previamente extraídos en ELE según el nivel de inglés, solo *Alimentos y bebidas* es el centro de interés común en ELE y en CLM en cuanto a tener un alto grado de productividad léxica en el conjunto de los dieciséis centros.

Además, los estratos de nivel básico, intermedio y avanzado cuenta con una mayor producción léxica en CLM en *Partes del cuerpo humano*; los de nivel intermedio y superior, en *La ciudad*; los de nivel avanzado y superior, en *El campo*; mientras únicamente el estrato de nivel básico tiene un mayor promedio de palabras en CLM también en *Ropa y complementos*.

En lo que se refiere a los campos semánticos menos productivos en CLM, la Tabla 5.47 informa de que, en los alumnos de nivel básico de inglés, la menor producción léxica se registra en *Trabajos del campo y del jardín*, seguido por *Objetos colocados en la mesa para la comida*, *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*. En cambio, la menor productividad léxica del estrato del nivel intermedio se halla en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*,

Partes de la casa (sin muebles), Trabajos del campo y del jardín y Medios de transporte. Por otra parte, los estudiantes del nivel avanzado evocan la menor cantidad de palabras en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción, Trabajos del campo y del jardín, Partes de la casa (sin muebles) y La cocina y sus utensilios*. Por último, el estrato del nivel superior tiene el menor promedio de palabras en cuatro campos asociativos: *Partes de la casa (sin muebles), Los muebles de la casa, Objetos colocados en la mesa para la comida y La cocina y sus utensilios*, seguidos por *Iluminación y calefacción*.

Vemos que *Partes de la casa (sin muebles), Objetos colocados en la mesa para la comida e Iluminación y calefacción* son comunes en todos los niveles de inglés con un menor promedio de palabras por informante en CLM. Comparado estos resultados con los previamente extraídos en ELE según el nivel de inglés, sabemos que *Objetos colocados en la mesa para la comida e Iluminación y calefacción* son dos centros con una menor productividad léxica en todos los niveles de inglés tanto en ELE como en CLM.

En lo que se refiere a las diferencias detectadas sobre los centros menos productivos en CLM, *Trabajos del campo y del jardín* se halla en los estratos del nivel básico, intermedio y avanzado como otro campo léxico menos productivo. *La cocina y sus utensilios* se cuenta entre los centros menos productivos en CLM en los informantes del nivel básico, avanzado y superior. Como casos únicos, solo el estrato del nivel intermedio cuenta con *Medios de transporte* entre sus campos léxicos menos productivos en CLM, y solo el estrato del nivel superior, con *Los muebles de la casa*.

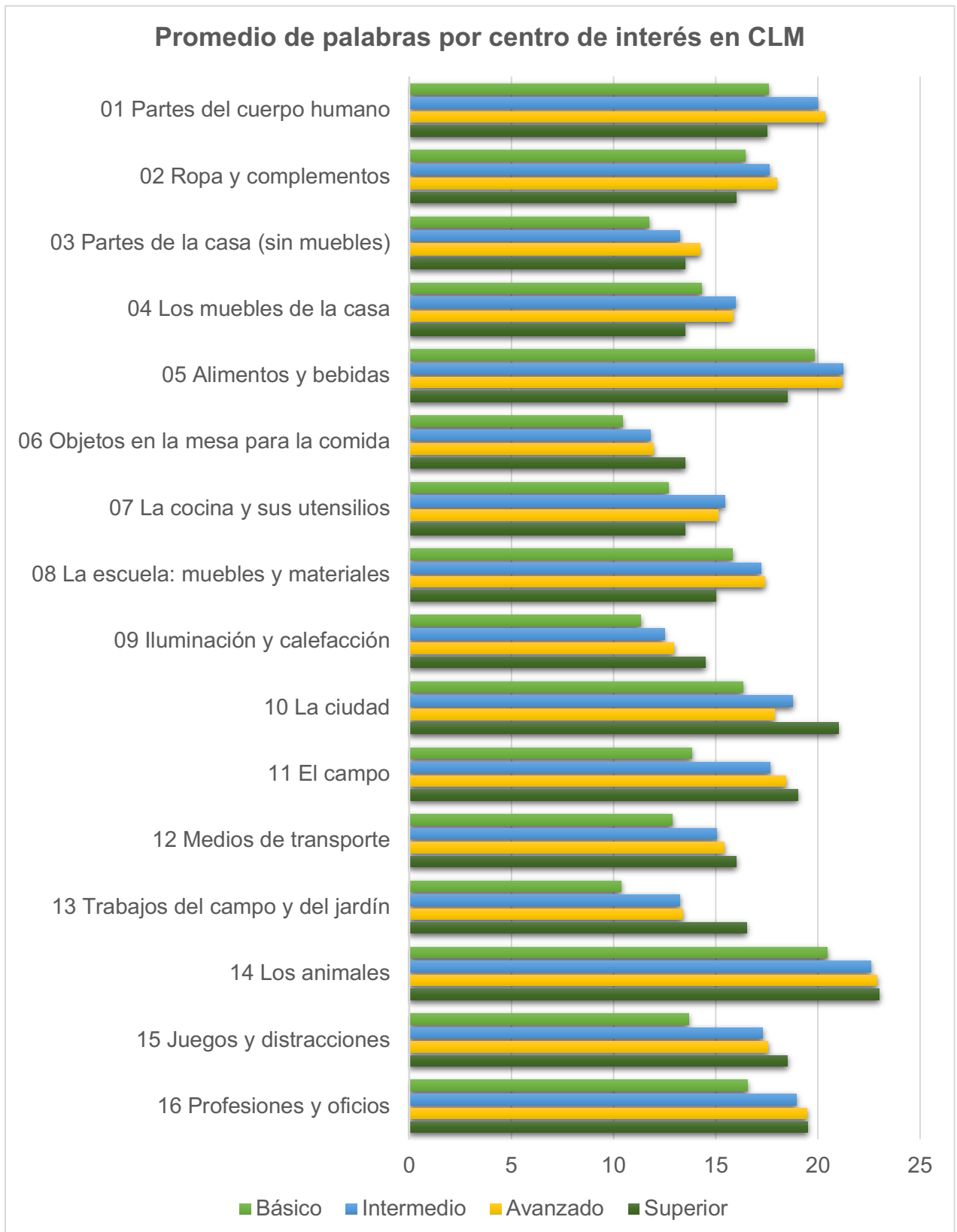


Gráfico 5.40. Promedio de palabras por informante por centro de interés en CLM según la variable nivel de inglés

Si nos fijamos en el Gráfico 5.40, que expone las diferencias entre los

sujetos de los cuatro niveles de inglés con respecto a la producción léxica en CLM en cada uno de centros de interés, parece que volvemos a observar la misma tendencia de evolución léxica que la que hemos encontrado en el corpus de ELE. Los resultados de los niveles básico, intermedio y avanzado evidencian que, en la inmensa mayoría de los centros de interés —o sea, en doce áreas nocionales—, la producción léxica en CLM incrementa de los estudiantes de nivel de inglés básico a los de nivel intermedio, y de los de nivel intermedio a los de nivel avanzado.

En los otros cuatro centros de interés —*Los muebles de la casa, Alimentos y bebidas, La cocina y sus utensilios y La ciudad*—, el incremento se detecta del nivel básico al intermedio y se ve un retroceso del nivel intermedio al avanzado. Aun así, nos damos cuenta de que dicho retroceso es realmente ligero en tres de los cuatro centros, puesto que observamos un retroceso de 0.14, 0.03 y 0.32 palabras en *Los muebles de la casa, Alimentos y bebidas y La cocina y sus utensilios*, respectivamente. Únicamente en *La ciudad*, parece que el estrato de nivel avanzado de inglés tiene una deficiencia significativa, presentado un retroceso de 0.9 palabras.

Tal como hemos visto en el corpus de ELE, el estrato de nivel de inglés superior no arroja resultados en CLM del todo fiables por su reducido número de integrantes. Comprobamos que los índices del estrato de nivel superior no siempre manifiestan ventaja; pero, también superan los promedios del estrato de nivel avanzado en nueve centros de interés.

Teniendo en cuenta todo lo anterior, deducimos que la productividad léxica en CLM de los estudiantes incrementa a medida que aumenta su nivel de inglés, en términos generales. A continuación, a partir de los datos globales expuestos en las Tablas 5.44 y 5.46, presentamos en los Gráficos 5.41 y 5.42 el promedio de palabras global de cada uno de los cuatro niveles de inglés en los corpus de ELE y de CLM.

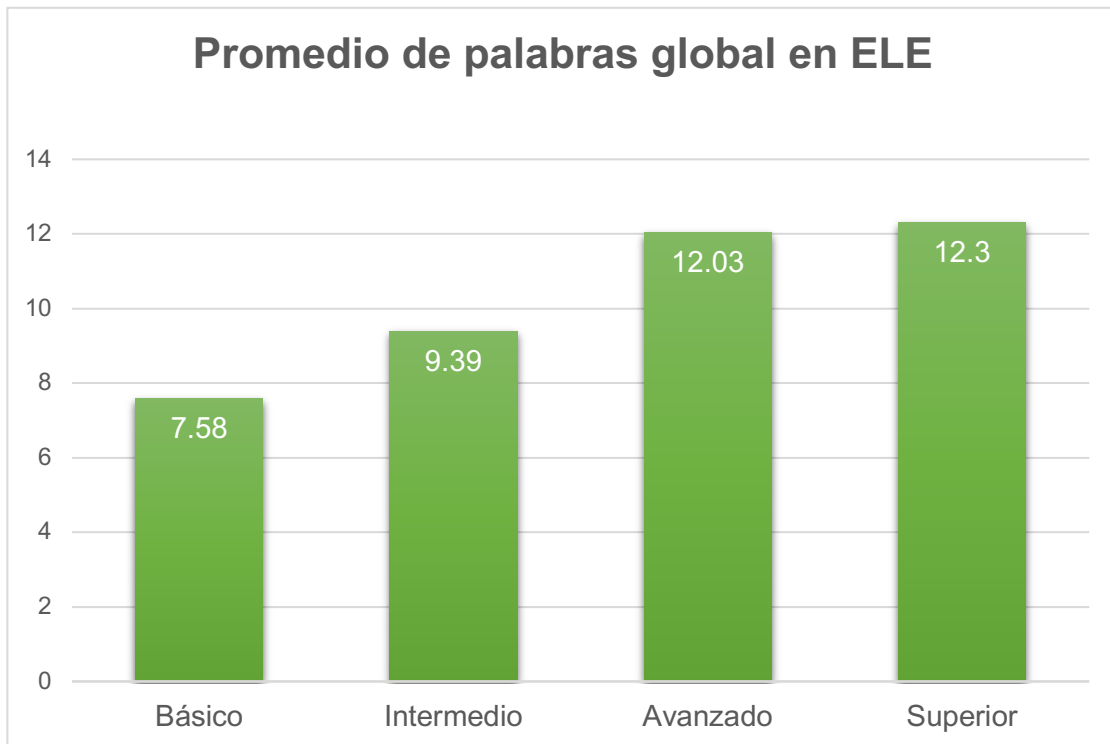


Gráfico 5.41. Promedio global de palabras por informante en ELE según la variable nivel de inglés

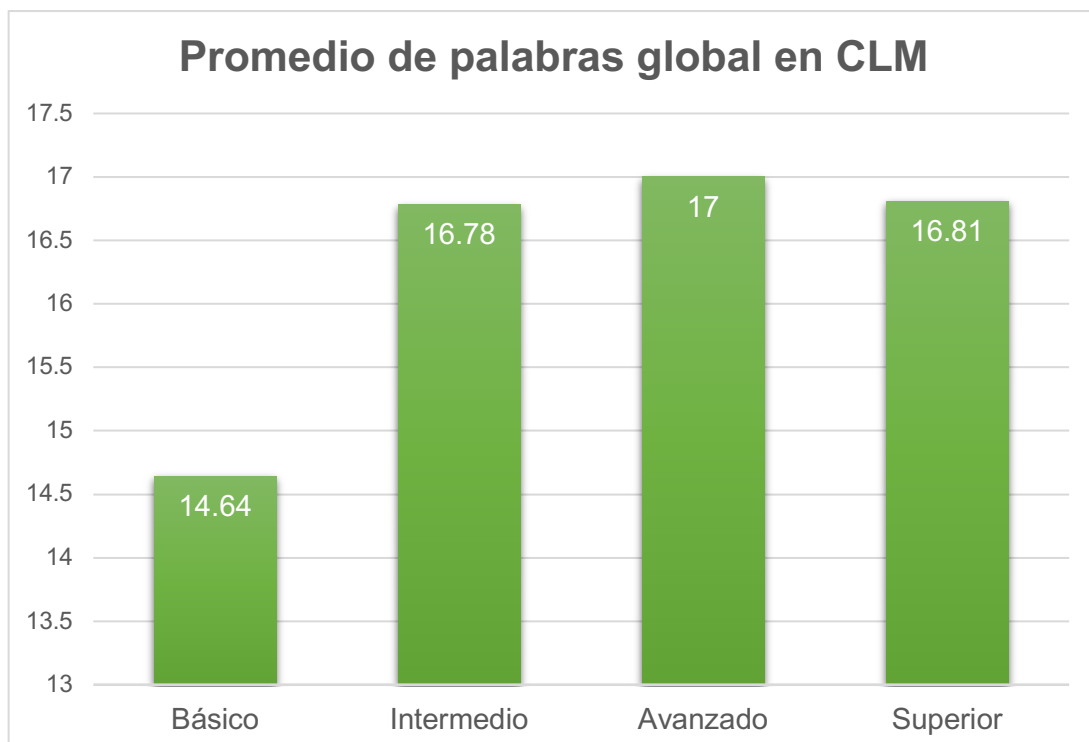


Gráfico 5.42. Promedio global de palabras por informante en CLM según la variable nivel de inglés

Como muestran los dos gráficos de arriba, vemos comprobadas las deducciones anteriormente sacadas por centro de interés sobre la evolución léxica que se corresponden con el aumento del nivel de inglés, pero, esta vez, de manera global e independientemente de las áreas temáticas. Podemos ver en ambos gráficos cómo la producción léxica media de los informantes asciende, en ELE y en CLM, del estrato del nivel de inglés bajo al estrato de nivel alto: concretamente, del estrato de nivel básico al de nivel intermedio y del estrato de este nivel al de nivel avanzado.

En el corpus de ELE, podemos ver dicho incremento global de la productividad léxica incluso en los informantes de nivel superior de inglés, hasta llegar a darse un promedio de 12.3 palabras por informante. Mientas tanto, en el corpus de CLM, a pesar de que los sujetos de nivel superior no muestran ventaja global sobre los de nivel avanzado, sí estamos advertidos de una evolución léxica en el nivel superior con respecto a los niveles básico e intermedio.

Por otra parte, comparando los Gráficos 5.41 y 5.42 entre ellos mismos, parece que los estudiantes de niveles de inglés distintos no revelan una misma divergencia de producción léxica entre ELE y CLM. De modo que vamos a examinar en el siguiente apartado, por nivel de inglés, la distancia de producción léxica entre las dos lenguas.

5.2.4.3. La Tabla 5.48 reúne los datos de las Tablas 5.44 y 5.46, presentando, a su vez, los resultados completos sobre el promedio de palabras por informante por nivel de inglés y en las dos lenguas de nuestra investigación. A partir de ella, obtenemos la Tabla 5.49, que muestra, por centro de interés y por nivel de inglés, las diferencias cuantitativas entre el promedio en ELE y el promedio en CLM.

Centro de interés	Promedio de palabras							
	Básico		Intermedio		Avanzado		Superior	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	9.40	17.59	11.12	19.98	15.90	20.35	16.38	17.50
02 Ropa	6.94	16.45	8.22	17.63	10.57	17.98	12.38	16.00
03 Casa	7.07	11.73	8.47	13.24	9.65	14.23	9.63	13.50
04 Muebles	7.76	14.30	9.15	15.98	9.95	15.84	9.25	13.50
05 Alimentos	12.59	19.84	15.20	21.22	17.41	21.19	20.00	18.50
06 Objetos	4.93	10.45	5.36	11.80	7.83	11.96	8.00	13.50
07 Cocina	5.82	12.67	6.14	15.43	7.77	15.11	9.38	13.50
08 Escuela	9.64	15.81	11.20	17.20	12.49	17.40	13.38	15.00
09 Iluminación	3.62	11.33	5.31	12.49	8.01	12.96	6.75	14.50
10 Ciudad	11.01	16.33	13.16	18.76	16.78	17.86	18.13	21.00
11 Campo	6.71	13.84	10.26	17.67	13.53	18.44	11.38	19.00
12 Transporte	6.93	12.88	8.82	15.06	11.32	15.40	11.50	16.00
13 Trabajos	5.61	10.38	7.68	13.24	9.67	13.37	9.75	16.50
14 Animales	6.95	20.44	9.51	22.59	15.05	22.88	17.13	23.00
15 Juegos	6.99	13.67	9.49	17.29	12.07	17.56	12.00	18.50
16 Profesiones	9.30	16.55	11.16	18.94	14.55	19.46	11.75	19.50
Global	7.58	14.64	9.39	16.78	12.03	17.00	12.30	16.81

Tabla 5.48. Promedio de palabras por informante por centro de interés en ELE y en CLM según la variable nivel de inglés

Centro de interés	Diferencia de promedio de palabras entre ELE y CLM			
	Básico	Intermedio	Avanzado	Superior
01 Partes del cuerpo humano	8.19	8.86	4.45	1.12
02 Ropa y complementos	9.51	9.41	7.41	3.62
03 Partes de la casa (sin muebles)	4.66	4.77	4.58	3.87
04 Los muebles de la casa	6.54	6.83	5.89	4.25
05 Alimentos y bebidas	7.25	6.02	3.78	-1.50
06 Objetos en la mesa para la comida	5.52	6.44	4.13	5.50
07 La cocina y sus utensilios	6.85	9.29	7.34	4.12
08 La escuela: muebles y materiales	6.17	6.00	4.91	1.62
09 Iluminación y calefacción	7.71	7.18	4.95	7.75
10 La ciudad	5.32	5.60	1.08	2.87
11 El campo	7.13	7.41	4.91	7.62
12 Medios de transporte	5.95	6.24	4.08	4.50
13 Trabajos del campo y del jardín	4.77	5.56	3.70	6.75
14 Los animales	13.49	13.08	7.83	5.87
15 Juegos y distracciones	6.68	7.80	5.49	6.50
16 Profesiones y oficios	7.25	7.78	4.91	7.75
Global	7.06	7.39	4.97	4.51

Tabla 5.49. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable nivel de inglés

Como podemos ver en la Tabla 5.49, las cifras revelan las distancias de producción léxica entre las dos lenguas, con respecto al promedio de palabras en CLM. La producción léxica en ELE es inferior a la que hay en CLM en todos los centros de interés y en todos los niveles de inglés, con excepción de *Alimentos y bebidas* del estrato del nivel superior (-1.5), donde el promedio en ELE supera dicho índice en CLM con 1.5 palabras más. Hemos visto en los apartados anteriores que los informantes de nivel superior de inglés muestran resultados sobresalientes en ELE. Ahora observamos que esos informantes tienen una productividad léxica aún más relevante en ELE en *Alimentos y bebidas*, puesto que es la única área nocional en la que un estrato presenta un promedio de palabras en ELE superior al promedio en CLM.

Además, la tabla nos informa de que la mayor diferencia entre los promedios de ELE y de CLM se halla en *Los animales* en el estrato del nivel de inglés básico (13.49). Cabe destacar que la segunda mayor diferencia también se registra en este centro de interés en el estrato del nivel intermedio (13.08).

Las celdas marcadas en fondo gris en la Tabla 5.49 indican los campos semánticos en los que la diferencia cuantitativa está por encima del promedio global o, en otras palabras, campos en los que la producción léxica en ELE de los informantes dista enormemente de la que se da en CLM. Como hemos recordado en los apartados anteriores, sugerimos que se preste atención a estas áreas semánticas en la enseñanza-aprendizaje de la lengua extranjera, ya que la inferioridad en ELE en ellas puede compensarse a través de adquirir los homólogos en español de las palabras que los estudiantes son capaces de evocar en CLM.

Los animales es el centro de interés en el que los alumnos de todos los niveles de inglés coinciden al respecto; es decir, todos los estudiantes, independientemente de su nivel de inglés, muestran considerable debilidad en *Los animales* en ELE, tomando como referencia su producción léxica en CLM. Además, hay tres campos léxicos que implican deficiencia en ELE con respecto a la producción en CLM en tres de los estratos. Señalamos que los estratos del nivel básico, intermedio y avanzado tienen mucha deficiencia en ELE en *Ropa y complementos*; los estratos del nivel intermedio, avanzado y superior, en *Juegos y distracciones*; y los estratos del nivel básico, intermedio y superior, en *Profesiones y oficios*. Como casos únicos, solo los alumnos de nivel básico de inglés presentan debilidad en *Alimentos y bebidas*; solo los de nivel avanzado, en *Los muebles de la casa*; y solo los de nivel superior, en *Objetos colocados en la mesa para la comida y Trabajos del campo y del jardín*.

Así estando las cosas, pasamos a fijarnos en la diferencia global de promedio de palabras entre ELE y CLM por nivel de inglés, la cual se expone

en la última fila de la Tabla 5.49 y, de manera ilustrativa, en el Gráfico 5.43. Con dicha diferencia global.

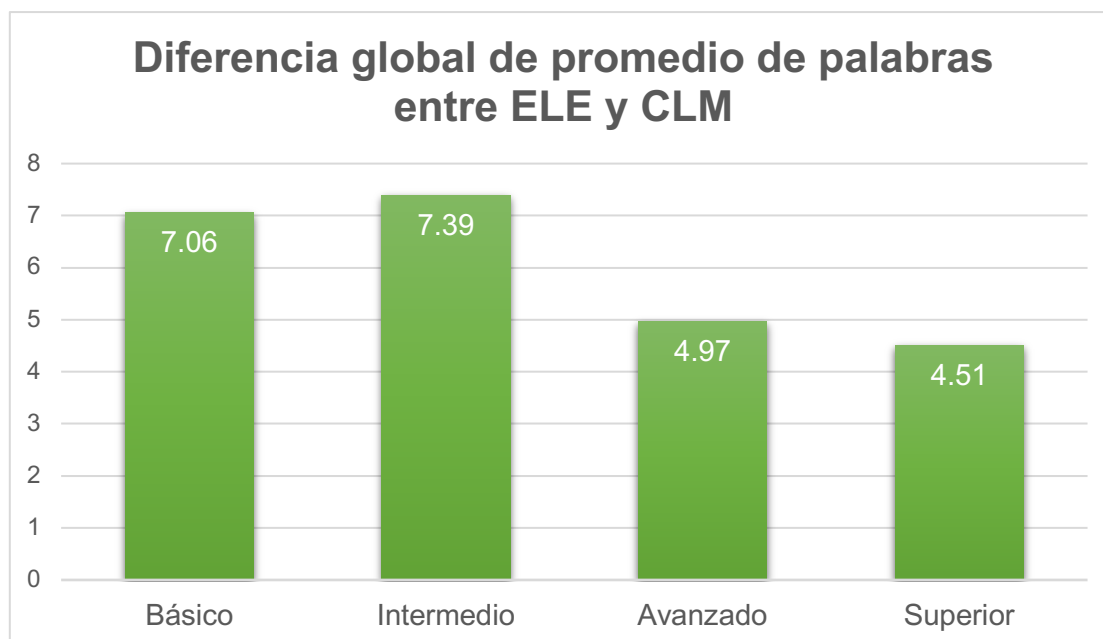


Gráfico 5.43. Diferencia global de promedio de palabras entre ELE y CLM según la variable nivel de inglés

Como podemos ver en el Gráfico 5.43, los estratos del nivel básico e intermedio arrojan resultados parecidos (7.06 y 7.39 respectivamente) con respecto a la diferencia global entre los promedios de palabras en ELE y en CLM. De igual modo, se observan dos valores de diferencia próximos (4.97 y 4.51, respectivamente) en los estratos del nivel avanzado y superior. Concretamente, los alumnos de nivel de inglés básico e intermedio aportan un poco más de 7 palabras menos en ELE que en CLM en un campo semántico medio; pero, en los estudiantes del nivel de inglés avanzado y superior, solo se detectan entre 4 y 5 palabras menos en ELE que en CLM. Aunque las cifras globales del Gráfico 5.43 no van disminuyendo a medida que incrementa el nivel de inglés de los informantes —puesto que se registra un repunte leve de 0.33 del nivel básico al intermedio—, sí podemos comprobar que la diferencia que existe entre la producción léxica en ELE y en CLM va disminuyendo considerablemente de los estudiantes de niveles de inglés bajos a los de niveles de inglés altos. A continuación, vamos a ver si esta tendencia también se percibe por centro de interés. Para ello, observaremos el Gráfico 5.44, que

resulta ser una visualización de los datos presentados en la Tabla 5.49.

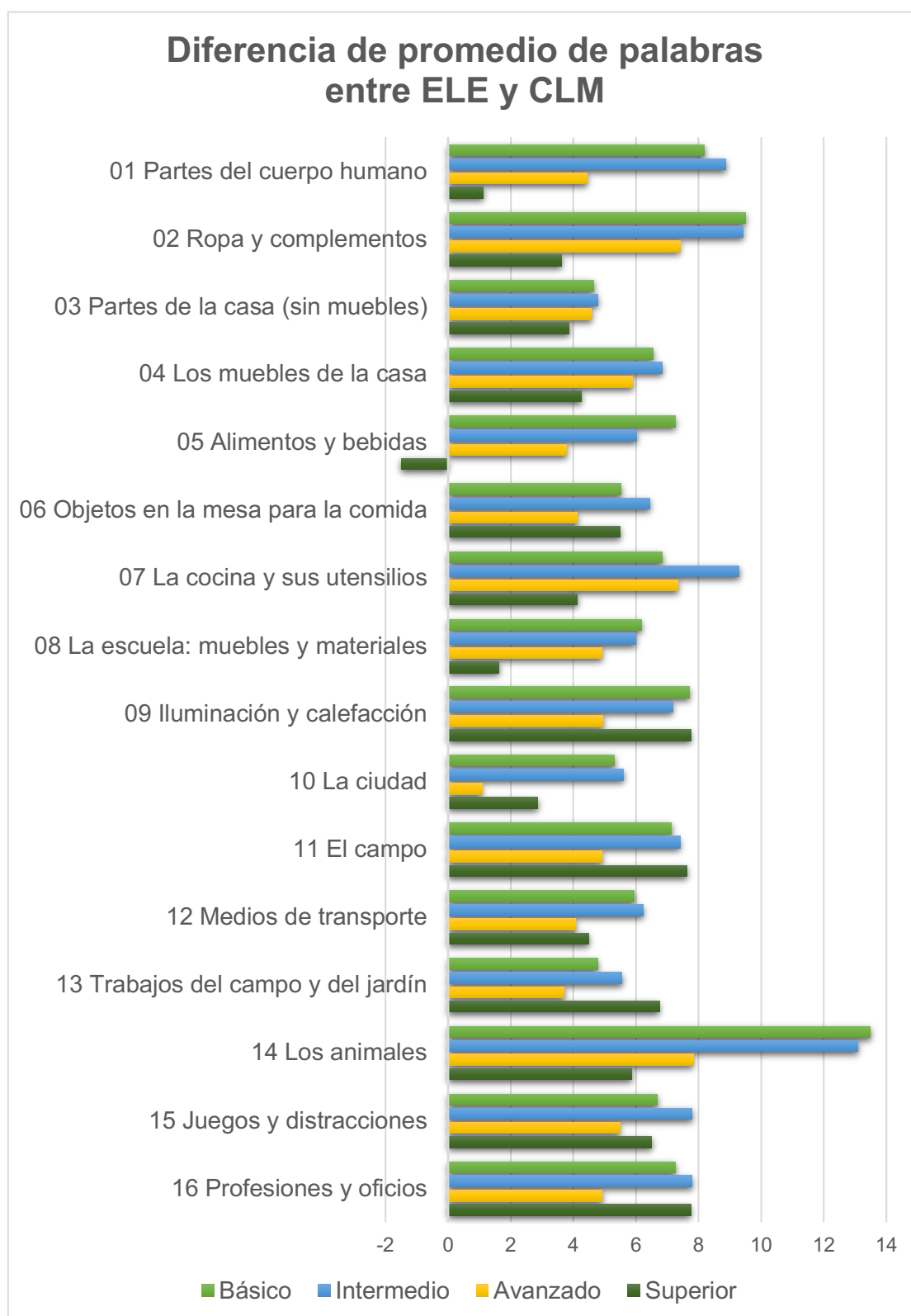


Gráfico 5.44. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable nivel de inglés

En el Gráfico 5.44, podemos observar que las diferencias de producción léxica entre ELE y CLM van reduciéndose desde niveles de inglés bajos hasta niveles altos en la mayoría de las áreas nocionales —excepto en *La cocina y sus utensilios*, *Iluminación y calefacción*, *El campo*, *Trabajos del campo y del jardín* y *Profesiones y oficios*—. Si exceptuamos los resultados del estrato del nivel superior dado su limitado número de integrantes, vemos que, únicamente en *La cocina y sus utensilios*, no resulta muy evidente la reducción de diferencia de producción léxica entre las dos lenguas desde niveles de inglés bajos hasta niveles altos. Parece que en *La cocina y sus utensilios* los alumnos de nivel básico de inglés muestran una ventaja en su productividad léxica en ELE en comparación con la en CLM, pero en el resto de niveles de inglés en ese centro de interés, volvemos a ver dicha reducción de diferencia relacionada con el nivel de inglés.

Por lo tanto, llegamos a ver que, en referencia a la productividad léxica que tienen en CLM, los informantes de niveles de inglés altos señalan mejores resultados en ELE que los de niveles bajos. Así pues, confirmamos el hecho de que los estudiantes generan una producción léxica en ELE cada vez más aproximada a la que tienen en CLM, a medida que van ascendiendo de nivel de inglés.

5.2.4.4. A continuación, vamos a examinar, por nivel de inglés y según la lengua de la encuesta, el total de vocablos que se recaban en cada uno de los centros de interés.

Centro de interés	Total de vocablos							
	Básico		Intermedio		Avanzado		Superior	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	113	162	167	128	161	159	58	22
02 Ropa	128	220	165	198	181	208	62	27
03 Casa	97	131	118	142	155	144	38	23
04 Muebles	103	158	115	154	146	166	38	22
05 Alimentos	175	394	244	355	270	377	91	34
06 Objetos	64	105	74	109	131	117	24	21
07 Cocina	115	154	136	162	163	164	39	25
08 Escuela	144	192	183	197	244	210	66	28
09 Iluminación	73	144	124	147	164	156	30	25
10 Ciudad	294	439	368	412	452	432	109	37
11 Campo	241	378	285	342	354	411	66	33
12 Transporte	109	153	179	166	205	201	43	27
13 Trabajos	189	212	238	232	292	248	63	30
14 Animales	114	212	161	184	188	214	76	37
15 Juegos	153	209	233	225	255	251	63	31
16 Profesiones	151	252	235	249	283	264	60	36
Global	2263	3515	3025	3402	3644	3722	926	458

Tabla 5.50. Total de vocablos por centro de interés en ELE y en CLM según la variable nivel de inglés

En la Tabla 5.50 se expone el número total de vocablos de cada uno de los dieciséis centros de interés según sea el nivel de inglés y la lengua. A partir de ahí, podemos darnos cuenta de la variedad léxica de los estudiantes de cada nivel de inglés en cada una de las dos lenguas contempladas.

Recordamos que, ante sub-muestras que implican entre sí unas diferencias considerables que se dan en el número de informantes, el total de vocablos no resulta un índice del todo fiable. Como hemos llamado la atención a lo largo de los apartados, en el caso del nivel de inglés, los estudiantes de nivel superior solo suponen un 3% —ocho estudiantes— del total de

encuestados en ELE y un 1% —dos estudiantes—, en CLM. Podemos ver cómo la diferencia enorme en el número de informantes repercute en los resultados de los vocablos según el nivel de inglés. Es evidente que el menor número de informantes en el estrato del nivel superior se relaciona con una menor riqueza léxica, tanto en ELE como en CLM y tanto en los resultados globales como en los resultados por centro de interés. De modo que, a continuación, nos centramos en los resultados de los otros tres estratos —los de nivel de inglés básico, intermedio y avanzado—.

Los resultados globales que se encuentran en la última fila de la tabla nos permiten sacar las siguientes conclusiones. Por un lado, la evolución de la riqueza léxica en ELE de los estudiantes manifiesta una tendencia igual que la que hemos observado en la productividad léxica en ELE de los mismos; es decir, va ascendiendo del nivel básico (2263) al intermedio (3025), y del intermedio al avanzado (3644). Cuanto más alto es su nivel de inglés, mayor riqueza léxica en ELE alcanzan los estudiantes. De modo similar se presenta la evolución de riqueza léxica en CLM: en los niveles básico e intermedio, los alumnos muestran una riqueza léxica relativamente estable (3515 y 3402, respectivamente), mientras que, al pasar al nivel avanzado, se halla un progreso considerable de la variedad léxica (3722).

Por otro lado, si comparamos, por nivel de inglés, la riqueza léxica en ELE y en CLM, podemos ver que, en el estrato del nivel básico, la riqueza léxica en CLM (3515) es abrumadoramente mayor que la que hay en ELE (2263); pero, al pasar al nivel intermedio, los estudiantes empiezan a mostrar menor distancia entre ELE (3025) y CLM (3402). Aún más, en el estrato del nivel avanzado, los informantes incluso presentan resultados próximos en ELE (3644) y en CLM (3722). Ello nos conduce a una deducción similar a la que hemos tenido al analizar la productividad léxica, esto es, a medida que aumenta el nivel de inglés, la diferencia de riqueza léxica entre ELE y CLM se convierte en cada vez menor. Los alumnos de niveles bajos de inglés no tienen en ELE una riqueza léxica comparable con la que tienen en CLM; en cambio, los de niveles altos dominan una variedad de vocablos más próximo entre ELE

y CLM.

Ahora nos centramos en los resultados de las distintas áreas nocionales por nivel de inglés. Como es lógico, el estrato del nivel básico manifiesta una variedad léxica menor en ELE que en CLM en todos los centros de interés. Esta tendencia se mantiene en el estrato del nivel intermedio en general, pero, los alumnos de este nivel de inglés ya señalan mejores resultados en ELE que en CLM en cuatro centros de interés: *Partes del cuerpo humano*, *Medios de transporte*, *Trabajos del campo y del jardín* y *Juegos y distracciones* —aunque levemente en estas últimas tres áreas—. En cambio, en el estrato del nivel avanzado de inglés, los alumnos obtienen mejores datos en ELE que en CLM ya en la mayoría de los campos semánticos —concretamente en diez campos léxicos: *Partes del cuerpo humano*, *Partes de la casa (sin muebles)*, *Objetos colocados en la mesa para la comida*, *La escuela: sus muebles y materiales*, *Iluminación y calefacción*, *La ciudad*, *Medios de transporte*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*—. Vemos, otra vez, cómo los alumnos van alcanzando una mayor riqueza léxica en ELE —con respecto a su competencia léxica en CLM— desde los niveles de inglés bajos hasta los niveles altos.

Centro de interés	Rango según el total de vocablos							
	Básico		Intermedio		Avanzado		Superior	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	11	10	9	15	13	13	10	14
02 Ropa	8	5	10	7	10	9	8	9
03 Casa	14	15	14	14	14	15	13	13
04 Muebles	13	11	15	12	15	11	13	14
05 Alimentos	4	2	3	2	5	3	2	4
06 Objetos	16	16	16	16	16	16	16	16
07 Cocina	9	12	12	11	12	12	12	11
08 Escuela	7	9	7	8	7	8	4	8
09 Iluminación	15	14	13	13	11	14	15	11
10 Ciudad	1	1	1	1	1	1	1	1
11 Campo	2	3	2	3	2	2	4	5
12 Transporte	12	13	8	10	8	10	11	9
13 Trabajos	3	6	4	5	3	6	6	7
14 Animales	10	6	11	9	9	7	3	1
15 Juegos	5	8	6	6	6	5	6	6
16 Profesiones	6	4	5	4	4	4	9	3

Tabla 5.51. Rango de los centros de interés por nivel de inglés según el total de vocablos en ELE y en CLM

La Tabla 5.51 nos informa del rango de los centros de interés por nivel de inglés según cuánta sea la riqueza léxica en las dos lenguas respectivas. De un vistazo podemos ver que *La ciudad* es el campo léxico más rico en vocabulario en las dos lenguas unánimemente en todos los estratos de niveles de inglés. *Objetos colocados en la mesa para la comida* es el área nocional menos rica en vocabulario en las dos lenguas también en todos los estratos.

Aparte de dichas áreas semánticas, vemos que el estrato del nivel básico muestra una mayor riqueza léxica en ELE también en *El campo*, *Trabajos del campo y del jardín*, *Alimentos bebidas y Juegos y distracciones*; el estrato del nivel intermedio, en *El campo*, *Alimentos bebidas*, *Trabajos del campo y del*

jardín y Profesiones y oficios; el estrato del nivel avanzado, en *El campo, Trabajos del campo y del jardín, Profesiones y oficios y Alimentos bebidas*; y el estrato del nivel superior, en *Alimentos bebidas, Los animales, La escuela: sus muebles y materiales y El campo*. De ahí, vemos que los informantes de los cuatro niveles de inglés coinciden en tres centros más ricos en vocabulario en ELE: *La ciudad, El campo y Alimentos y bebidas*. Los estudiantes de nivel de inglés básico, intermedio y avanzado también tienen un alto grado de variedad léxica en ELE en *Trabajos del campo y del jardín*; los de nivel intermedio y avanzado, en *Profesiones y oficios*; solo los de nivel básico, en *Juegos y distracciones*; mientras solo los de nivel superior, en *Los animales y La escuela: sus muebles y materiales*.

Por el contrario, además de *Objetos colocados en la mesa para la comida*, el estrato del nivel básico también tiene una menor variedad léxica en ELE en *Iluminación y calefacción, Partes de la casa (sin muebles), Los muebles de la casa y Medios de transporte*; los alumnos de nivel de inglés intermedio, en *Los muebles de la casa, Partes de la casa (sin muebles), Iluminación y calefacción y La cocina y sus utensilios*; los de nivel avanzado, en *Los muebles de la casa, Partes de la casa (sin muebles), Partes del cuerpo humano y La cocina y sus utensilios*; y los de nivel superior, en *Iluminación y calefacción, Partes de la casa (sin muebles), Los muebles de la casa y La cocina y sus utensilios*. Así es que los informantes de los cuatro niveles de inglés coinciden en tres centros menos ricos en vocabulario en ELE: *Objetos colocados en la mesa para la comida, Partes de la casa (sin muebles) y Los muebles de la casa*. En cuanto a las diferencias, *Iluminación y calefacción* implica una menor riqueza léxica en ELE en tres estratos: los niveles de inglés básico, intermedio y superior; *La cocina y sus utensilios* también en tres estratos: los niveles intermedio, avanzado y superior. Como casos únicos, solo el estrato del nivel básico señala una menor riqueza léxica en ELE en *Medios de transporte*, y solo el del nivel avanzado, en *Partes del cuerpo humano*.

Centro de interés	Índice de cohesión							
	Básico		Intermedio		Avanzado		Superior	
	ELE	CLM	ELE	CLM	ELE	CLM	ELE	CLM
01 Cuerpo	0.08	0.11	0.07	0.16	0.10	0.13	0.28	0.80
02 Ropa	0.05	0.07	0.05	0.09	0.06	0.09	0.20	0.59
03 Casa	0.07	0.09	0.07	0.09	0.06	0.10	0.25	0.59
04 Muebles	0.08	0.09	0.08	0.10	0.07	0.10	0.24	0.61
05 Alimentos	0.07	0.05	0.06	0.06	0.06	0.06	0.22	0.54
06 Objetos	0.08	0.10	0.07	0.11	0.06	0.10	0.33	0.64
07 Cocina	0.05	0.08	0.05	0.10	0.05	0.09	0.24	0.54
08 Escuela	0.07	0.08	0.06	0.09	0.05	0.08	0.20	0.54
09 Iluminación	0.05	0.08	0.04	0.08	0.05	0.08	0.23	0.58
10 Ciudad	0.04	0.04	0.04	0.05	0.04	0.04	0.17	0.57
11 Campo	0.03	0.04	0.04	0.05	0.04	0.04	0.17	0.58
12 Transporte	0.06	0.08	0.05	0.09	0.06	0.08	0.27	0.59
13 Trabajos	0.03	0.05	0.03	0.06	0.03	0.05	0.15	0.55
14 Animales	0.06	0.10	0.06	0.12	0.08	0.11	0.23	0.62
15 Juegos	0.05	0.07	0.04	0.08	0.05	0.07	0.19	0.60
16 Profesiones	0.06	0.07	0.05	0.08	0.05	0.07	0.20	0.54

Tabla 5.52. Índice de cohesión por centro de interés en ELE y en CLM según la variable nivel de inglés

Por último, presentamos el índice de cohesión, que relaciona los dos índices anteriormente estudiados —el promedio de palabras por informante y el total de vocablos—. La Tabla 5.52 exhibe el índice de cohesión de cada centro de interés por nivel de inglés y según cuál sea la lengua de la encuesta. Volvemos a advertir de que la sub-muestra del nivel superior de inglés cuenta con un número reducido de informantes. Podemos ver cómo la cantidad de encuestados influye en los resultados obtenidos en la Tabla 5.52. Se observan unos índices de cohesión mayores en el nivel de inglés superior, puesto que, cuantos menos informantes haya en una sub-muestra, mayor índice de cohesión puede darse en la misma. Entre los alumnos de niveles de inglés básico, intermedio y avanzado, parece que los distintos estratos tienen unos

índices de cohesión próximos; es decir, el índice no parece cambiar según el nivel de inglés de los estudiantes.

Entre las dos lenguas, podemos ver que, en las sub-muestras de las dos lenguas bajo cualquier nivel de inglés, el índice de cohesión de cualquier área nocional en ELE es igual o menor que el de la misma área en CLM (solo con excepción de *Alimentos y bebidas* del estrato del nivel básico). Ello se debe probablemente a que en las sub-muestras en CLM hay un menor número de informantes. Sin embargo, comprobamos que los campos léxicos más compactos en una sub-muestra (con valores mayores del índice) también son los más compactos en la otra, y viceversa.

5.2.5. Incidencias cuantitativas de la variable “tipo de centro”

El tipo de centro es la última variable que vamos a analizar en esta tesis. Como hemos comentado en el apartado 4.4.5, bajo esta variable, vamos a confrontar los resultados obtenidos en las universidades públicas y los obtenidos en la privada. Recordamos que las muestras tanto en ELE como en CLM están constituidas por informantes de dos universidades públicas y una universidad privada. El cotejo entre dichos resultados será interesante, porque las investigaciones de disponibilidad léxica en el mundo hispánico destacan que se trata de una variable que implica enormes diferencias entre sus variantes y, por eso, consideran que el tipo de centro es una variable altamente influyente en el léxico disponible. Pretendemos averiguar, en este apartado, de qué modo el tipo de centro influye en los estudiantes de ELE en Tianjin. Vamos a ver si existe alguna ventaja en la disponibilidad léxica promovida por un tipo de centro sobre el otro y, en su caso, si la ventaja es estadísticamente significativa.

5.2.5.1. Para empezar, en la Tabla 5.53, presentamos, por centro de interés, el promedio de palabras por informante del corpus en ELE según se trate de un centro público o privado. Después de esta tabla, figura la Tabla 5.54 que nos permite ver el rango de los centros de interés en cada uno de los dos

tipos de centro según el promedio de palabras por informante en ELE, de modo que se perciba más claramente el mayor o menor grado del valor de las cifras expuestas en la Tabla 5.53.

Centro de interés	Promedio de palabras en ELE	
	Público	Privado
01 Partes del cuerpo humano	13.05	10.03
02 Ropa y complementos	8.44	9.03
03 Partes de la casa (sin muebles)	8.55	8.00
04 Los muebles de la casa	8.92	8.93
05 Alimentos y bebidas	15.96	13.20
06 Objetos en la mesa para la comida	6.21	5.64
07 La cocina y sus utensilios	6.62	6.61
08 La escuela: muebles y materiales	11.22	10.90
09 Iluminación y calefacción	5.88	4.97
10 La ciudad	13.86	13.25
11 El campo	10.39	9.37
12 Medios de transporte	9.04	8.94
13 Trabajos del campo y del jardín	7.91	7.02
14 Los animales	11.30	8.79
15 Juegos y distracciones	9.29	9.98
16 Profesiones y oficios	11.72	11.25
Global	9.90	9.12

Tabla 5.53. Promedio de palabras por informante por centro de interés en ELE según la variable tipo de centro

Centro de interés	Rango según el promedio de palabras por informante en ELE	
	Público	Privado
01 Partes del cuerpo humano	3	5
02 Ropa y complementos	12	8
03 Partes de la casa (sin muebles)	11	12
04 Los muebles de la casa	10	10
05 Alimentos y bebidas	1	2
06 Objetos en la mesa para la comida	15	15
07 La cocina y sus utensilios	14	14
08 La escuela: muebles y materiales	6	4
09 Iluminación y calefacción	16	16
10 La ciudad	2	1
11 El campo	7	7
12 Medios de transporte	9	9
13 Trabajos del campo y del jardín	13	13
14 Los animales	5	11
15 Juegos y distracciones	8	6
16 Profesiones y oficios	4	3

Tabla 5.54. Rango de los centros de interés por tipo de centro según el promedio de palabras por informante en ELE

Al echar un vistazo a la Tabla 5.53, podemos ver que la mayor producción léxica en ELE de todos los centros de interés y de ambos tipos de centro se detecta en *Alimento y bebidas* de los centros públicos, que ostenta una media de 15.96 palabras. Por el contrario, la menor producción léxica en ELE se halla en *Iluminación y calefacción* del centro privado, con solamente un promedio de 4.97 palabras.

A continuación, vamos a analizar por tipo de centro el promedio de palabras por informante en ELE. Como podemos comprobar en la Tabla 5.54, los centros públicos tienen la mayor producción léxica en ELE en *Alimentos y bebidas*, seguido por *La ciudad*, *Partes del cuerpo humano*, *Profesiones y*

oficios y *Los animales*; mientras que el centro privado registra su mayor promedio de palabras por informante en ELE en *La ciudad*, seguido por *Alimentos y bebidas*, *Profesiones y oficios*, *La escuela: sus muebles y materiales* y *Partes del cuerpo humano*. Como resultado, los dos tipos de centro coinciden en cuatro áreas temáticas con una mayor productividad léxica en ELE, las cuales son *Alimentos y bebidas*, *La ciudad*, *Partes del cuerpo humano* y *Profesiones y oficios*. En cuanto a las diferencias, señalamos que los centros públicos cuentan *Los animales* entre sus campos semánticos más productivos en ELE; y el centro privado, *La escuela: sus muebles y materiales*.

En lo que se refiere a los campos semánticos menos productivos en ELE, la Tabla 5.54 informa de que tanto el centro público como el privado coinciden totalmente en los primeros cuatro campos semánticos con una menor producción léxica, los cuales son *Iluminación y calefacción*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios* y *Trabajos del campo y del jardín*. La diferencia al respecto solo se da en el quinto campo léxico menos productivo. Concretamente, los centros públicos cuentan con *Ropa y complementos* entre sus campos nocionales menos productivos en ELE; y el centro privado, con *Partes de la casa (sin muebles)*.

Si nos fijamos en el rango de los demás centros de interés, nos damos cuenta de que, en realidad, los dos tipos de centro coinciden en siete áreas temáticas (*Los muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *Iluminación y calefacción*, *El campo*, *Medios de transporte* y *Trabajos del campo y del jardín*) en lo referente al mayor o menor grado de productividad léxica en ELE.

Promedio de palabras por centro de interés en ELE

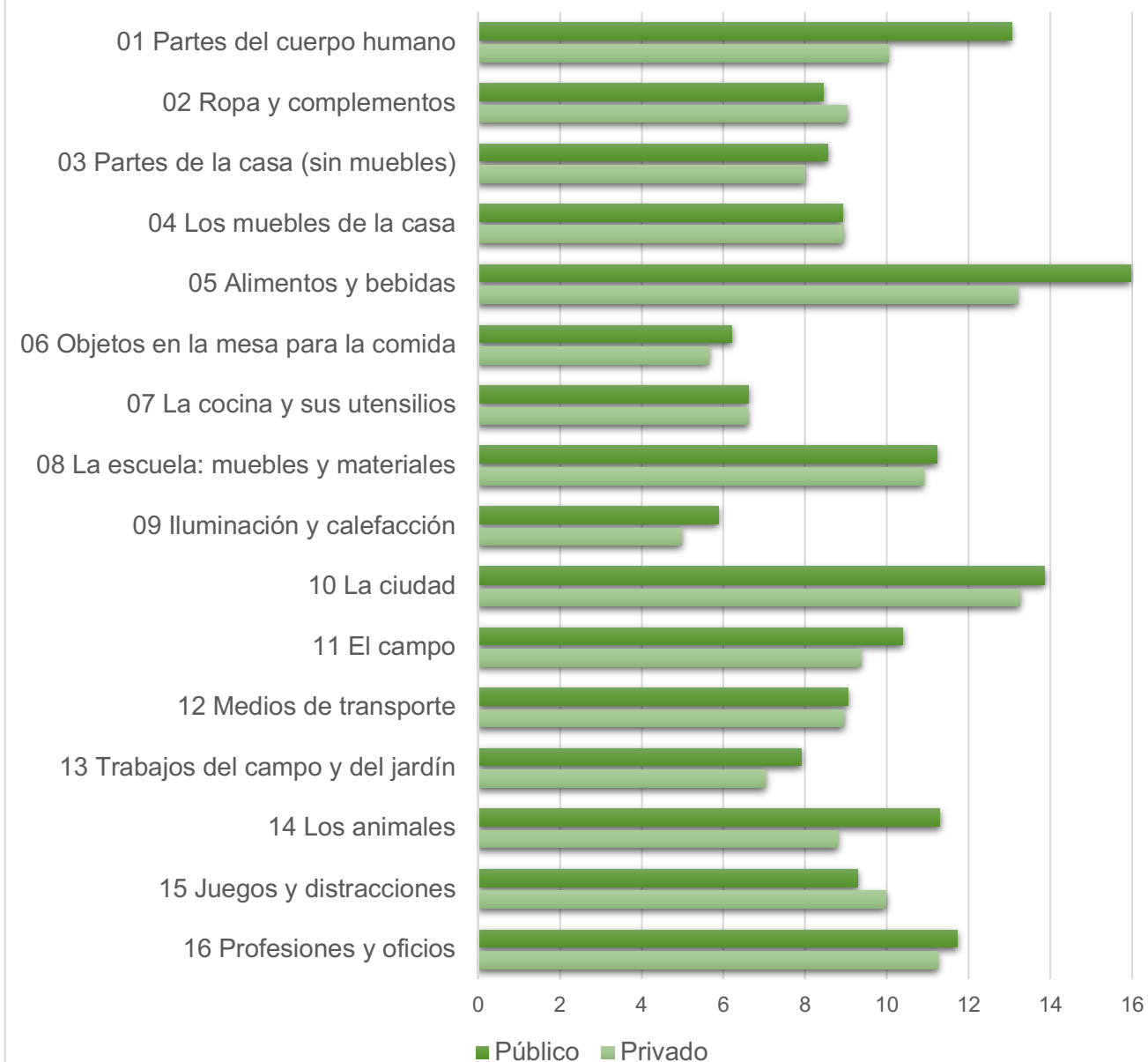


Gráfico 5.45. Promedio de palabras por informante por centro de interés en ELE según la variable tipo de centro

Ahora nos centramos en el Gráfico 5.45, que muestra las diferencias entre los centros públicos y el centro privado con respecto al promedio de palabras por informante en ELE en cada una de las áreas nocionales. Como podemos ver, el estrato de centro público señala mejores resultados de producción léxica en ELE que el estrato de centro privado en trece campos asociativos: *Partes del cuerpo humano*, *Partes de la casa (sin muebles)*, *Alimentos y bebidas*,

Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela: sus muebles y materiales, Iluminación y calefacción, La ciudad, El campo, Medios de transporte, Trabajos del campo y del jardín, Los animales y Profesiones y oficios. Por lo tanto, solo hay tres campos semánticos en los que el estrato de centro privado indica producción léxica en ELE superior al estrato de centro público: *Ropa y complementos, Los muebles de la casa y Juegos y distracciones.* Vemos que, en la inmensa mayoría de las áreas nocionales, los centros públicos muestran mayor productividad léxica en ELE que el centro privado. A continuación, pasamos a analizar si dichas diferencias cuantitativas entre los dos tipos de centro son significativas o, al contrario, muy leves.

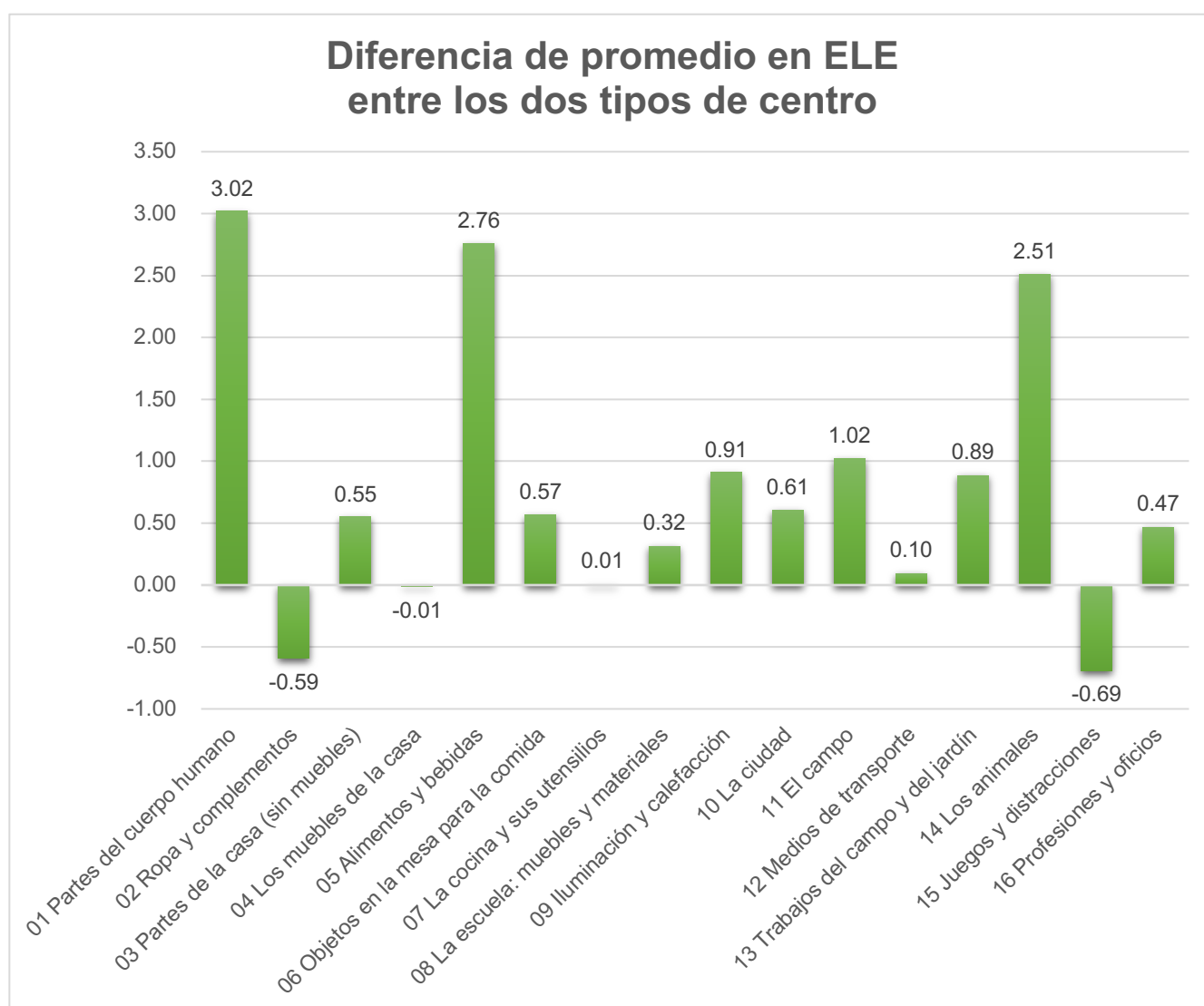


Gráfico 5.46. Diferencia de promedio de palabras en ELE entre los dos tipos de centro por centro de interés

El Gráfico 5.46 exhibe, por centro de interés, las diferencias cuantitativas que se dan entre los centros públicos y el privado con respecto al promedio de palabras por informante en ELE. Se elabora a partir de los datos de la Tabla 5.53 y con respecto a los resultados del estrato de centro público, es decir, las diferencias resultan de restar las cifras del centro privado de las de los públicos. Como podemos ver, la mayor divergencia se halla en *Partes del cuerpo humano* (3.02), donde los estudiantes de los centros públicos aportan 3.02 palabras más que los del centro privado en promedio.

Observamos que solamente existen cuatro áreas nocionales en las que la diferencia cuantitativa entre los dos tipos de centro supera una palabra: *Partes del cuerpo humano* (3.02), *Alimentos y bebidas* (2.76), *Los animales* (2.51) y *El campo* (1.02). Estos cuatro centros de interés indican todos una superioridad de los centros públicos sobre el privado. Como demuestra el gráfico, los centros que señalan una diferencia de entre 0.5 y 1 palabras son los siguientes: *Partes de la casa (sin muebles)* (0.55), *Objetos colocados en la mesa para la comida* (0.57), *Iluminación y calefacción* (0.91), *La ciudad* (0.61) y *Trabajos del campo y del jardín* (0.89), en los que los centros públicos cuentan con mayor producción léxica en ELE que el centro privado; y *Ropa y complementos* (-0.59) y *Juegos y distracciones* (-0.69), en los que el centro privado tiene mejores resultados que los públicos. En las demás áreas temáticas, las diferencias cuantitativas que implican la superioridad de un tipo de centro sobre el otro son bastantes leves, con una diferencia menor de 0.5 palabras.

Dado lo anterior, creemos que dichos resultados según el tipo de centro universitario favorecen al estrato de los centros públicos; es decir, los alumnos de los centros públicos disponen de una ventaja sobre los del centro privado, en relación a la productividad léxica en ELE.

5.2.5.2. De aquí en adelante, exponemos, como referencia de comparación, los resultados sobre el promedio de palabras por informante del corpus en CLM según sea centro público o privado. Del mismo modo como

hemos hecho en los párrafos previos sobre ELE, presentamos abajo la Tabla 5.55, en la que se ve dicho índice en CLM en cada centro de interés por tipo de centro universitario. Al mismo tiempo, en la Tabla 5.56 se puede comprobar el rango de las áreas nocionales por tipo de centro de enseñanza según el promedio de palabras por informante en CLM.

Centro de interés	Promedio de palabras en CLM	
	Público	Privado
01 Partes del cuerpo humano	20.03	16.87
02 Ropa y complementos	17.78	15.96
03 Partes de la casa (sin muebles)	13.61	11.37
04 Los muebles de la casa	15.68	14.17
05 Alimentos y bebidas	21.41	18.63
06 Objetos en la mesa para la comida	11.71	10.43
07 La cocina y sus utensilios	14.80	12.83
08 La escuela: muebles y materiales	17.15	15.57
09 Iluminación y calefacción	12.87	10.50
10 La ciudad	18.11	16.13
11 El campo	17.44	14.00
12 Medios de transporte	14.87	13.02
13 Trabajos del campo y del jardín	13.20	9.67
14 Los animales	22.78	19.46
15 Juegos y distracciones	17.03	13.35
16 Profesiones y oficios	18.94	16.26
Global	16.71	14.26

Tabla 5.55. Promedio de palabras por informante por centro de interés en CLM según la variable tipo de centro

Centro de interés	Rango según el promedio de palabras por informante en CLM	
	Público	Privado
01 Partes del cuerpo humano	3	3
02 Ropa y complementos	6	6
03 Partes de la casa (sin muebles)	13	13
04 Los muebles de la casa	10	8
05 Alimentos y bebidas	2	2
06 Objetos en la mesa para la comida	16	15
07 La cocina y sus utensilios	12	12
08 La escuela: muebles y materiales	8	7
09 Iluminación y calefacción	15	14
10 La ciudad	5	5
11 El campo	7	9
12 Medios de transporte	11	11
13 Trabajos del campo y del jardín	14	16
14 Los animales	1	1
15 Juegos y distracciones	9	10
16 Profesiones y oficios	4	4

Tabla 5.56. Rango de los centros de interés por tipo de centro según el promedio de palabras por informante en CLM

Una ojeada a las cifras expuestas en la Tabla 5.55 nos revela que la mayor producción léxica en CLM de todos los centros de interés de los dos tipos de centro se encuentra en *Los animales* del estrato de los centros públicos —con un promedio de 22.78 palabras—. Por el contrario, la menor producción léxica en CLM se halla en *Trabajos del campo y del jardín* del estrato del centro privado —con una media de 9.67 palabras—. Estos resultados en CLM no contradicen los resultados previamente extraídos en ELE según el tipo de centro universitario, ya que en ambas lenguas los centros públicos arrojan el mayor promedio de palabras por informante y el centro privado, en cambio, el menor promedio.

Ahora, vamos a analizar el promedio de palabras por informante en CLM en cada uno de los tipos de centro de enseñanza. Como podemos ver en la Tabla 5.56, los centros públicos y el centro privado coinciden totalmente en los primeros cinco campos semánticos con una mayor producción léxica en CLM, los cuales son *Los animales*, *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad*. Cabe destacar que, comparando estos resultados con los previamente extraídos en ELE según el tipo de centro universitario, vemos que, en ambos tipos de centro, *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad* son los centros de interés comunes en ELE y en CLM en cuanto a su carácter de alto grado de productividad léxica en el conjunto de los dieciséis campos asociativos.

En lo que se refiere a los campos semánticos menos productivos en CLM, la Tabla 5.56 informa de que los centros públicos tienen la menor producción léxica en CLM en *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *Trabajos del campo y del jardín*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*; mientras que el centro privado muestra su menor promedio de palabras por informante en CLM en *Trabajos del campo y del jardín*, seguido por *Objetos colocados en la mesa para la comida*, *Iluminación y calefacción*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*. De igual modo se observa una coincidencia considerable, puesto que, en ambos tipos de centro universitario, las áreas nocionales menos productivas en CLM son las mismas, aunque no tienen las mismas posiciones en el rango de cada tipo de centro. Comparando estos resultados con los previamente extraídos en ELE según el tipo de centro, sabemos que cuatro de las cinco áreas nocionales menos productivas en CLM, comunes para los centros públicos y el privado, también se hallan entre las menos productivas en ELE: *Objetos colocados en la mesa para la comida*, *Iluminación y calefacción*, *Trabajos del campo y del jardín* y *La cocina y sus utensilios*.

Comprobamos de nuevo —esta vez desde la perspectiva de los dos tipos de centro universitario— cómo las áreas semánticas más y menos productivas en CLM resultan ser también las más y menos productivas en ELE,

respectivamente.

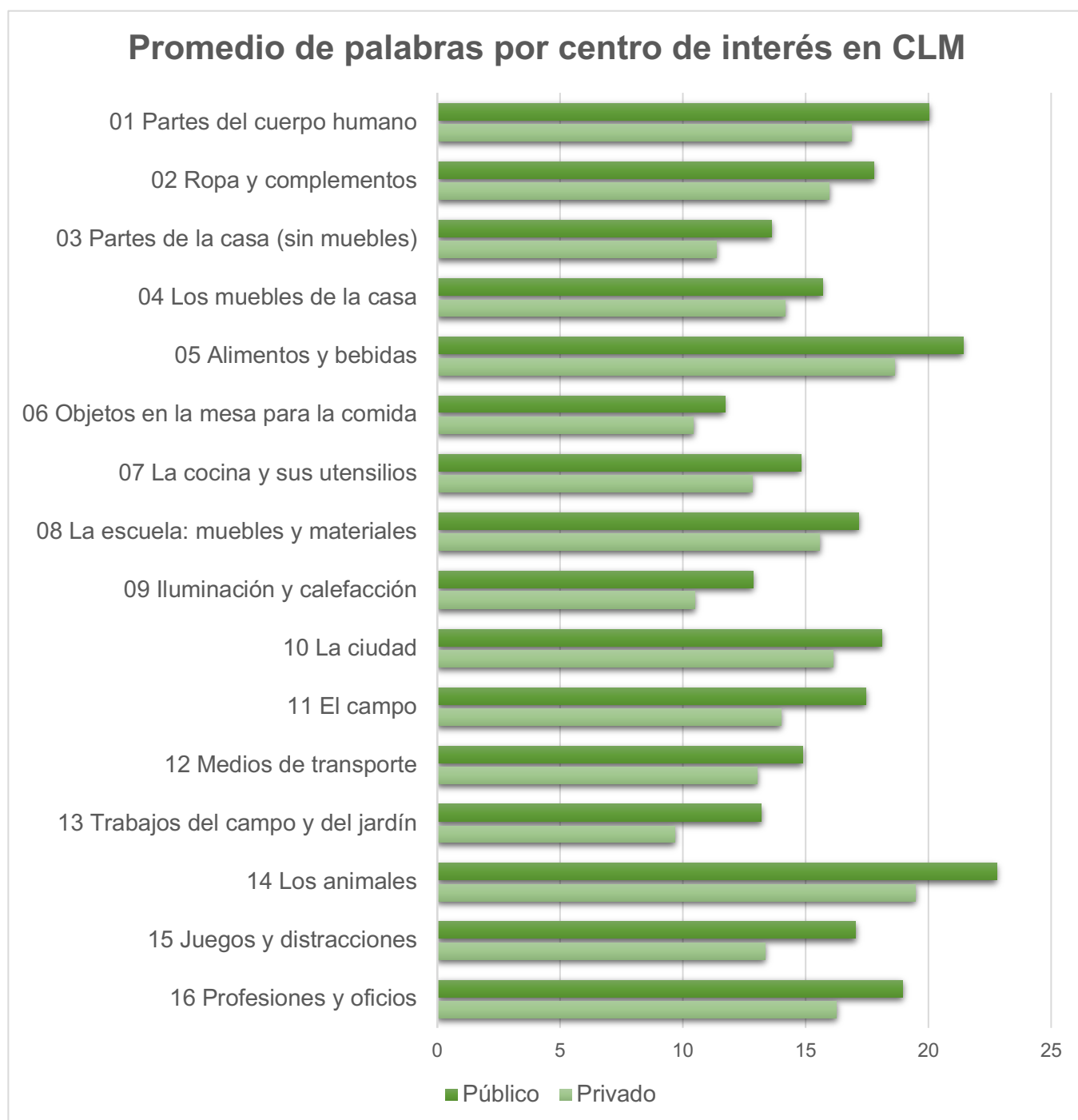


Gráfico 5.47. Promedio de palabras por informante por centro de interés en CLM según la variable tipo de centro

Si nos fijamos en el Gráfico 5.47, que expone las diferencias entre los centros públicos y el privado con respecto a la producción léxica en CLM en cada uno de centros de interés, podemos ver que existe una superioridad

evidente de un tipo de centro al otro. Concretamente, el estrato de los centros públicos señala mejores resultados de producción léxica en CLM que el estrato del centro privado en todos los campos asociativos sin excepción. A continuación, pasamos a analizar si es significativa la superioridad cuantitativa de los centros públicos sobre el privado en cada uno de los dieciséis campos semánticos.

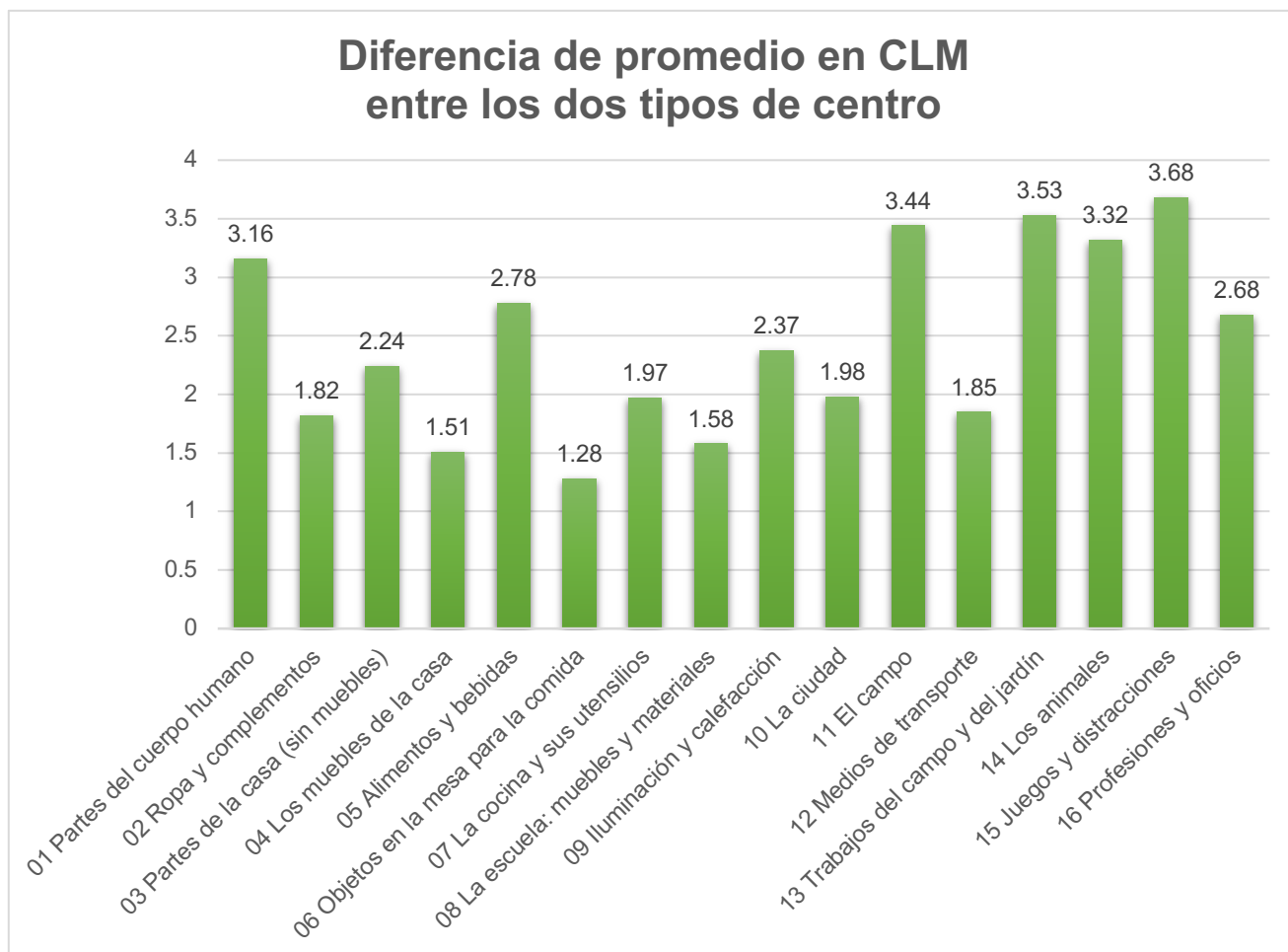


Gráfico 5.48. Diferencia de promedio de palabras en CLM entre los dos tipos de centro por centro de interés

El gráfico 5.48 exhibe, por centro de interés, las diferencias cuantitativas que se dan entre los dos tipos de centro universitario con respecto al promedio de palabras por informante en CLM. Es elaborado a partir de los datos aparecidos en la Tabla 5.55 y con respecto a los resultados de los centros públicos, es decir, las diferencias resultan de restar las cifras del centro privado de las de los públicos. Como podemos ver, la mayor divergencia se encuentra

en *Juegos y distracciones* (3.68), donde los sujetos de los centros públicos aportan 3.68 palabras en CLM más que los del centro privado en promedio.

Observamos que la diferencia cuantitativa entre los dos tipos de centro de enseñanza está por encima de una palabra en todos los centros de interés. Vale la pena señalar que dicha diferencia supera tres palabras en cinco campos léxicos —*Partes del cuerpo humano* (3.16), *El campo* (3.44), *Trabajos del campo y del jardín* (3.53), *Los animales* (3.32) y *Juegos y distracciones* (3.68)—. Así pues, estos cinco son los campos léxicos en los que los alumnos de los centros públicos llevan una ventaja considerable sobre los alumnos del centro privado, con relación a la productividad léxica en CLM. Las áreas nocionales donde dicha diferencia se queda entre dos y tres palabras son cuatro: *Partes de la casa (sin muebles)* (2.24), *Alimentos y bebidas* (2.78), *Iluminación y calefacción* (2.37) y *Profesiones y oficios* (2.68). El resto de centros de interés —siete— presentan una superioridad menor y que supone entre una y dos palabras: *Ropa y complementos* (1.82), *Los muebles de la casa* (1.51), *Objetos colocados en la mesa para la comida* (1.28), *La cocina y sus utensilios* (1.97), *La escuela: sus muebles y materiales* (1.58), *La ciudad* (1.98) y *Medios de transporte* (1.85). Por eso, deducimos que, sistemáticamente, los alumnos de los centros públicos disponen de mayor productividad léxica en CLM que los del centro privado.

A continuación, a partir de los datos globales expuestos en las Tablas 5.53 y 5.55, presentamos en los Gráficos 5.49 y 5.50 el promedio de palabras global de los dos tipos de centro universitario en los corpus de ELE y de CLM.

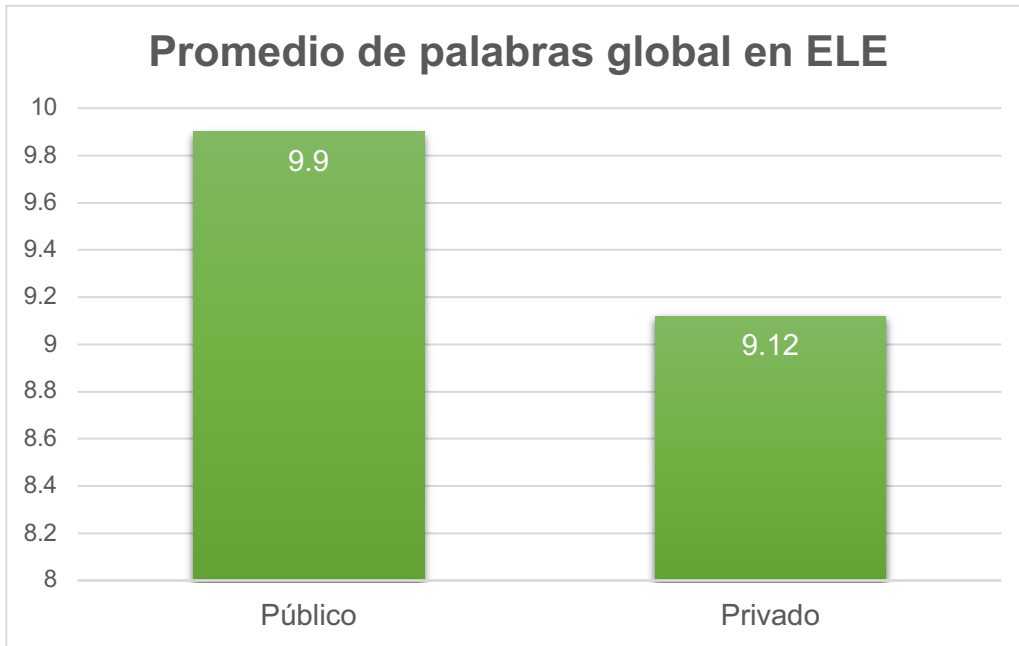


Gráfico 5.49. Promedio global de palabras por informante en ELE según la variable tipo de centro

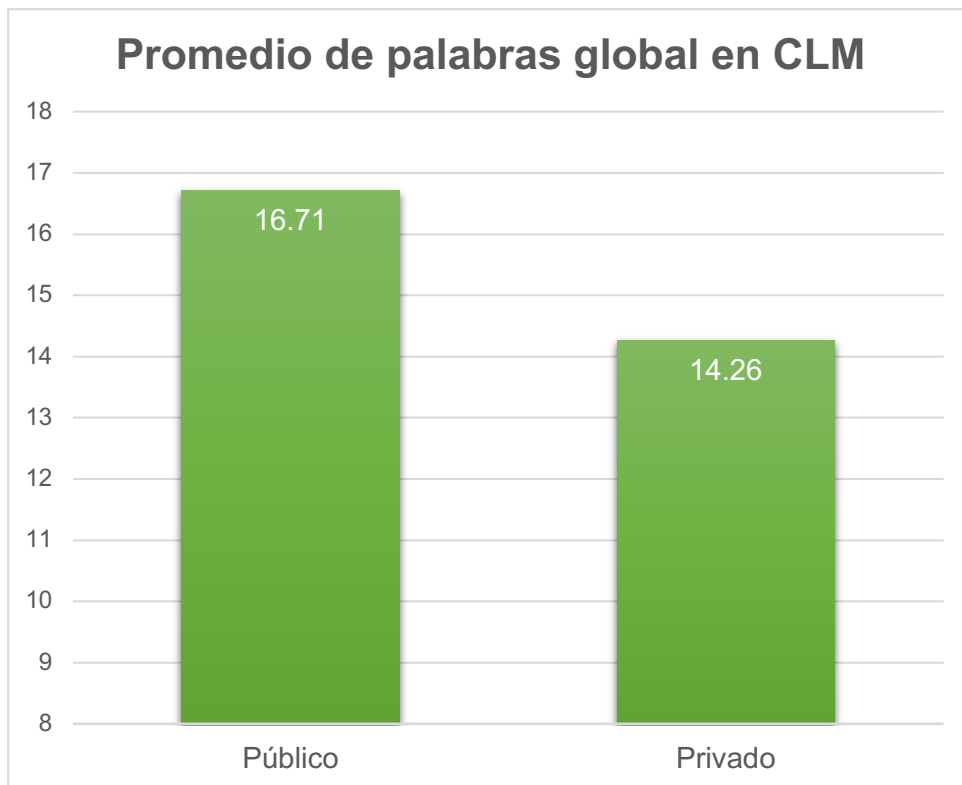


Gráfico 5.50. Promedio global de palabras por informante en CLM según la variable tipo de centro

Como muestran los dos gráficos de arriba, vemos comprobadas las

deducciones anteriormente extraídas por centro de interés sobre la ventaja de productividad léxica de un tipo de centro universitario sobre el otro, pero, esta vez, de manera global e independientemente de los campos asociativos. En el Gráfico 5.49, podemos observar que los informantes de los centros públicos aportan, globalmente, 0.78 palabras en ELE más que los informantes del centro privado (9.9, en los centros públicos, frente a 9.12, en el centro privado). En cambio, la diferencia del promedio de palabras global en CLM entre los dos tipos de centro es mucho mayor que la que se da en ELE. Como podemos ver en el Gráfico 5.50, dicha diferencia en CLM es de 2.45 palabras (16.71, en los centros públicos, frente a 14.26, en el centro privado). Así pues, tal y como hemos comentado en los párrafos anteriores, los resultados de producción léxica según el tipo de centro favorecen al estrato de los centros públicos. En las dos lenguas, los alumnos de los centros públicos disponen de mayor productividad léxica que los alumnos del centro privado.

Por otra parte, comparando los Gráficos 5.49 y 5.50 entre sí, vemos que, tanto en el estrato de los centros públicos como en el del privado, la producción léxica en ELE dista enormemente de la que se muestra en CLM. Además, parece que los estudiantes de cada tipo de centro universitario no revelan una misma divergencia de producción léxica entre ELE y CLM (9.9 frente a 16.71, en los centros públicos, y 9.12 frente a 14.26, en el centro privado). De modo que vamos a examinar en el siguiente apartado, por tipo de centro universitario, la distancia de producción léxica entre las dos lenguas.

5.2.5.3. La Tabla 5.57 reúne los datos de las Tablas 5.53 y 5.55, presentando, a su vez, los resultados completos sobre el promedio de palabras por informante por tipo de centro de enseñanza y en las dos lenguas de nuestra investigación. A partir de ella, obtenemos la Tabla 5.58, que muestra, por centro de interés y por tipo de centro universitario, las diferencias cuantitativas entre el promedio en ELE y el promedio en CLM.

Centro de interés	Promedio de palabras			
	Público		Privado	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	13.05	20.03	10.03	16.87
02 Ropa y complementos	8.44	17.78	9.03	15.96
03 Partes de la casa (sin muebles)	8.55	13.61	8.00	11.37
04 Los muebles de la casa	8.92	15.68	8.93	14.17
05 Alimentos y bebidas	15.96	21.41	13.20	18.63
06 Objetos en la mesa para la comida	6.21	11.71	5.64	10.43
07 La cocina y sus utensilios	6.62	14.80	6.61	12.83
08 La escuela: muebles y materiales	11.22	17.15	10.90	15.57
09 Iluminación y calefacción	5.88	12.87	4.97	10.50
10 La ciudad	13.86	18.11	13.25	16.13
11 El campo	10.39	17.44	9.37	14.00
12 Medios de transporte	9.04	14.87	8.94	13.02
13 Trabajos del campo y del jardín	7.91	13.20	7.02	9.67
14 Los animales	11.30	22.78	8.79	19.46
15 Juegos y distracciones	9.29	17.03	9.98	13.35
16 Profesiones y oficios	11.72	18.94	11.25	16.26
Global	9.90	16.71	9.12	14.26

Tabla 5.57. Promedio de palabras por informante por centro de interés en ELE y en CLM según la variable tipo de centro

Centro de interés	Diferencia de promedio de palabras entre ELE y CLM	
	Público	Privado
01 Partes del cuerpo humano	6.98	6.84
02 Ropa y complementos	9.34	6.93
03 Partes de la casa (sin muebles)	5.06	3.37
04 Los muebles de la casa	6.76	5.24
05 Alimentos y bebidas	5.45	5.43
06 Objetos en la mesa para la comida	5.50	4.79
07 La cocina y sus utensilios	8.18	6.22
08 La escuela: muebles y materiales	5.93	4.67
09 Iluminación y calefacción	6.99	5.53
10 La ciudad	4.25	2.88
11 El campo	7.05	4.63
12 Medios de transporte	5.83	4.08
13 Trabajos del campo y del jardín	5.29	2.65
14 Los animales	11.48	10.67
15 Juegos y distracciones	7.74	3.37
16 Profesiones y oficios	7.22	5.01
Global	6.81	5.14

Tabla 5.58. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable tipo de centro

Otra vez vemos que el promedio de palabras en ELE es inferior al obtenido en CLM en todos los campos léxicos de ambos tipos de centro universitario. La Tabla 5.58 informa de que la mayor diferencia entre los promedios de ELE y de CLM se halla en *Los animales* en los centros públicos (11.48). Cabe destacar que la segunda mayor diferencia también se registra en este centro de interés, en el centro privado (10.67).

Las celdas marcadas en fondo gris en la Tabla 5.58 indican aquellas áreas nocionales en las que la diferencia cuantitativa está por encima del promedio global o, en otras palabras, aquellas áreas en las que la producción léxica en

ELE de los informantes dista enormemente de la que se da en CLM. Volvemos a recordar que hay que prestar atención a estas áreas semánticas, toda vez que la inferioridad en ELE en ellas puede compensarse en el aula a través de adquirir los homólogos en español de las palabras que los estudiantes son capaces de evocar en CLM.

Existen cinco centros de interés en los que tanto los estudiantes de los centros públicos como los del privado señalan gran divergencia de producción léxica entre las dos lenguas: *Partes del cuerpo humano*, *Ropa y complementos*, *La cocina y sus utensilios*, *Iluminación y calefacción* y *Los animales*. Esto quiere decir que los estudiantes de ambos tipos de centro universitario muestran una considerable carencia léxica en estas cinco áreas temáticas en ELE, tomando como referencia su producción de palabras en CLM. A este respecto, también señalamos que solo los centros públicos tienen mucha escasez de vocabulario en ELE en *El campo*, *Juegos y distracciones* y *Profesiones y oficios*. Únicamente en el centro privado, sucede con *Los muebles de la casa* y *Alimentos y bebidas* en los que la deficiencia de la producción en ELE se ve considerable con respecto a la producción en CLM.

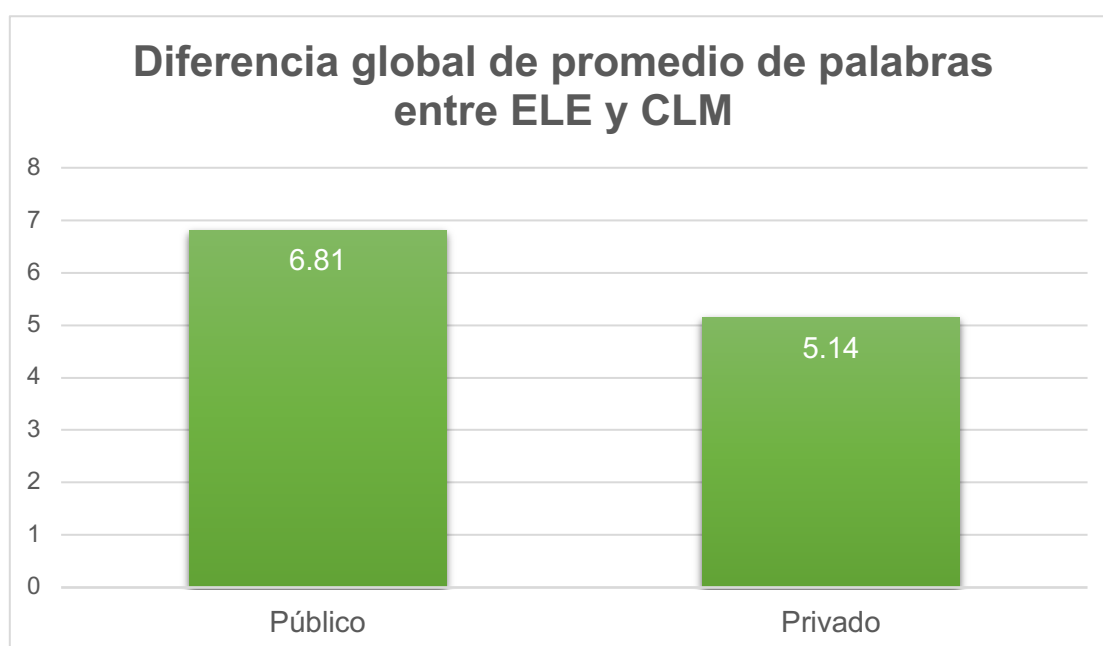


Gráfico 5.51. Diferencia global de promedio de palabras entre ELE y CLM según la variable tipo de centro

Si nos fijamos en la diferencia global de promedio de palabras entre ELE y CLM por tipo de centro universitario —la que se expone en la última fila de la Tabla 5.58 y, de manera ilustrativa, en el Gráfico 5.51—, podemos saber, en ambos tipos de centro de enseñanza, cómo la producción léxica en una lengua dista globalmente de la producción en la otra. Tal y como muestra el Gráfico 5.51, los informantes de los centros públicos aportan 6.81 palabras menos en ELE que en CLM en un campo semántico medio; mientras que los informantes del centro privado escriben 5.14 palabras menos en ELE que en CLM.

Al parecer, los informantes del centro privado muestran menor distancia de producción léxica entre las dos lenguas. No obstante, no podemos llegar a la conclusión de que ellos tienen una ventaja en ELE, aun tomando como referencia su productividad léxica en CLM. Pensamos que la menor distancia entre las dos lenguas del estrato del centro privado se debe a su productividad léxica demasiado baja en CLM, tal como hemos comentado en torno al Gráfico 5.48. Así pues, teniendo en cuenta los resultados excepcionales del centro privado en cuanto a la producción léxica en CLM, creemos inadecuado afirmar que los informantes del centro privado muestran una ventaja en ELE tomando como referencia su productividad léxica en CLM.

A continuación, vamos a ver si el distinto grado de distancia global en los dos tipos de centro universitario también se percibe por área nocional. Para ello, observaremos el Gráfico 5.52, que resulta ser una visualización de los datos presentados en la Tabla 5.58.

Diferencia de promedio de palabras entre ELE y CLM

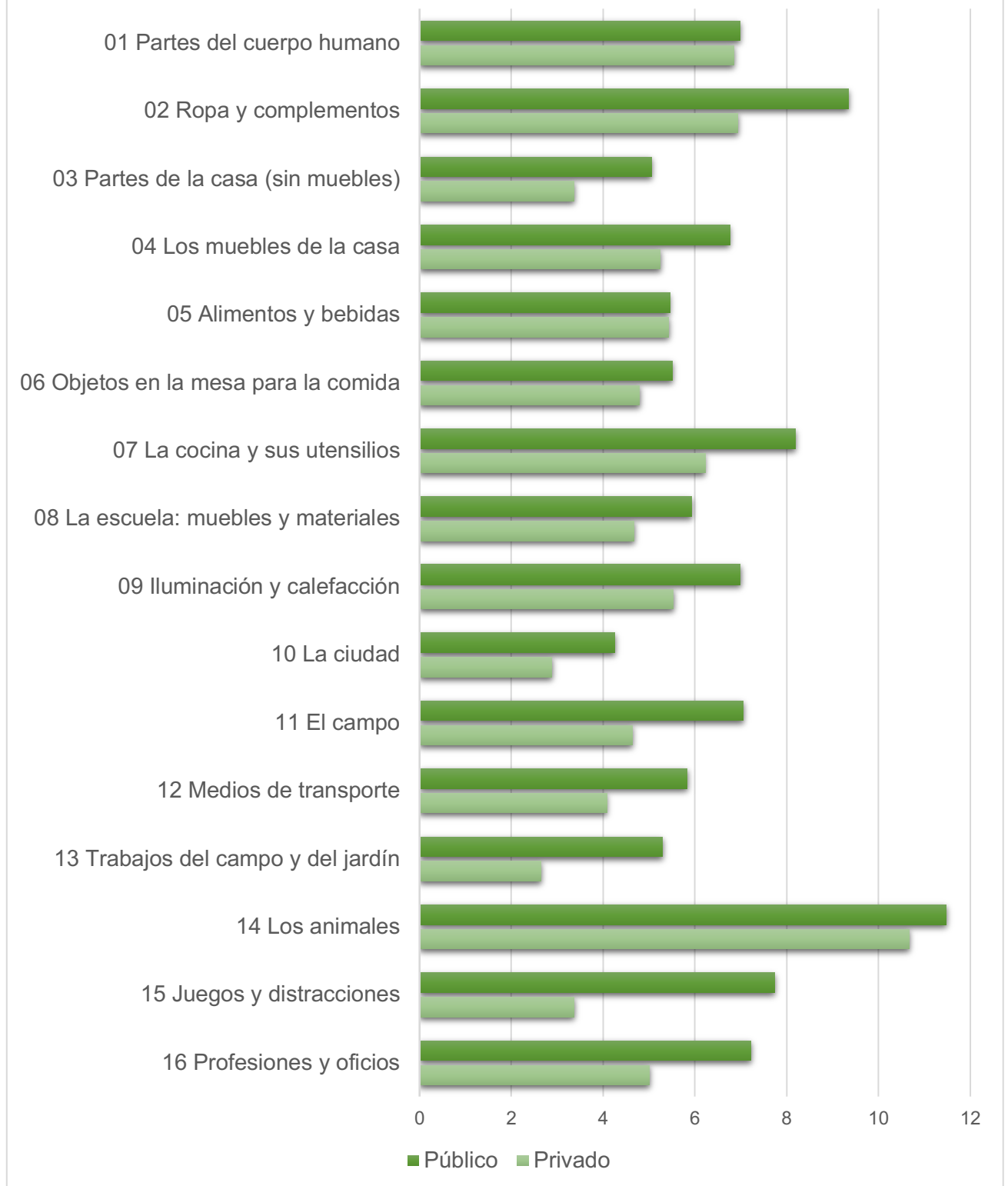


Gráfico 5.52. Diferencia de promedio de palabras entre ELE y CLM por centro de interés según la variable tipo de centro

A partir del Gráfico 5.52, nos damos cuenta de que los alumnos de los centros públicos muestran una mayor distancia de producción léxica entre ELE y CLM, en beneficio de la segunda, que los alumnos del centro privado en todos los campos semánticos; lo cual coincide con la observación global que hemos tenido del Gráfico 5.51. Al igual que lo considerado en los párrafos anteriores, suponemos que la menor distancia entre las dos lenguas que manifiesta el estrato del centro privado no se traduce en que dicho estrato cuente con una ventaja de producción léxica en ELE. En cambio, debido a que los estudiantes del centro privado tienen una productividad léxica demasiado baja en CLM, muestran una ventaja relativa en ELE al tomarse como referencia su productividad léxica en CLM. Es decir, que el hecho de que tengan una producción léxica más cercana en CLM y ELE no significa que los estudiantes del centro privado hayan generado mucho vocabulario en español, sino que han actualizado muchos menos términos en chino de lo esperable.

A continuación, pasamos a analizar si las diferencias cuantitativas entre ELE y CLM que arrojan los dos tipos de centro son afines o dispares.

Afinidad o disparidad entre los dos tipos de centro en sus diferencias de promedio entre ELE y CLM

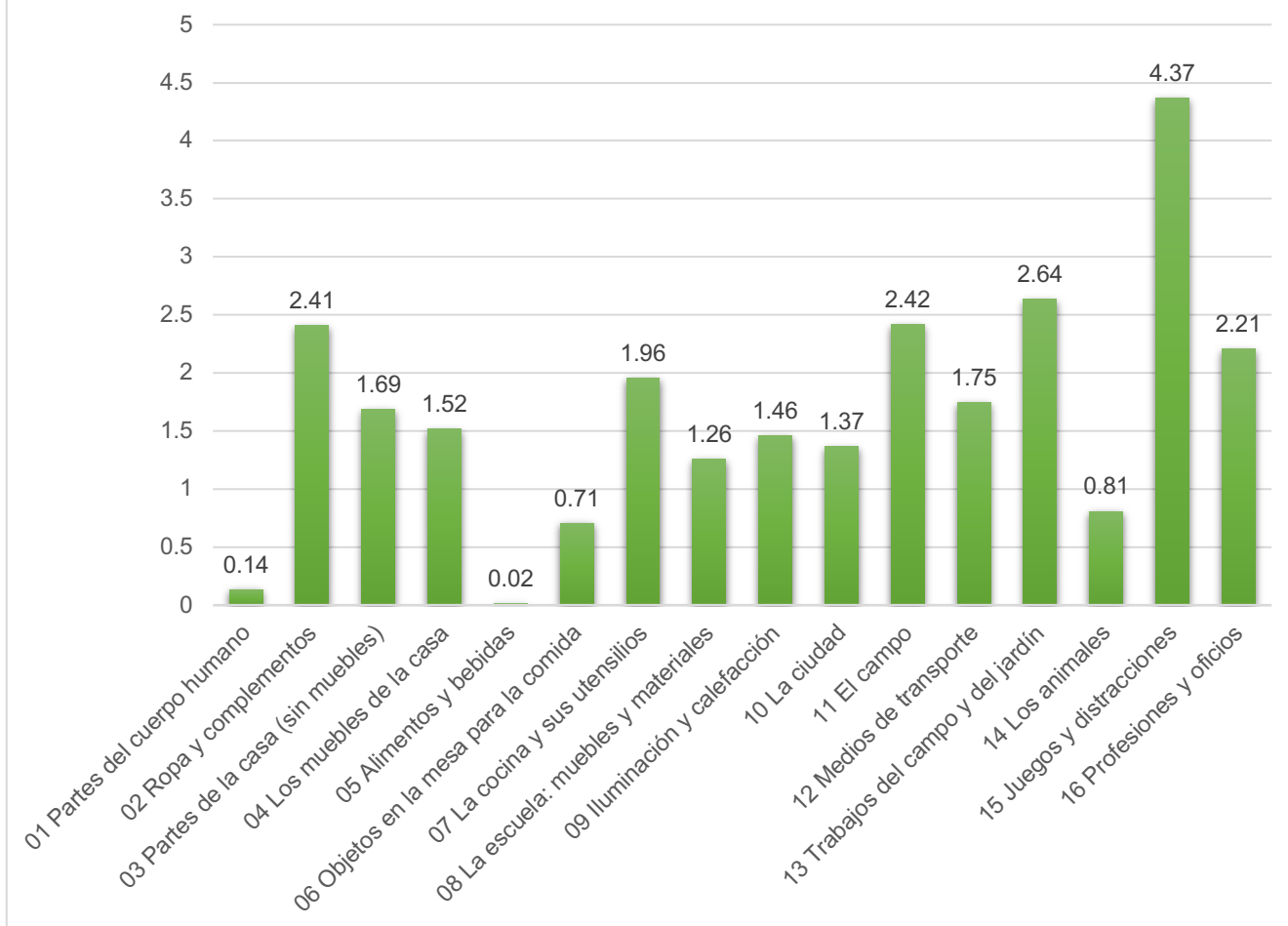


Gráfico 5.53. Afinidad o disparidad entre los dos tipos de centro en sus diferencias de promedio de palabras entre ELE y CLM

El Gráfico 5.53 manifiesta la diferencia entre la barra de los centros públicos y la barra del centro privado en cada centro de interés del Gráfico 5.52. En otras palabras, los valores en el Gráfico 5.53 resultan ser la resta (con respecto a las cifras de los centros públicos) de los datos de cada centro de interés expuestos en la Tabla 5.58. Así pues, nos informa de si son afines o dispares, cuantitativamente, las diferencias de producción léxica entre ELE y CLM en los dos tipos de centro universitario. Tal como presenta el Gráfico 5.53, el valor 0 del eje de ordenadas indica que es cuantitativamente igual dicha diferencia entre los dos tipos de centro universitario en un área nocional, y, por eso, cuanto más se aleje la barra del eje de las abscisas —o lo que es lo mismo, cuanto mayor sea el valor absoluto de la barra—, mayor disparidad se da.

Como podemos ver, la mayor afinidad se da en *Alimentos y bebidas* (0.02), donde el estrato de los centros públicos muestra 0.02 palabras más que el estrato del centro privado en cuanto a la diferencia de promedio de palabras entre las dos lenguas. Esto quiere decir que la distancia de producción léxica entre ELE y CLM en dicha área semántica de los alumnos de los centros públicos supone solamente 0.02 palabras más que la distancia de los alumnos del centro privado. Por lo tanto, los dos estratos tienen en *Alimentos y bebidas* casi la misma diferencia cuantitativa de productividad léxica entre las dos lenguas; o sea la productividad léxica en ELE de ambos tipos de centro de enseñanza en esa área semántica dista de igual modo de su productividad en CLM. Se puede decir que, en *Alimentos y bebidas*, la variable del tipo de centro tiene una incidencia casi nula con relación a la diferencia de productividad léxica entre las dos lenguas.

Siguiendo esa lógica, comprobamos que *Partes del cuerpo humano* (0.14), *Objetos colocados en la mesa para la comida* (0.71) y *Los animales* (0.81) son campos asociativos en los que el valor absoluto de la afinidad resulta menor de 1, aparte de *Alimentos y bebidas* (0.02). En consecuencia, deducimos que, en esos cuatro campos asociativos, la variable del tipo de centro tiene una incidencia poco considerable o muy pequeña con relación a la diferencia de productividad léxica entre las dos lenguas. Por otra parte, *La escuela: sus muebles y materiales* (1.26), *La ciudad* (1.37), *Iluminación y calefacción* (1.46), *Los muebles de la casa* (1.52), *Partes de la casa (sin muebles)* (1.69), *Medios de transporte* (1.75) y *La cocina y sus utensilios* (1.96) son áreas nocionales en las que el valor absoluto de la afinidad está entre 1 y 2, en este orden. Por eso, en esas siete áreas nocionales, el tipo de centro tiene una incidencia pequeña, pero considerable al respecto.

El valor absoluto de la afinidad supera 2, en este orden, en *Profesiones y oficios* (2.21), *Ropa y complementos* (2.41), *El campo* (2.42), *Trabajos del campo y del jardín* (2.64) y *Juegos y distracciones* (4.37). De modo que el tipo de centro tiene una incidencia evidente en esos cinco campos semánticos en

cuanto a la diferencia de producción léxica entre ELE y CLM. Cabe destacar que, en *Juegos y distracciones*, dicho valor llega a 4.37, por lo que es donde los centros públicos se distancian más del privado en la diferencia de productividad léxica entre ELE y CLM, o sea, es donde la productividad léxica en ELE dista de la productividad en CLM en los centros públicos en un grado mayor que como dista en el centro privado.

5.2.5.4. A continuación, vamos a examinar, por tipo de centro universitario y según la lengua de la encuesta, el total de vocablos que se recaban en cada uno de los centros de interés. Recordamos que, ante submuestras que implican entre sí unas diferencias considerables que se dan en el número de informantes, el total de vocablos no resulta un índice del todo fiable. En este caso del tipo de centro universitario, los integrantes del centro privado solo ocupan más o menos un 30% de la muestra de ELE y de CLM, frente a la mayoría —alrededor de un 70%— que suponen los sujetos de los centros públicos (véase el apartado 4.4.5). Podemos ver cómo la diferencia enorme en el número de informantes repercute en los resultados de los vocablos según el tipo de centro.

Centro de interés	Total de vocablos			
	Público		Privado	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	239	226	137	119
02 Ropa y complementos	292	308	125	187
03 Partes de la casa (sin muebles)	210	213	107	105
04 Los muebles de la casa	187	250	107	134
05 Alimentos y bebidas	355	600	191	306
06 Objetos en la mesa para la comida	147	173	74	87
07 La cocina y sus utensilios	213	226	129	129
08 La escuela: muebles y materiales	308	286	166	168
09 Iluminación y calefacción	218	228	92	107
10 La ciudad	595	670	336	361
11 El campo	464	607	296	311
12 Medios de transporte	281	278	153	129
13 Trabajos del campo y del jardín	404	361	210	179
14 Los animales	263	299	127	161
15 Juegos y distracciones	342	361	206	184
16 Profesiones y oficios	378	381	183	199
Global	4896	5467	2639	2866

Tabla 5.59. Total de vocablos por centro de interés en ELE y en CLM según la variable tipo de centro

En la Tabla 5.59 se exhibe el número total de vocablos de cada uno de los dieciséis campos asociativos según el tipo de centro universitario y según pertenezca al corpus de ELE o al de CLM. A partir de ahí, podemos darnos cuenta de la variedad léxica de los estudiantes de ambos tipos de centro de enseñanza en cada una de las dos lenguas contempladas. En primer lugar, los resultados globales que se encuentran en la última fila de la tabla ya evidencian cómo el mayor número de informantes en el estrato de los centros públicos se relaciona con una mayor riqueza léxica, tanto en ELE (4896, en los centros públicos, frente a 2639, en el privado) como en CLM (5467, en los centros públicos, frente a 2866, en el privado). La ventaja de los centros públicos sobre

el privado en ambas lenguas es completamente comprensible en el sentido de que, cuantos más informantes haya, más probablemente encontremos a informantes atípicos que contribuyan al aumento de la variedad del vocabulario.

Por otra parte, si comparamos, en cada tipo de centro universitario, la riqueza léxica global en ELE y en CLM, podemos ver que en los centros públicos la riqueza léxica en ELE (4896) es menor que la que hay en CLM (5467); y en el centro privado sucede más o menos lo mismo, es decir la riqueza léxica en ELE (2639) es un poco menor que la que hay en CLM (2866). Así, deducimos que, en ambos tipos de centro, los alumnos tienen en ELE una riqueza léxica inferior a la que tienen en CLM.

Si pasamos a mirar los resultados de las distintas áreas nocionales, vemos que no contradicen las observaciones desde la perspectiva de los datos globales. Por un lado, en todos los centros de interés, los sujetos de los centros públicos aventajan a los del centro privado en riqueza léxica tanto en ELE como en CLM, lo cual se debe principalmente al mayor número de integrantes de los primeros. Por otro lado, al comparar las dos lenguas, vemos que los estudiantes de los centros públicos disponen de una menor riqueza léxica en ELE que en CLM en la inmensa mayoría de los centros de interés, salvo *Partes del cuerpo humano*, *La escuela: sus muebles y materiales*, *Medios de transporte* y *Trabajos del campo y del jardín*. De igual modo, los estudiantes del centro privado tienen una menor o igual riqueza léxica en ELE que en CLM en la gran mayoría de los centros de interés, salvo en *Partes del cuerpo humano*, *Partes de la casa (sin muebles)*, *Medios de transporte*, *Trabajos del campo y del jardín* y *Juegos y distracciones*. Comprobamos que, en ambos tipos de centro universitario, los alumnos tienen en ELE una riqueza léxica inferior a la que tienen en CLM.

Centro de interés	Rango según el total de vocablos			
	Público		Privado	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	11	13	9	13
02 Ropa y complementos	8	7	12	5
03 Partes de la casa (sin muebles)	14	15	13	15
04 Los muebles de la casa	15	11	13	10
05 Alimentos y bebidas	5	3	5	3
06 Objetos en la mesa para la comida	16	16	16	16
07 La cocina y sus utensilios	13	13	10	11
08 La escuela: muebles y materiales	7	9	7	8
09 Iluminación y calefacción	12	12	15	14
10 La ciudad	1	1	1	1
11 El campo	2	2	2	2
12 Medios de transporte	9	10	8	11
13 Trabajos del campo y del jardín	3	5	3	7
14 Los animales	10	8	11	9
15 Juegos y distracciones	6	5	4	6
16 Profesiones y oficios	4	4	6	4

Tabla 5.60. Rango de los centros de interés por tipo de centro según el total de vocablos en ELE y en CLM

La Tabla 5.60 nos informa del rango de los centros de interés por tipo de centro universitario según la riqueza léxica en las dos lenguas respectivas. Un vistazo a las cifras en la tabla nos advierte de que, unánimemente en ambos tipos de centro de enseñanza y en ambas lenguas, *La ciudad* es el área semántica más rica en vocabulario; y *El campo* es el segundo centro de interés más rico. Del mismo modo, *Objetos colocados en la mesa para la comida* es el campo asociativo menos rico en vocabulario.

Aparte de dichas áreas semánticas, la Tabla 5.60 señala que los centros públicos muestran una mayor riqueza léxica en ELE también en *Trabajos del campo y del jardín*, *Profesiones y oficios* y *Alimentos y bebidas*; mientras que

el centro privado lo hace en *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Alimentos y bebidas*. Así vemos que los dos tipos de centro universitario coinciden en cuatro centros más ricos en vocabulario en ELE: *La ciudad*, *El campo*, *Trabajos del campo y del jardín* y *Alimentos y bebidas*. Como diferencia, los centros públicos también tienen un alto grado de variedad léxica en ELE en *Profesiones y oficios*; mientras que el centro privado, en *Juegos y distracciones*.

Por el contrario, además de *Objetos colocados en la mesa para la comida*, los centros públicos también tienen una menor variedad léxica en ELE en *Los muebles de la casa*, *Partes de la casa (sin muebles)*, *La cocina y sus utensilios* e *Iluminación y calefacción*; mientras que el centro privado la tiene en *Iluminación y calefacción*, *Los muebles de la casa*, *Partes de la casa (sin muebles)* y *Ropa y complementos*. De modo que los informantes de ambos tipos de centro universitario coinciden en cuatro centros menos ricos en vocabulario en ELE: *Objetos colocados en la mesa para la comida*, *Los muebles de la casa*, *Partes de la casa (sin muebles)* e *Iluminación y calefacción*. Como diferencia, los centros públicos señalan una menor riqueza léxica en ELE en *La cocina y sus utensilios*; mientras que el centro privado lo hace en *Ropa y complementos*.

Centro de interés	Índice de cohesión			
	Público		Privado	
	ELE	CLM	ELE	CLM
01 Partes del cuerpo humano	0.05	0.09	0.07	0.14
02 Ropa y complementos	0.03	0.06	0.07	0.09
03 Partes de la casa (sin muebles)	0.04	0.06	0.07	0.11
04 Los muebles de la casa	0.05	0.06	0.08	0.11
05 Alimentos y bebidas	0.04	0.04	0.07	0.06
06 Objetos en la mesa para la comida	0.04	0.07	0.08	0.12
07 La cocina y sus utensilios	0.03	0.07	0.05	0.10
08 La escuela: muebles y materiales	0.04	0.06	0.07	0.09
09 Iluminación y calefacción	0.03	0.06	0.05	0.10
10 La ciudad	0.02	0.03	0.04	0.04
11 El campo	0.02	0.03	0.03	0.05
12 Medios de transporte	0.03	0.05	0.06	0.10
13 Trabajos del campo y del jardín	0.02	0.04	0.03	0.05
14 Los animales	0.04	0.08	0.07	0.12
15 Juegos y distracciones	0.03	0.05	0.05	0.07
16 Profesiones y oficios	0.03	0.05	0.06	0.08

Tabla 5.61. Índice de cohesión por centro de interés en ELE y en CLM según la variable tipo de centro

Por último, presentamos el índice de cohesión, que relaciona los dos índices que acabamos de observar —el promedio de palabras por informante y el total de vocablos—. La Tabla 5.61 expone el índice de cohesión de cada centro de interés por tipo de centro universitario y según cuál sea la lengua de la encuesta. Volvemos a ver cómo la diferencia en el número de informantes afecta los resultados de dicho índice en cada sub-muestra.

Por un lado, según cuál sea el tipo de centro, en ambos corpus de ELE y de CLM, se observan unos índices de cohesión mayores en el centro privado que en los públicos. Esto se debe a que, en el conjunto de cada corpus por separado, hay menos estudiantes de los primeros que de los segundos (véase

el apartado 4.4.5) y es más probable que los alumnos del centro privado coincidan en sus respuestas, lo que conduce al aumento del valor del índice de cohesión. Por otro lado, según cuál sea la lengua de la encuesta, podemos ver que, tanto en la sub-muestra de los centros públicos como en la del privado, el índice de cohesión de cualquier área nocional en ELE es menor o igual que el de la misma área en CLM, salvo en *Alimentos y bebidas* del centro privado. Creemos que, de igual modo, es por causa de que en las sub-muestras en CLM se halla un menor número de informantes. De todos modos, sí comprobamos que los campos léxicos más compactos en una sub-muestra (con valores mayores del índice) también son los más compactos en la otra, y viceversa.

6. ANÁLISIS CUALITATIVO DE LOS DATOS

Este capítulo se ocupa de los resultados léxicos obtenidos desde una perspectiva cualitativa. Concretamente, vamos a hacer un análisis cualitativo de los vocablos disponibles que se han recaudado en el conjunto de las encuestas. Como señala Gómez Devís (2004: 181),

“los materiales nos permiten observar distintos aspectos del comportamiento léxico: 1) las deficiencias lingüísticas tanto léxicas como ortográficas; 2) las transferencias léxicas: cantidad y procedencia de extranjerismos, adaptación lingüística como préstamos, etc.; 3) los procesos de lexificación: composición, derivación, sinonimia, antonimia; 4) los léxicos específicos, dialectalismos y palabras patrimoniales; 5) el léxico como diagnóstico del desarrollo etnolingüístico: influencia del entorno, evolución cultural; 6) el estudio de las categorías gramaticales así como de los grupos léxicos; y 7) el grado de convergencia conceptual en los repertorios léxicos de los diferentes subgrupos”.

En lo que respecta a este estudio del español como lengua extranjera (ELE) y del chino como lengua materna (CLM) de los mismos informantes, nos interesan, sobre todo, dos de dichos aspectos: 1) la convergencia y divergencia conceptuales entre las listas de palabras de cada centro de interés según cada una de las dos lenguas contempladas y 2) la deficiencia lingüística en ELE de los estudiantes, tomando como referencia su conocimiento léxico en CLM. Para ello, primeramente, vamos a comparar los vocablos disponibles de los alumnos en ELE y en CLM bajo cada una de las dieciséis áreas nocionales consideradas. Después, buscaremos los homólogos —o, lo que es lo mismo, los sinónimos conceptuales— de los vocablos de una lengua en la otra. Así, podremos obtener dos tipos de observación cualitativa. El primero consiste en cuáles son los vocablos que los estudiantes son capaces de actualizar en ELE bajo una circunstancia comunicativa determinada, aun cuando los conceptos de dichos vocablos no se les ocurran en la misma circunstancia en CLM. Estos vocablos divergentes que solo aparecen en las listas de ELE y que no son

aportados en CLM pueden dar a conocer el fruto del aprendizaje en el aula del vocabulario español.

El segundo tipo de observación cualitativa —que es el más importante— es cuáles son los vocablos que los estudiantes pueden evocar en CLM cuyos conceptos resultan ausentes en las listas de palabras en ELE. La ausencia de estos vocablos en el repertorio en ELE puede deberse, generalmente, a una deficiencia en el aprendizaje. Por eso, el averiguar cuáles son esos vocablos tiene un sentido didáctico práctico, puesto que, en el capítulo anterior, hemos hablado mucho de las deficiencias cuantitativas que muestran los alumnos en ELE con respecto a su competencia léxica en CLM, mientras tal defecto puede compensarse en el aula, adquiriendo los homólogos en español de las palabras que los estudiantes son capaces de actualizar solamente en CLM.

Existen varios criterios de corte en lo que se refiere a qué vocablos de los repertorios léxicos se deben someter al análisis cualitativo. Tal como hemos comentado en el apartado 5.1.1.3, en algunas investigaciones se consideran el 75% de la frecuencia acumulada (Samper Padilla, 1999) o el índice de disponibilidad igual a o mayor que 0.02 (Bartol Hernández, 2001). Por otra parte, otros trabajos que analizan de manera pormenorizada las palabras disponibles también plantean números de vocablos concretos como medida de corte. Por ejemplo, en Gómez Devís (2004) y Martínez Olmos (2008) se examinan los primeros cincuenta vocablos más disponibles. En cuanto al ámbito de disponibilidad léxica en ELE, Carcedo González (2000) revisa los primeros veinte vocablos más disponibles de los preuniversitarios fineses, pero se limita a comparar los primeros diez vocablos más disponibles de los cuatro diferentes estratos según el nivel de estudios. Samper Hernández (2002) coteja los primeros diez vocablos más disponibles de los alumnos extranjeros de ELE según su lengua materna. De modo similar, Sifrar Kalan (2009) compara los primeros diez vocablos más disponibles en ELE de Eslovenia, Salamanca y Finlandia, mientras tanto la misma autora (2012) opta por examinar los veinte vocablos más disponibles de los alumnos de ELE en Eslovenia para compararlos con los de otras sintopías de español como lengua

materna.

Dado el análisis cuantitativo que hemos hecho en el apartado 5.1.1.3 de esta tesis, nuestra consideración al respecto resulta analizar cualitativamente los vocablos que alcanzan un índice de disponibilidad igual o mayor que 0.1, mientras las listas completas de disponibilidad léxica se pueden consultar en los apéndices que figuran al final de la tesis. A nuestro juicio, los vocablos con índice de disponibilidad superior a 0.1 constituyen las palabras altamente disponibles de nuestros informantes y, además, implica una cantidad de palabras justa (véase la Tabla 5.12) que permite llevar a cabo actividades docentes prácticas para enseñar el vocabulario en el aula. Así pues, vamos a realizar, en adelante, un análisis cualitativo por cada centro de interés, averiguando la convergencia y divergencia conceptuales entre los vocablos altamente disponibles de las dos lenguas y, por consiguiente, conociendo la naturaleza de la deficiencia lingüística en ELE de nuestros estudiantes.

6.1. Resultados cualitativos generales

Los resultados cualitativos de esta tesis tienen una orientación didáctica. En este apartado, vamos a revisar, por centro de interés, los vocablos altamente disponibles (con índice de disponibilidad igual o mayor que 0.1) en las dos lenguas de nuestra investigación. Exponemos, bajo cada área nocional, los vocablos con un índice de disponibilidad superior a 0.1 en ELE y en CLM del centro de interés tratado, en el orden de disponibilidad de mayor a menor.

Así pues, observamos la convergencia y divergencia conceptuales que presentan las dos listas de vocablos en ELE y en CLM. Para esto, buscamos los homólogos de los vocablos de una lengua en la lista de la otra lengua. Como podemos ver en las tablas que siguen, utilizamos el fondo gris para marcar aquellos vocablos que no tienen su sinónimo conceptual en la lista de la otra lengua. Dicho de otra manera, el fondo gris indica los conceptos únicamente presentes en la sub-muestra de una lengua y ausentes en la de la otra.

Concretamente, bajo cada uno de los dieciséis centros de interés, presentamos una tabla con tres columnas de palabras. Como podemos ver en adelante, en las columnas con los epígrafes “ELE” y “CLM” se exponen los vocablos altamente disponibles de las dos lenguas respectivas. Además, buscamos, entre las palabras de la columna “ELE”, los sinónimos conceptuales de los vocablos de la columna “CLM” y los apuntamos en su celda correspondiente en la columna con el epígrafe “Homólogo en ELE”. En el caso de que un vocablo en chino no tenga sinónimos en español en la columna “ELE” —marcado en fondo gris—, dejamos, en su celda correspondiente en la columna “Homólogo en ELE”, sus posibles traducciones al español.

Llamamos la atención sobre la importancia de estas traducciones al español, o sea, los conceptos que los alumnos saben actualizar en chino y no en español. Serán palabras que implican una deficiencia lingüística en ELE de los informantes. Así pues, a esto es adonde se puede dirigir la orientación didáctica de nuestro análisis cualitativo, puesto que recomendamos que los profesores de ELE en Tianjin se fijen en esas palabras para poder ampliar el vocabulario en español de sus estudiantes.

Como conclusión de cada área nocional, después de cada tabla, resumimos los resultados cualitativos obtenidos a partir de los datos recogidos en ella. Para ello, exponemos tres tipos de palabras según el orden de disponibilidad:

- 1) los conceptos coincidentes en las dos lenguas, o sea, conceptos que se expresan tanto en términos en ELE como en términos homólogos en CLM;
- 2) los vocablos únicamente presentes en ELE, es decir, los que no tienen su homólogo en CLM;
- 3) los vocablos ausentes en ELE, que son los homólogos en español de aquellos vocablos únicamente presentes en CLM. Estos últimos constituyen el objeto fundamental de nuestro análisis cualitativo, toda vez que son vocablos que los estudiantes deberían aprender en el aula

para compensar su deficiencia léxica en ELE.

Por último, señalamos que, como se puede ver en adelante, en la presentación de los vocablos, utilizamos el punto y coma (;) para separar los diferentes vocablos disponibles y empleamos la coma (,) para segregar los distintos homólogos en español de un mismo vocablo en chino, esto es, cuando un término chino engloba voces diferentes en español bien porque son sinónimos bien porque se refieren a distintas manifestaciones del concepto que en chino no han sido considerados con etiquetas distintas.

6.1.1. Partes del cuerpo humano

01 Partes del cuerpo humano			
	ELE	CLM	Homólogo en ELE
1	ojo	眼睛	ojo
2	mano	鼻子	nariz
3	pie	手	mano
4	nariz	耳朵	oreja
5	pelo	嘴	boca
6	cabeza	头	cabeza
7	boca	脚	pie
8	pierna	腿	pierna
9	oreja	胳膊	brazo
10	brazo	头发	pelo
11	cara	脖子	cuello
12	dedo	手指	dedo
13	corazón	胸	pecho
14	pecho	脸	cara
15	diente	眉毛	ceja
16	estómago	膝盖	rodilla
17	lengua	腰	cintura
18		肩	hombro
19		心	corazón
20		背	espalda
21		指甲	uña
22		肚子	vientre, abdomen, barriga
23		手臂	brazo
24		脚趾	dedo del pie
25		脚踝	tobillo
26		牙齿	diente
27		屁股	culo
28		大腿	muslo
29		肝	hígado
30		胃	estómago
31		肘	codo

Tabla 6.1. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes del cuerpo humano

- 1) Conceptos coincidentes en las dos lenguas:
ojo; mano; pie; nariz; pelo; cabeza; boca; pierna; oreja; brazo; cara; dedo; corazón; pecho; diente; estómago.
- 2) Vocablos únicamente presentes en ELE:
lengua.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
cuello; ceja; rodilla; cintura; hombro; espalda; uña; vientre, abdomen, barriga; dedo del pie; tobillo; culo; muslo; hígado; codo.

6.1.2. Ropa y complementos

	ELE	02 Ropa y complementos CLM	Homólogo en ELE
1	zapato	大衣	abrigo
2	pantalón	毛衣	jersey
3	camisa	项链	collar
4	chaqueta	帽子	sombrero, gorro, gorra
5	abrigo	耳环	arete
6	camiseta	围巾	bufanda, fular
7	ropa	衬衫	camisa, blusa
8	jersey	牛仔裤	vaquero
9	gafa	裙子	falda
10	falda	裤子	pantalón
11	vestido	袜子	calcetín
12	blusa	手链	pulsera
13	zapatilla	戒指	anillo
14		T 恤	camiseta
15		外套	chaqueta
16		羽绒服	plumífero
17		手套	guante
18		夹克	chaqueta
19		鞋	zapato, zapatilla
20		眼镜	gafa
21		内衣	ropa interior
22		短裤	pantalón corto, bermuda
23		靴子	bota
24		短裙	falda corta

Tabla 6.2. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Ropa y complementos

1) Conceptos coincidentes en las dos lenguas:
zapato; pantalón; camisa; chaqueta; abrigo; camiseta; jersey; gafa; falda; blusa; zapatilla.

2) Vocablos únicamente presentes en ELE:
ropa; vestido.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
collar; sombrero, gorro, gorra; arete; bufanda, fular; vaquero; calcetín; pulsera; anillo; plumífero; guante; ropa interior; pantalón corto, bermuda; bota; falda corta.

6.1.3. Partes de la casa (sin muebles)

03 Partes de la casa (sin muebles)			
	ELE	CLM	Homólogo en ELE
1	(cuarto de) baño	厨房	cocina
2	cocina	卧室	dormitorio
3	habitación	客厅	salón, sala de estar
4	dormitorio	阳台	balcón
5	puerta	窗户	ventana
6	ventana	书房	estudio
7	sala	墙壁	pared
8	suelo	地板	suelo
9	salón	卫生间	(cuarto de) baño
10	balcón	房顶	tejado
11	sala de estar	门	puerta
12	estudio	天花板	techo
13	pared	餐厅	comedor
14	cuarto	厕所	(cuarto de) baño
15	techo	房梁	viga
16		楼梯	escalera
17		走廊	pasillo
18		阁楼	ático
19		花园	jardín
20		地下室	sótano

Tabla 6.3. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes de la casa (sin muebles)

- 1) Conceptos coincidentes en las dos lenguas:
(cuarto de) baño; cocina; dormitorio; puerta; ventana; suelo; salón; balcón; sala de estar; estudio; pared; techo.
- 2) Vocablos únicamente presentes en ELE:
habitación; sala; cuarto.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
tejado; comedor; viga; escalera; pasillo; ático; jardín; sótano.

6.1.4. Los muebles de la casa

04 Los muebles de la casa			
	ELE	CLM	Homólogo en ELE
1	mesa	沙发	sofá
2	silla	椅子	silla
3	sofá	床	cama
4	cama	桌子	mesa
5	tele(visión)	衣柜	armario
6	armario	电视	tele(visión), televisor
7	lámpara	茶几	mesa de centro
8	estantería	餐桌	mesa de comedor
9	televisor	书桌	escritorio
10	puerta	冰箱	frigorífico, nevera
11	luz	台灯	lámpara de mesa
12	ordenador	床头柜	mesita de noche
13		洗衣机	lavadora
14		书柜	librería
15		空调	aire acondicionado
16		电脑	ordenador
17		书架	estantería
18		灯	lámpara, luz
19		电视柜	mueble de televisión
20		橱柜	alacena, armario de cocina
21		鞋柜	zapatero
22		马桶	inodoro, retrete, váter
23		柜子	aparador

Tabla 6.4. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los muebles de la casa

- 1) Conceptos coincidentes en las dos lenguas:
mesa; silla; sofá; cama; tele(visión); armario; lámpara; estantería; televisor; luz; ordenador.
- 2) Vocablos únicamente presentes en ELE:
puerta.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
mesa de centro; mesa de comedor; escritorio; frigorífico, nevera; lámpara de mesa; mesita de noche; lavadora; librería; aire acondicionado; mueble de televisión; alacena, armario de cocina; zapatero; inodoro, retrete, váter; aparador.

6.1.5. Alimentos y bebidas

05 Alimentos y bebidas			
	ELE	CLM	Homólogo en ELE
1	agua	可乐	coca cola
2	carne	米饭	arroz
3	vino	面条	pasta, fideo, espagueti, tallarín
4	café	面包	pan
5	pan	牛奶	leche
6	arroz	咖啡	café
7	leche	果汁	zumos
8	coca cola	汉堡	hamburguesa
9	pescado	牛肉	carne de ternera
10	verdura	奶茶	té con leche
11	tomate	雪碧	sprite
12	paella	鸡肉	pollo
13	sopa	猪肉	carne de cerdo
14	patata	鱼肉	pescado
15	naranja	水	agua
16	manzana	橙汁	zumos de naranja
17	ensalada	茶	té
18	zumos	羊肉	carne de cordero
19	vino tinto	啤酒	cerveza
20	pollo	红酒	vino tinto
21	fruta	火锅	olla caliente
22	té	包子	baozi (bollo relleno al vapor)
23	jamón	蛋糕	pastel, tarta
24	agua mineral	馒头	mantou (panecillo al vapor)
25	comida	白酒	baijiu (licor blanco chino)
26	cerveza	苹果	manzana
27		巧克力	chocolate
28		薯条	patatas fritas
29		饺子	jjaozi (ravioli chino)

Tabla 6.5. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Alimentos y bebidas

1) Conceptos coincidentes en las dos lenguas:

agua; café; pan; arroz; leche; coca cola; pescado; manzana; zumos; vino tinto; pollo; té; cerveza.

2) Vocablos únicamente presentes en ELE:

carne; vino; verdura; tomate; paella; sopa; patata; naranja; ensalada; fruta; jamón; agua mineral; comida.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

pasta, fideo, espagueti, tallarín; hamburguesa; carne de ternera; té con leche; sprite; carne de cerdo; zumo de naranja; carne de cordero; olla caliente; baozi (bollo relleno al vapor); pastel, tarta; mantou (panecillo al vapor); baijiu (licor blanco chino); chocolate; patatas fritas; jiaozi (ravioli chino).

6.1.6. Objetos colocados en la mesa para la comida

06 Objetos colocados en la mesa para la comida			
	ELE	CLM	Homólogo en ELE
1	plato	筷子	palillo
2	vaso	叉子	tenedor
3	copa	刀	cuchillo
4	cuchillo	勺子	cuchara
5	botella	盘子	plato
6	cuchara	碗	cuenco
7	palillo	杯子	vaso, taza
8	taza	碟子	platillo
9	tenedor	纸巾	pañuelo de papel
10	papel	高脚杯	copa
11	mesa	桌布	mantel
12		餐巾纸	servilleta de papel
13		盆	cuenco grande, plato hondo
14		餐巾	servilleta

Tabla 6.6. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Objetos colocados en la mesa para la comida

1) Conceptos coincidentes en las dos lenguas:

plato; vaso; copa; cuchillo; cuchara; palillo; taza; tenedor.

2) Vocablos únicamente presentes en ELE:

botella; papel; mesa.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

cuenco; platillo; pañuelo de papel; mantel; servilleta de papel; cuenco grande, plato hondo; servilleta.

6.1.7. La cocina y sus utensilios

07 La cocina y sus utensilios			
	ELE	CLM	Homólogo en ELE
1	plato	锅	olla
2	cocina	抽油烟机	campana extractora
3	cuchillo	微波炉	microondas
4	vaso	碗	cuenco
5	olla	平底锅	sartén
6	sartén	冰箱	frigorífico, nevera
7	copa	菜刀	cuchillo de cocina
8	agua	刀	cuchillo
9	fuego	锅铲	espátula
10	botella	电饭煲	olla arrocera
11	microondas	烤箱	horno
12		电磁炉	cocina de inducción
13		盆	cuenco grande, plato hondo
14		灶台	encimera de cocina
15		盘子	plato
16		勺子	cuchara
17		高压锅	olla a presión
18		洗碗机	lavavajillas
19		瓢	cucharón
20		煤气灶	fogón de gas
21		菜板	tabla de cortar
22		筷子	palillos
23		案板	tabla de cortar

Tabla 6.7. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La cocina y sus utensilios

- 1) Conceptos coincidentes en las dos lenguas:
plato; cuchillo; olla; sartén; microondas.
- 2) Vocablos únicamente presentes en ELE:
cocina; vaso; copa; agua; fuego; botella.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
campana extractora; cuenco; frigorífico, nevera; cuchillo de cocina; espátula; olla arrocera; horno; cocina de inducción; cuenco grande, plato hondo; encimera de cocina; cuchara; olla a presión; lavavajillas; cucharón; fogón de gas; tabla de cortar; palillos.

6.1.8. La escuela: sus muebles y materiales

08 La escuela: sus muebles y materiales			
	ELE	CLM	Homólogo en ELE
1	mesa	黑板	pizarra
2	silla	椅子	silla
3	libro	粉笔	tiza
4	profesor	电脑	ordenador
5	lápiz	投影仪	proyector
6	ordenador	笔	lápiz, bolí(grafo), pluma
7	papel	桌子	mesa
8	biblioteca	橡皮	goma
9	bolí(grafo)	书	libro
10	aula	多媒体	multimedia
11	pluma	铅笔	lápiz
12	pizarra	钢笔	pluma estilográfica
13	universidad	课桌	pupitre
14	clase	本子	cuaderno, libreta
15	escuela	板擦	borrador
16	estudiante	灯	lámpara
17		尺子	regla
18		讲台	tarima
19		教室	aula, clase
20		教学楼	edificio docente
21		纸	papel
22		圆珠笔	bolí(grafo)
23		书桌	pupitre
24		操场	patio de recreo
25		图书馆	biblioteca
26		书包	mochila
27		宿舍	residencia universitaria

Tabla 6.8. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La escuela: sus muebles y materiales

1) Conceptos coincidentes en las dos lenguas:

mesa; silla; libro; lápiz; ordenador; papel; biblioteca; bolí(grafo); aula; pluma; pizarra; clase.

2) Vocablos únicamente presentes en ELE:

profesor; universidad; escuela; estudiante.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

tiza; proyector; goma; multimedia; pluma estilográfica; pupitre; cuaderno, libreta; borrador; lámpara; regla; tarima; edificio docente; patio de recreo; mochila; residencia universitaria.

6.1.9. Iluminación y calefacción

09 Iluminación y calefacción			
	ELE	CLM	Homólogo en ELE
1	luz	暖气	calefacción
2	lámpara	灯	luz, lámpara
3	calefacción	台灯	lámpara de mesa
4	sol	手电筒	linterna
5	aire acondicionado	空调	aire acondicionado
6	fuego	地暖	suelo radiante
7	móvil	太阳	sol
8		蜡烛	vela, candela
9		暖气片	radiador
10		暖宝宝	parche térmico
11		吊灯	lámpara colgante
12		火炉	estufa
13		电	electricidad
14		电热毯	manta eléctrica
15		灯泡	bombilla
16		热水袋	bolsa de agua caliente
17		电暖器	calefactor
18		白炽灯	lámpara incandescente
19		手机	móvil
20		LED	LED

Tabla 6.9. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Iluminación y calefacción

1) Conceptos coincidentes en las dos lenguas:

luz; lámpara; calefacción; sol; aire acondicionado; móvil.

2) Vocablos únicamente presentes en ELE:

fuego.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

lámpara de mesa; linterna; suelo radiante; vela, candela; radiador; parche térmico; lámpara colgante; estufa; electricidad; manta eléctrica; bombilla; bolsa de agua caliente; calefactor; lámpara incandescente; LED.

6.1.10. La ciudad

	ELE	10 La ciudad CLM	Homólogo en ELE
1	coche	北京	Beijing
2	Madrid	上海	Shanghai (Shanghái)
3	edificio	天津	Tianjin
4	autobús	地铁	metro
5	ciudad	学校	escuela
6	universidad	公交车	autobús
7	calle	公园	parque
8	persona	医院	hospital
9	hospital	汽车	coche
10	Barcelona	楼房	edificio
11	Beijing	街道	calle
12	escuela	交通	tráfico
13	metro	马德里	Madrid
14	Tianjin	商场	centro comercial
15	casa		

Tabla 6.10. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La ciudad

1) Conceptos coincidentes en las dos lenguas:

coche; Madrid; edificio; autobús; calle; hospital; Beijing; escuela; metro;

Tianjin.

2) Vocablos únicamente presentes en ELE:

ciudad; universidad; persona; Barcelona; casa.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Shanghai; parque; tráfico; centro comercial.

6.1.11. El campo

	ELE	11 El campo CLM	Homólogo en ELE
1	río	土地	tierra
2	campo	田	campo
3	árbol	农民	campesino
4	campesino	田野	campo
5	flor	河	río
6	tierra	山	montaña
7	perro	树	árbol
8	casa	牛	toro, buey, vaca
9	montaña	羊	cabra, oveja
10	animal	猪	cerdo
11		鸡	gallo, gallina, pollo
12		花	flor
13		平房	casa
14		狗	perro
15		小溪	arroyo
16		草	hierba, yerba, pasto
17		鸭	pato
18		水稻	arroz
19		小麦	trigo

Tabla 6.11. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en El campo

- 1) Conceptos coincidentes en las dos lenguas:
río; campo; árbol; campesino; flor; tierra; perro; casa; montaña.
- 2) Vocablos únicamente presentes en ELE:
animal.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
toro, buey, vaca; cabra, oveja; cerdo; gallo, gallina, pollo; arroyo; hierba, yerba, pasto; pato; arroz; trigo.

6.1.12. Medios de transporte

12 Medios de transporte			
	ELE	CLM	Homólogo en ELE
1	autobús	飞机	avión
2	coche	自行车	bici(cleta)
3	tren	火车	tren
4	bici(cleta)	汽车	coche, carro
5	metro	地铁	metro
6	avión	公交车	autobús
7	a pie	高铁	tren de alta velocidad
8	barco	摩托车	moto
9	taxi	出租车	taxi
10	moto	轮船	buque
11	carro	电动车	automóvil eléctrico
12		三轮车	triciclo
13		船	barco
14		轿车	sedán, berlina
15		马车	coche de caballos
16		共享单车	bicicleta compartida

Tabla 6.12. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Medios de transporte

1) Conceptos coincidentes en las dos lenguas:

autobús; coche; tren; bici(cleta); metro; avión; barco; taxi; moto; carro.

2) Vocablos únicamente presentes en ELE:

a pie.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

tren de alta velocidad; buque; automóvil eléctrico; triciclo; sedán, berlina; coche de caballos; bicicleta compartida.

6.1.13. Trabajos del campo y del jardín

13 Trabajos del campo y del jardín			
	ELE	CLM	Homólogo en ELE
1	flor	浇水	regar
2	árbol	播种	sembrar, siembra
3	agua	施肥	poner abono
4	tierra	除草	arrancar hierbas, cortar hierbas
5	cultivar	锄头	azada
6	campesino	修剪	recortar, podar
7	campo	收割	cosechar, cosecha
8	jardín	农民	campesino
9	regar	树	árbol
10	arroz	插秧	trasplantar plántulas de arroz
11	trabajar	耕地	trabajar la tierra
12		喷农药	tirar plaguicida
13		灌溉	regar
14		花	flor
15		除虫	matar insectos
16		种树	plantar árboles

Tabla 6.13. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Trabajos del campo y del jardín

1) Conceptos coincidentes en las dos lenguas:

flor; árbol; campesino; regar.

2) Vocablos únicamente presentes en ELE:

agua; tierra; cultivar; campo; jardín; arroz; trabajar.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

sembrar, siembra; poner abono; arrancar hierbas, cortar hierbas; azada; recortar, podar; cosechar, cosecha; trasplantar plántulas de arroz; trabajar la tierra; tirar plaguicida; matar insectos; plantar árboles.

6.1.14. Los animales

14 Los animales			
	ELE	CLM	Homólogo en ELE
1	perro	狗	perro
2	gato	猫	gato
3	tigre	牛	vaca
4	león	老虎	tigre
5	pez	羊	cabra, oveja
6	mono	兔子	conejo
7	cerdo	猪	cerdo
8	caballo	鸡	pollo, gallo
9	ratón	马	caballo
10	conejo	蛇	serpiente
11	serpiente	老鼠	ratón
12	pájaro	猴子	mono
13	vaca	狮子	león
14	pollo	鸭	pato
15	oso	鱼	pez, pescado
16	gallo	熊	oso
17	zorro	大象	elefante
18	elefante	熊猫	panda
19	pescado	龙	dragón
20	panda	鸟	pájaro
21		狼	lobo
22		豹	leopardo
23		海豚	delfín
24		鲸鱼	ballena

Tabla 6.14. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los animales

1) Conceptos coincidentes en las dos lenguas:

perro; gato; tigre; león; pez; mono; cerdo; caballo; ratón; conejo; serpiente; pájaro; vaca; pollo; oso; gallo; elefante; pescado; panda.

2) Vocablos únicamente presentes en ELE:

zorro.

3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

cabra, oveja; pato; dragón; lobo; leopardo; delfín; ballena.

6.1.15. Juegos y distracciones

15 Juegos y distracciones			
	ELE	CLM	Homólogo en ELE
1	fútbol	篮球	baloncesto
2	correr	跑步	correr
3	baloncesto	足球	fútbol
4	deporte	游泳	nadar
5	nadar	羽毛球	bádminton
6	jugar	网球	tenis
7	tenis	乒乓球	ping-pong, tenis de mesa
8	juego	排球	voleibol
9	móvil	王者荣耀	Rey de Gloria (videojuego)
10	voleibol	跳绳	salto a la comba
11	cantar	跳舞	bailar
12	carta	电影	película, cine
13	bailar	唱歌	cantar
14	videojuego	台球	billar
15		纸牌	carta
16		瑜伽	yoga
17		麻将	mah-jong
18		击剑	esgrima

Tabla 6.15. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Juegos y distracciones

- 1) Conceptos coincidentes en las dos lenguas:
fútbol; correr; baloncesto; nadar; tenis; voleibol; cantar; carta; bailar.
- 2) Vocablos únicamente presentes en ELE:
deporte; jugar; juego; móvil; videojuego.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
bádminton; ping-pong, tenis de mesa; Rey de Gloria (videojuego); salto a la comba; película, cine; billar; yoga; mah-jong; esgrima.

6.1.16. Profesiones y oficios

16 Profesiones y oficios			
	ELE	CLM	Homólogo en ELE
1	profesor	老师	profesor
2	estudiante	医生	médico, doctor
3	médico	学生	estudiante, alumno
4	policía	律师	abogado
5	enfermero	护士	enfermero
6	camarero	翻译	traductor, intérprete
7	alumno	教师	profesor
8	cocinero	警察	policía
9	abogado	清洁工	limpiador, barrendero
10	doctor	厨师	cocinero
11	secretario	司机	conductor
12	actor	公务员	funcionario
13	cantante	演员	actor
14	conductor	服务员	camarero
15		工人	obrero, trabajador
16		经理	gerente
17		会计	contable
18		工程师	ingeniero
19		歌手	cantante
20		消防员	bombero
21		商人	comerciante
22		记者	periodista, reportero, corresponsal
23		作家	escritor

Tabla 6.16. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Profesiones y oficios

- 1) Conceptos coincidentes en las dos lenguas:
profesor; estudiante; médico; policía; enfermero; camarero; alumno; cocinero; abogado; doctor; actor; cantante; conductor.
- 2) Vocablos únicamente presentes en ELE:
secretario.
- 3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):
traductor, intérprete; limpiador, barrendero; funcionario; obrero, trabajador; gerente; contable; ingeniero; bombero; comerciante; periodista, reportero, corresponsal; escritor.

6.2. Incidencias cualitativas de la variable "nivel de estudio"

En el capítulo sobre el análisis cuantitativo hemos analizado las incidencias de las cinco variables extralingüísticas contempladas en nuestras encuestas: el nivel de estudio, el sexo, la procedencia, el nivel de inglés y el tipo de centro. De modo similar, nos interesa saber qué influencia ejercen estas variables sobre los resultados cualitativos. Pero nos reducimos, en esta tesis, a analizar las incidencias cualitativas de la variable "nivel de estudio", teniendo en cuenta la orientación didáctica de nuestra investigación. Tal y como hemos comentado anteriormente, el "nivel de estudio" difiere del resto de las variables —y constituye la más importante— en ser la única variable según la que los estudiantes están estratificados por naturaleza. Dicho de otra manera, los alumnos están, de por sí, separados por cursos académicos para vehicular los contenidos de enseñanza y aprendizaje que les correspondan. Así pues, los resultados cualitativos según las variantes del nivel de estudio —los cuatro cursos universitarios— pueden facilitar materiales y recursos léxicos que los profesores de ELE son capaces de utilizar directamente en el aula para sus actividades docentes sobre el vocabulario. En esto consiste el sentido didáctico práctico de esta parte de la tesis, toda vez que vamos a presentar, a continuación, resultados cualitativos de acuerdo con los cuatro estratos según el nivel de estudio, o lo que viene a ser lo mismo, en función de los grupos de estudiantes a los que los profesores imparten distintos grados de enseñanza.

Para ello, vamos a revisar, por centro de interés, los vocablos altamente disponibles (con índice de disponibilidad igual o mayor que 0.1) en las dos lenguas. Bajo cada centro de interés, exponemos cinco tablas, con los vocablos presentados según el orden de disponibilidad.

Por un lado, hemos elaborado una tabla general que permite comparar los vocablos altamente disponibles en ELE de los cuatro cursos. En ella, empleamos el **fondo gris** para indicar aquellos vocablos únicamente presentes en la sub-muestra de un curso y ausentes en las de los demás cursos, y la **negrita** para señalar los vocablos actualizados en más de una sub-muestra

pero no en todas. A partir de esta primera tabla, vamos a resumir los datos obtenidos en dos tipos de palabras:

- a1)** los vocablos en ELE coincidentes en los cuatro cursos, es decir, los que aparecen en todas las sub-muestras;
- a2)** los vocablos en ELE ausentes en cada uno de los cuatro cursos y presentes, en cambio, en otro o en otros cursos. de modo que estos vocablos señalan la deficiencia léxica de los alumnos de cada curso con respecto a los de otros cursos.

Por otro lado, hemos preparado cuatro tablas en las que se cotejan los vocablos altamente disponibles en ELE y en CLM de cada uno de los cuatro cursos. En estas tablas, utilizamos el fondo gris para marcar los vocablos que no tienen su sinónimo conceptual en la lista de la otra lengua. Es decir, en este caso, el fondo gris indica los conceptos únicamente presentes en la sub-muestra de una lengua y ausentes en la de la otra. Sobre la base de estas cuatro tablas, especificamos tres tipos de palabras al igual que en el apartado anterior:

- b1)** los conceptos coincidentes en las dos lenguas de cada curso, o sea, conceptos que se expresan tanto en términos en ELE como en términos homólogos en CLM;
- b2)** los vocablos únicamente presentes en ELE de cada curso, es decir, los que no tienen su homólogo en CLM;
- b3)** los vocablos ausentes en ELE de cada curso, que son los homólogos en español de aquellos vocablos únicamente presentes en CLM. Estos constituyen los vocablos que los estudiantes deberían aprender en el aula para compensar su deficiencia léxica en ELE con respecto a su disponibilidad léxica en CLM.

Al igual que en el apartado anterior, en el resumen de los resultados obtenidos, exhibimos los vocablos, utilizando el punto y coma (;) para separar los diferentes vocablos disponibles y empleando la coma (,) para segregar los distintos homólogos en español de un mismo vocablo en chino.

6.2.1. Partes del cuerpo humano

	01 Partes del cuerpo humano			
	Primero	Segundo	Tercero	Cuarto
1	mano	ojo	ojo	ojo
2	ojo	mano	mano	mano
3	pelo	cabeza	pie	nariz
4	pie	pie	nariz	pie
5	boca	nariz	cabeza	oreja
6	nariz	boca	boca	cabeza
7	cabeza	pelo	pierna	boca
8	cara	pierna	pelo	pierna
9	brazo	cara	oreja	pelo
10	pecho	brazo	dedo	dedo
11	oreja	dedo	brazo	brazo
12	diente	estómago	corazón	corazón
13	pierna	oreja	cara	cara
14	lengua	labio	estómago	diente
15		corazón	piel	estómago
16		diente	pecho	pecho
17			uña	uña
18			cuerpo	cuerpo
19			espalda	ceja
20			labio	cuello
21			diente	piel
22			cuello	lengua
23			lengua	labio
24				espalda

Tabla 6.17. Nociones con índice de disponibilidad superior a 0.1 en ELE en Partes del cuerpo humano en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

mano; ojo; pelo; pie; boca; nariz; cabeza; cara; brazo; oreja; diente; pierna.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos) (en comparación con los demás cursos):

Primer curso:

dedo; estómago; labio; corazón; piel; uña; cuerpo; espalda; cuello; ceja.

Segundo curso:

pecho; lengua; labio; piel; uña; cuerpo; espalda; cuello; ceja.

Tercer curso:

ceja.

Cuarto curso:

-

01 Partes del cuerpo humano			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	mano	眼睛	ojo
2	ojo	手	mano
3	pelo	鼻子	nariz
4	pie	嘴	boca
5	boca	耳朵	oreja
6	nariz	脚	pie
7	cabeza	头	cabeza
8	cara	腿	pierna
9	brazo	头发	pelo
10	pecho	胳膊	brazo
11	oreja	手指	dedo
12	diente	脖子	cuello
13	pierna	眉毛	ceja
14	lengua	胸	pecho
15		脸	cara
16		心	corazón
17		指甲	uña
18		脚踝	tobillo
19		膝盖	rodilla
20		背	espalda
21		肩	hombro
22		牙齿	diente
23		肚子	vientre, abdomen, barriga
24		脚趾	dedo del pie
25		手臂	brazo
26		腰	cintura
27		大腿	muslo

Tabla 6.18. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes del cuerpo humano en el primer curso

01 Partes del cuerpo humano Segundo curso			
	ELE	CLM	Homólogo en ELE
1	ojo	眼睛	ojo
2	mano	鼻子	nariz
3	cabeza	头	cabeza
4	pie	嘴	boca
5	nariz	手	mano
6	boca	耳朵	oreja
7	pelo	脚	pie
8	pierna	腿	pierna
9	cara	脖子	cuello
10	brazo	胳膊	brazo
11	dedo	头发	pelo
12	estómago	手指	dedo
13	oreja	脸	cara
14	labio	胸	pecho
15	corazón	眉毛	ceja
16	diente	膝盖	rodalli
17		腰	cintura
18		肩	hombro
19		背	espalda
20		肚子	vientre, abdomen, barriga
21		心	corazón
22		牙齿	diente
23		手臂	brazo
24		肘	codo
25		大腿	muslo
26		指甲	uña
27		脚踝	tobillo
28		胃	estómago
29		屁股	culo
30		肺	pulmón
31		舌头	lengua
32		肝	hígado
33		脚趾	dedo del pie
34		小腿	pierna

Tabla 6.19. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes del cuerpo humano en el segundo curso

01 Partes del cuerpo humano Tercer curso			
	ELE	CLM	Homólogo en ELE
1	ojo	眼睛	ojo
2	mano	鼻子	nariz
3	pie	耳朵	oreja
4	nariz	手	mano
5	cabeza	嘴	boca
6	boca	脚	pie
7	pierna	头	cabeza
8	pelo	腿	pierna
9	oreja	胸	pecho
10	dedo	胳膊	brazo
11	brazo	手指	dedo
12	corazón	头发	pelo
13	cara	脖子	cuello
14	estómago	背	espalda
15	piel	屁股	culo
16	pecho	腰	cintura
17	uña	指甲	uña
18	cuerpo	膝盖	rodilla
19	espalda	眉毛	ceja
20	labio	心	corazón
21	diente	肚子	vientre, abdomen, barriga
22	cuello	大腿	muslo
23	lengua	脚趾	dedo del pie
24		肾	riñón
25		肝	hígado
26		小腿	pierna
27		脑	cerebro
28		口	boca
29		胃	estómago
30		肩	hombro
31		手腕	muñeca

Tabla 6.20. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes del cuerpo humano en el tercer curso

01 Partes del cuerpo humano			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	ojo	眼睛	ojo
2	mano	鼻子	nariz
3	nariz	脚	pie
4	pie	耳朵	oreja
5	oreja	手	mano
6	cabeza	头	cabeza
7	boca	腿	pierna
8	pierna	嘴	boca
9	pelo	胳膊	brazo
10	dedo	脖子	cuello
11	brazo	手指	dedo
12	corazón	脸	cara
13	cara	头发	pelo
14	diente	腰	cintura
15	estómago	肩	hombro
16	pecho	胸	pecho
17	uña	手臂	brazo
18	cuerpo	口	boca
19	ceja	膝盖	rodilla
20	cuello	脚趾	dedo del pie
21	piel	指甲	uña
22	lengua	心	corazón
23	labio	肝	hígado
24	espalda	眉毛	ceja
25		背	espalda
26		肾	riñón
27		屁股	culo
28		脚踝	tobillo
29		脸颊	mejilla
30		肺	pulmón
31		牙齿	diente

Tabla 6.21. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes del cuerpo humano en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

mano; ojo; pelo; pie; boca; nariz; cabeza; cara; brazo; pecho; oreja; diente; pierna.

Segundo curso:

ojo; mano; cabeza; pie; nariz; boca; pelo; pierna; cara; brazo; dedo; estómago; oreja; labio; corazón; diente.

Tercer curso:

ojo; mano; pie; nariz; cabeza; boca; pierna; pelo; oreja; dedo; brazo;

corazón; cara; estómago; pecho; uña; espalda; cuello.

Cuarto curso:

*ojo; mano; nariz; pie; oreja; cabeza; boca; pierna; pelo; dedo; brazo;
corazón; cara; diente; pecho; uña; ceja; cuello; espalda.*

b2) Vocablos únicamente presentes en ELE:

Primer curso:

lengua;

Segundo curso:

-

Tercer curso:

piel; cuerpo; labio; diente; lengua.

Cuarto curso:

estómago; cuerpo; piel; lengua; labio.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

*dedo; cuello; ceja; corazón; uña; tobillo; rodilla; espalda; hombro; vientre,
abdomen, barriga; dedo del pie; cintura; muslo.*

Segundo curso:

*cuello; pecho; ceja; rodilla; cintura; hombro; espalda; vientre, abdomen,
barriga; codo; muslo; uña; tobillo; culo; pulmón; lengua; hígado; dedo del pie.*

Tercer curso:

*culo; cintura; rodilla; ceja; vientre, abdomen, barriga; muslo; dedo del pie;
riñón; hígado; cerebro; hombro; muñeca.*

Cuarto curso:

*cintura; hombro; rodilla; dedo del pie; hígado; riñón; culo, nalga; tobillo;
mejilla; pulmón.*

6.2.2. Ropa y complementos

02 Ropa y complementos				
	Primero	Segundo	Tercero	Cuarto
1	zapato	zapato	pantalón	zapato
2	ropa	pantalón	zapato	pantalón
3	chaqueta	abrigo	abrigo	abrigo
4	pantalón	camiseta	camiseta	camisa
5	camisa	camisa	camisa	chaqueta
6	jersey	chaqueta	chaqueta	jersey
7	gafa	gafa	ropa	blusa
8	camiseta	jersey	gafa	vestido
9	blusa	ropa	falda	calcetín
10		zapatilla	jersey	falda
11		vestido	bolsa	guante
12		falda	pendiente	camiseta
13		suéter	traje	ropa
14		sombrero	anillo	bufanda
15			sombrero	sombrero
16				gafa
17				pendiente

Tabla 6.22. Nociones con índice de disponibilidad superior a 0.1 en ELE en Ropa y complementos en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

zapato; ropa; chaqueta; pantalón; camisa; jersey; gafa; camiseta.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

abrigo; zapatilla; vestido; falda; suéter; sombrero; bolsa; pendiente; traje; anillo; calcetín; guante; bufanda.

Segundo curso:

blusa; bolsa; pendiente; traje; anillo; calcetín; guante; bufanda.

Tercer curso:

blusa; zapatilla; vestido; suéter; calcetín; guante; bufanda.

Cuarto curso:

zapatilla; suéter; bolsa; traje; anillo.

02 Ropa y complementos			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	zapato	项链	collar
2	ropa	大衣	abrigo
3	chaqueta	毛衣	jersey
4	pantalón	耳环	arete
5	camisa	手链	pulsera
6	jersey	衬衫	camisa, blusa
7	gafa	帽子	sombrero, gorro, gorra
8	camiseta	围巾	bufanda, fular
9	blusa	T 恤	camiseta
10		裙子	falda
11		外套	chaqueta
12		牛仔裤	vaquero
13		夹克	chaqueta
14		鞋	zapato
15		裤子	pantalón
16		戒指	anillo
17		眼镜	gafa
18		袜子	calcetín
19		手套	guante
20		短裤	pantalón corto, bermuda
21		胸针	broche
22		卫衣	sudadera
23		羽绒服	plumífero
24		耳坠	pendiente

Tabla 6.23. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Ropa y complementos en el primer curso

02 Ropa y complementos			
Segundo curso			
	ELE	CLM	Homólogo en ELE
1	zapato	大衣	abrigo
2	pantalón	毛衣	jersey, suéter
3	abrigo	帽子	sombrero
4	camiseta	衬衫	camisa
5	camisa	牛仔裤	vaquero
6	chaqueta	围巾	bufanda, fular
7	gafa	羽绒服	plumífero
8	jersey	耳环	arete
9	ropa	袜子	calcetín
10	zapatilla	项链	collar
11	vestido	裤子	pantalón
12	falda	手套	guante
13	suéter	靴子	bota
14	sombrero	夹克	chaqueta
15		短裙	falda corta
16		裙子	falda
17		戒指	anillo
18		眼镜	gafa
19		外套	chaqueta
20		连衣裙	vestido
21		短裤	pantalón corto, bermuda
22		T 恤	camiseta
23		背心	camiseta sin mangas
24		内衣	ropa interior
25		衬衣	camisa
26		运动鞋	zapato deportivo
27		鞋	zapato, zapatilla
28		长裙	falda
29		短袖	camiseta
30		风衣	gabardina
31		手链	pulsera
32		卫衣	sudadera

Tabla 6.24. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Ropa y complementos en el segundo curso

02 Ropa y complementos Tercer curso			
	ELE	CLM	Homólogo en ELE
1	pantalón	帽子	sombrero
2	zapato	毛衣	jersey
3	abrigo	项链	collar
4	camiseta	耳环	arete
5	camisa	围巾	bufanda, fular
6	chaqueta	大衣	abrigo
7	ropa	牛仔裤	vaquero
8	gafa	羽绒服	plumífero
9	falda	眼镜	gafa
10	jersey	外套	chaqueta
11	bolsa	内衣	ropa interior
12	pendiente	领带	corbata
13	traje	丝巾	bufanda de seda
14	anillo	戒指	anillo
15	sombrero	T 恤	camiseta
16		裙子	falda
17		裤子	pantalón
18		夹克	chaqueta
19		手链	pulsera
20		衬衫	camisa
21		手镯	brazalete
22		鞋	zapato
23		内裤	calzón, calzoncillo, braga
24		扣子	botón

Tabla 6.25. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Ropa y complementos en el tercer curso

02 Ropa y complementos			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	zapato	耳环	arete
2	pantalón	项链	collar
3	abrigo	大衣	abrigo
4	camisa	毛衣	jersey
5	chaqueta	帽子	sombrero
6	jersey	围巾	bufanda
7	blusa	裙子	falda
8	vestido	袜子	calcetín
9	calcetín	牛仔裤	vaquero
10	falda	裤子	pantalón
11	guante	手套	guante
12	camiseta	戒指	anillo
13	ropa	衬衫	camisa, blusa
14	bufanda	手链	pulsera
15	sombrero	T 恤	camiseta
16	gafa	靴子	bota
17	pendiente	胸针	broche
18		鞋	zapato
19		内衣	ropa interior
20		羽绒服	plumífero
21		外套	chaqueta
22		短裙	falda corta
23		手镯	brazaletes

Tabla 6.26. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Ropa y complementos en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

zapato; chaqueta; pantalón; camisa; jersey; gafa; camiseta; blusa.

Segundo curso:

zapato; pantalón; abrigo; camiseta; camisa; chaqueta; gafa; jersey; zapatilla; vestido; falda; suéter; sombrero.

Tercer curso:

pantalón; zapato; abrigo; camiseta; camisa; chaqueta; gafa; falda; jersey; anillo; sombrero.

Cuarto curso:

zapato; pantalón; abrigo; camisa; chaqueta; jersey; blusa; vestido; calcetín; falda; guante; camiseta; bufanda; sombrero.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

ropa.

Segundo curso:

ropa.

Tercer curso:

ropa; bolsa; pendiente; traje.

Cuarto curso:

ropa; gafa; pendiente.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

collar; abrigo; arete; pulsera; sombrero, gorro, gorra; bufanda, fular; falda; vaquero; anillo; calcetín; guante; pantalón corto, bermuda; broche; sudadera; plumífero; pendiente.

Segundo curso:

vaquero; bufanda, fular; plumífero; arete; calcetín; collar; guante; bota; falda corta; anillo; pantalón corto, bermuda; camiseta sin mangas; ropa interior; zapato deportivo; gabardina; pulsera; sudadera.

Tercer curso:

collar; arete; bufanda, fular; vaquero; plumífero; ropa interior; corbata; bufanda de seda; pulsera; brazalete; calzón, calzoncillo, braga; botón.

Cuarto curso:

arete; collar; vaquero; anillo; pulsera; bota; broche; ropa interior; plumífero; falda corta; brazalete.

6.2.3. Partes de la casa (sin muebles)

03 Partes de la casa (sin muebles)				
	Primero	Segundo	Tercero	Cuarto
1	(cuarto de) baño	(cuarto de) baño	(cuarto de) baño	(cuarto de) baño
2	cocina	cocina	habitación	cocina
3	dormitorio	habitación	cocina	dormitorio
4	habitación	ventana	ventana	puerta
5	sala de estar	balcón	puerta	habitación
6	estudio	puerta	pared	salón
7	sala	sala	dormitorio	techo
8	puerta	dormitorio	suelo	ventana
9	suelo	suelo	balcón	pared
10	salón	salón	sala de estar	sala
11	casa	pared	salón	balcón
12	ventana	estudio	cuarto	suelo
13	cuarto	jardín	techo	comedor
14			sala	escalera
15			comedor	cuarto
16			escalera	casa
17			jardín	

Tabla 6.27. Nociones con índice de disponibilidad superior a 0.1 en ELE en Partes de la casa (sin muebles) en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

(cuarto de) baño; cocina; dormitorio; habitación; sala; puerta; suelo; salón; ventana.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

balcón; pared; jardín; techo; comedor; escalera.

Segundo curso:

sala de estar; casa; cuarto; techo; comedor; escalera.

Tercer curso:

estudio; casa.

Cuarto curso:

sala de estar; estudio; jardín.

03 Partes de la casa (sin muebles)			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	(cuarto de) baño	厨房	cocina
2	cocina	卧室	dormitorio
3	dormitorio	客厅	sala de estar, salón
4	habitación	阳台	balcón
5	sala de estar	书房	estudio
6	estudio	卫生间	(cuarto de) baño
7	sala	房顶	tejado
8	puerta	窗户	ventana
9	suelo	地板	suelo
10	salón	门	puerta
11	casa	墙壁	pared
12	ventana	天花板	techo
13	cuarto	餐厅	comedor
14		厕所	(cuarto de) baño
15		房梁	viga
16		浴室	(cuarto de) baño
17		洗手间	(cuarto de) baño

Tabla 6.28. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes de la casa (sin muebles) en el primer curso

03 Partes de la casa (sin muebles) Segundo curso			
	ELE	CLM	Homólogo en ELE
1	(cuarto de) baño	卧室	dormitorio
2	cocina	墙壁	pared
3	habitación	客厅	salón
4	ventana	厨房	cocina
5	balcón	阳台	balcón
6	puerta	窗户	ventana
7	sala	房顶	tejado
8	dormitorio	门	puerta
9	suelo	天花板	techo
10	salón	地板	suelo
11	pared	书房	estudio
12	estudio	卫生间	(cuarto de) baño
13	jardín	楼梯	escalera
14		餐厅	comedor
15		房梁	viga
16		阁楼	ático
17		烟囱	chimenea
18		厕所	(cuarto de) baño
19		花园	jardín
20		门厅	vestíbulo
21		走廊	pasillo
22		地下室	sótano

Tabla 6.29. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes de la casa (sin muebles) en el segundo curso

03 Partes de la casa (sin muebles)			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	(cuarto de) baño	厨房	cocina
2	habitación	窗户	ventana
3	cocina	卧室	dormitorio
4	ventana	阳台	balcón
5	puerta	客厅	sala de estar, salón
6	pared	墙壁	pared
7	dormitorio	天花板	techo
8	suelo	书房	estudio
9	balcón	地板	suelo
10	sala de estar	门	puerta
11	salón	卫生间	(cuarto de) baño
12	cuarto	房顶	tejado
13	techo	厕所	(cuarto de) baño
14	sala	花园	jardín
15	comedor	楼梯	escalera
16	escalera	餐厅	comedor
17	jardín	浴室	(cuarto de) baño
18		阁楼	ático
19		车库	garaje
20		房间	habitación, cuarto, sala

Tabla 6.30. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes de la casa (sin muebles) en el tercer curso

03 Partes de la casa (sin muebles)			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	(cuarto de) baño	厨房	cocina
2	cocina	卧室	dormitorio
3	dormitorio	客厅	salón
4	puerta	地板	suelo
5	habitación	阳台	balcón
6	salón	天花板	techo
7	techo	卫生间	(cuarto de) baño
8	ventana	房顶	tejado
9	pared	窗户	ventana
10	sala	墙壁	pared
11	balcón	门	puerta
12	suelo	书房	estudio
13	comedor	走廊	pasillo
14	escalera	厕所	(cuarto de) baño
15	cuarto	房梁	viga
16	casa	地下室	sótano
17		玄关	zaguán
18		餐厅	comedor
19		楼梯	escalera
20		花园	jardín

Tabla 6.31. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Partes de la casa (sin muebles) en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

(cuarto de) baño; cocina; dormitorio; sala de estar; estudio; puerta; suelo; salón; ventana.

Segundo curso:

(cuarto de) baño; cocina; ventana; balcón; puerta; dormitorio; suelo; salón; pared; estudio; jardín.

Tercer curso:

(cuarto de) baño; habitación; cocina; ventana; puerta; pared; dormitorio; suelo; balcón; sala de estar; salón; cuarto; techo; sala; comedor; escalera; jardín.

Cuarto curso:

(cuarto de) baño; cocina; dormitorio; puerta; salón; techo; ventana; pared; balcón; suelo; comedor; escalera.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

habitación; sala; casa; cuarto.

Segundo curso:

habitación; sala.

Tercer curso:

-

Cuarto curso:

habitación; sala; cuarto, casa.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

balcón; tejado; pared; techo; comedor; viga.

Segundo curso:

tejado; techo; escalera; comedor; viga; ático; chimenea; vestíbulo; pasillo; sótano.

Tercer curso:

estudio; tejado; ático; garaje.

Cuarto curso:

tejado; estudio; pasillo; viga; sótano; zaguán; jardín.

6.2.4. Los muebles de la casa

04 Los muebles de la casa				
	Primero	Segundo	Tercero	Cuarto
1	mesa	mesa	mesa	mesa
2	silla	silla	cama	silla
3	sofá	cama	sofá	cama
4	cama	sofá	silla	sofá
5	estantería	armario	tele(visión)	armario
6	tele(visión)	tele(visión)	armario	tele(visión)
7	lámpara	lámpara	lámpara	lámpara
8	puerta	luz	ordenador	televisor
9	armario	televisor	estantería	nevera
10	vaso	estantería	televisor	aire acondicionado
11	suelo	sillón	nevera	lavadora
12		ordenador	lavadora	puerta
13		puerta	aire acondicionado	calefacción
14			luz	
15			refrigerador	

Tabla 6.32. Nociones con índice de disponibilidad superior a 0.1 en ELE en Los muebles de la casa en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

mesa; silla; sofá; cama; tele(visión); lámpara; armario.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

luz; televisor; sillón; ordenador; nevera; lavadora; aire acondicionado; refrigerador; calefacción.

Segundo curso:

vaso; suelo; nevera; lavadora; aire acondicionado; refrigerador; calefacción.

Tercer curso:

puerta; vaso; suelo; sillón; calefacción.

Cuarto curso:

estantería; vaso; suelo; luz; sillón; ordenador; refrigerador.

04 Los muebles de la casa Primer curso			
	ELE	CLM	Homólogo en ELE
1	mesa	沙发	sofá
2	silla	椅子	silla
3	sofá	衣柜	armario
4	cama	床	cama
5	estantería	桌子	mesa
6	tele(visión)	电视	tele(visión)
7	lámpara	茶几	mesa de centro
8	puerta	书桌	escritorio
9	armario	冰箱	frigorífico, nevera
10	vaso	台灯	lámpara de mesa
11	suelo	餐桌	mesa de comedor
12		空调	aire acondicionado
13		电脑	ordenador, computadora
14		床头柜	mesita de noche
15		书柜	librería
16		鞋柜	zapatero
17		洗衣机	lavadora
18		书架	estantería
19		灯	lámpara
20		马桶	inodoro, retrete, váter
21		柜子	aparador
22		梳妆台	tocador

Tabla 6.33. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los muebles de la casa en el primer curso

04 Los muebles de la casa Segundo curso			
	ELE	CLM	Homólogo en ELE
1	mesa	沙发	sofá
2	silla	椅子	silla
3	cama	床	cama
4	sofá	衣柜	armario
5	armario	桌子	mesa
6	tele(visión)	电视	tele(visión), televisor
7	lámpara	茶几	mesa de centro
8	luz	餐桌	mesa de comedor
9	televisor	书桌	escritorio
10	estantería	床头柜	mesita de noche
11	sillón	台灯	lámpara de mesa
12	ordenador	冰箱	frigorífico, nevera
13	puerta	洗衣机	lavadora
14		书柜	librería
15		空调	aire acondicionado
16		电视柜	mueble de televisión
17		橱柜	alacena, armario de cocina
18		凳子	banco, banqueta, taburete
19		电脑	ordenador
20		书架	estantería
21		地毯	alfombra
22		窗帘	cortina
23		灯	lámpara, luz
24		鞋柜	zapatero
25		吊灯	lámpara colgante

Tabla 6.34. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los muebles de la casa en el segundo curso

04 Los muebles de la casa Tercer curso			
	ELE	CLM	Homólogo en ELE
1	mesa	桌子	mesa
2	cama	床	cama
3	sofá	沙发	sofá
4	silla	椅子	silla
5	tele(visión)	电视	tele(visión), televisor
6	armario	衣柜	armario
7	lámpara	冰箱	nevera, refrigerador
8	ordenador	洗衣机	lavadora
9	estantería	餐桌	mesa de comedor
10	televisor	灯	lámpara, luz
11	nevera	茶几	mesa de centro
12	lavadora	床头柜	mesita de noche
13	aire acondicionado	电脑	ordenador
14	luz	空调	aire acondicionado
15	refrigerador	电视柜	mueble de televisión
16		书柜	librería
17		台灯	lámpara de mesa
18		书桌	escritorio
19		微波炉	microondas
20		书架	estantería
21		梳妆台	tocador
22		窗帘	cortina
23		洗碗机	lavavajillas

Tabla 6.35. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los muebles de la casa en el tercer curso

04 Los muebles de la casa			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	mesa	沙发	sofá
2	silla	椅子	silla
3	cama	床	cama
4	sofá	桌子	mesa
5	armario	衣柜	armario
6	tele(visión)	电视	tele(visión), televisor
7	lámpara	茶几	mesa de centro
8	televisor	冰箱	nevera
9	nevera	台灯	lámpara de mesa
10	aire acondicionado	餐桌	mesa de comedor
11	lavadora	书柜	librería
12	puerta	书桌	escritorio
13	calefacción	洗衣机	lavadora
14		床头柜	mesita de noche
15		空调	aire acondicionado
16		马桶	inodoro, retrete, váter
17		书架	estantería
18		梳妆台	tocador
19		电视柜	mueble de televisión
20		电脑	ordenador, computadora
21		柜子	aparador
22		板凳	banco, banqueta, taburete
23		橱柜	alacena, armario de cocina

Tabla 6.36. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los muebles de la casa en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

mesa; silla; sofá; cama; tele(visión); lámpara; armario.

Segundo curso:

mesa; silla; sofá; cama; estantería; tele(visión); lámpara; armario.

Tercer curso:

mesa; silla; cama; sofá; armario; tele(visión); lámpara; luz; televisor; estantería; ordenador.

Cuarto curso:

mesa; cama; sofá; silla; tele(visión); armario; lámpara; ordenador; estantería; televisor; nevera; lavadora; aire acondicionado; luz; refrigerador.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

vaso; suelo.

Segundo curso:

puerta; vaso; suelo.

Tercer curso:

sillón; puerta.

Cuarto curso:

-

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

mesa de centro; escritorio; frigorífico, nevera; lámpara de mesa; mesa de comedor; aire acondicionado; ordenador, computadora; mesita de noche; librería; zapatero; lavadora; inodoro, retrete, váter; aparador; tocador.

Segundo curso:

mesa de centro; mesa de comedor; escritorio; mesita de noche; lámpara de mesa; frigorífico, nevera; lavadora; librería; aire acondicionado; mueble de televisión; alacena, armario de cocina; banco, banqueta, taburete; alfombra; cortina; zapatero; lámpara colgante.

Tercer curso:

mesa de comedor; mesa de centro; mesita de noche; mueble de televisión; librería; lámpara de mesa; escritorio; microondas; tocador; cortina; lavavajillas.

Cuarto curso:

mesa de centro; lámpara de mesa; mesa de comedor; librería; escritorio; mesita de noche; inodoro, retrete, váter; estantería; tocador; mueble de televisión; ordenador, computadora; aparador; banco, banqueta, taburete; alacena, armario de cocina.

6.2.5. Alimentos y bebidas

05 Alimentos y bebidas				
	Primero	Segundo	Tercero	Cuarto
1	carne	agua	agua	coca cola
2	agua	arroz	arroz	agua
3	vino	pan	carne	arroz
4	verdura	café	coca cola	café
5	ensalada	leche	zum	pan
6	pescado	carne	vino	vino
7	café	vino	pan	carne
8	sopa	paella	leche	patata
9	leche	naranja	manzana	tomate
10	pan	pescado	tomate	leche
11	agua mineral	coca cola	fruta	hamburguesa
12	vino tinto	manzana	paella	manzana
13	tomate	jamón	verdura	naranja
14	paella	tomate	patata	zum
15	marisco	patata	pescado	cerveza
16	pollo	huevo	cerveza	verdura
17	coca cola	tapa	sopa	pollo
18	patata	sopa	naranja	paella
19	comida	té	comida	fruta
20	fruta	refresco	pollo	té
21	arroz	vino tinto	bebida	sangría
22		verdura	limón	tortilla
23		zum	jamón	
24		tortilla	café	
		sangría	cerdo	
		pollo	huevo	
			tapa	
			refresco	
			té	

Tabla 6.37. Nociones con índice de disponibilidad superior a 0.1 en ELE en Alimentos y bebidas en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

carne; agua; vino; verdura; café; leche; pan; tomate; paella; pollo; coca cola; patata; arroz.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

naranja; manzana; jamón; huevo; tapa; té; refresco; zumo; tortilla; sangría; cerveza; bebida; limón; cerdo; hamburguesa.

Segundo curso:

ensalada; agua mineral; marisco; comida; fruta; cerveza; bebida; limón; cerdo; hamburguesa.

Tercer curso:

ensalada; agua mineral; vino tinto; marisco; tortilla; sangría; hamburguesa.

Cuarto curso:

ensalada; pescado; sopa; agua mineral; vino tinto; marisco; comida; jamón; huevo; tapa; refresco; bebida; limón; cerdo.

05 Alimentos y bebidas			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	carne	可乐	coca cola
2	agua	米饭	arroz
3	vino	面包	pan
4	verdura	面条	pasta, fideo, espagueti, tallarín
5	ensalada	果汁	zumo, jugo
6	pescado	牛奶	leche
7	café	水	agua
8	sopa	牛肉	carne de ternera
9	leche	鱼肉	pescado
10	pan	鸡肉	pollo
11	agua mineral	奶茶	té con leche
12	vino tinto	雪碧	sprite
13	tomate	猪肉	carne de cerdo
14	paella	羊肉	carne de cordero
15	marisco	咖啡	café
16	pollo	橙汁	zumo de naranja
17	coca cola	蛋糕	pastel, tarta
18	patata	饼干	galleta
19	comida	蔬菜	verdura
20	fruta	啤酒	cerveza
21	arroz	馒头	mantou (panecillo al vapor)
22		肉	carne
23		巧克力	chocolate
24		包子	baozi (bollo relleno al vapor)
25		汉堡	hamburguesa

Tabla 6.38. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Alimentos y bebidas en el primer curso

05 Alimentos y bebidas			
Segundo curso			
	ELE	CLM	Homólogo en ELE
1	agua	可乐	coca cola
2	arroz	米饭	arroz
3	pan	面条	pasta, fideo, espagueti, tallarín
4	café	咖啡	café
5	leche	汉堡	hamburguesa
6	carne	牛奶	leche
7	vino	火锅	olla caliente
8	paella	雪碧	sprite
9	naranja	面包	pan
10	pescado	牛肉	carne de ternera
11	coca cola	奶茶	té con leche
12	manzana	鸡肉	pollo
13	jamón	猪肉	carne de cerdo
14	tomate	橙汁	zumو de naranja
15	patata	果汁	zumو
16	huevo	红酒	vino tinto
17	tapa	鱼肉	pescado
18	sopa	白菜	col, repollo
19	té	啤酒	cerveza
20	refresco	羊肉	carne de cordero
21	vino tinto	茶	té
22	verdura	薯条	patatas fritas
23	zumو	苹果	manzana
24	tortilla	白酒	baijiu (licor blanco chino)
25	sangría	饺子	jiaozi (ravioli chino)
26	pollo	包子	baozi (bollo relleno al vapor)
27		巧克力	chocolate
28		饼干	galleta
29		蛋糕	pastel, tarta

Tabla 6.39. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Alimentos y bebidas en el segundo curso

05 Alimentos y bebidas			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	agua	米饭	arroz
2	arroz	可乐	coca cola
3	carne	汉堡	hamburguesa
4	coca cola	果汁	zumó
5	zumó	奶茶	té con leche
6	vino	咖啡	café
7	pan	茶	té
8	leche	面包	pan
9	manzana	牛奶	leche
10	tomate	雪碧	sprite
11	fruta	红酒	vino tinto
12	paella	啤酒	cerveza
13	verdura	水	agua
14	patata	水果	fruta
15	pescado	肉	carne
16	cerveza	白酒	baijiu (licor blanco chino)
17	sopa	面条	pasta, fideo, espagueti, tallarín
18	naranja	薯条	patatas fritas
19	comida	玉米	maíz
20	pollo	海鲜饭	paella
21	bebida	蛋糕	pastel, tarta
22	limón	酸奶	yogur
23	jamón	大米	arroz
24	café	煎饼	jianbing (tortita frita)
25	cerdo	火锅	olla caliente
26	huevo	苹果	manzana
27	tapa		
28	refresco		
29	té		

Tabla 6.40. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Alimentos y bebidas en el tercer curso

05 Alimentos y bebidas			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	coca cola	可乐	coca cola
2	agua	米饭	arroz
3	arroz	汉堡	hamburguesa
4	café	面条	pasta, fideo, espagueti, tallarín
5	pan	咖啡	café
6	vino	茶	té
7	carne	牛肉	carne de ternera
8	patata	猪肉	carne de cerdo
9	tomate	牛奶	leche
10	leche	鸡肉	pollo
11	hamburguesa	水	agua
12	manzana	馒头	mantou (panecillo al vapor)
13	naranja	雪碧	sprite
14	zummo	包子	baozi (bollo relleno al vapor)
15	cerveza	橙汁	zummo de naranja
16	verdura	奶茶	té con leche
17	pollo	鱼肉	pescado
18	paella	大饼	torta
19	fruta	饺子	jiaozi (ravioli chino)
20	té	果汁	zummo
21	sangría	羊肉	carne de cordero
22	tortilla	白酒	baijiu (licor blanco chino)
23		土豆	patata
24		红酒	vino tinto
25		面包	pan
26		披萨	pizza
27		啤酒	cerveza
28		苹果	manzana

Tabla 6.41. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Alimentos y bebidas en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

carne; agua; verdura; pescado; café; leche; pan; pollo; coca cola; arroz.

Segundo curso:

arroz; pan; café; leche; pescado; coca cola; manzana; té; vino tinto; zumo; pollo.

Tercer curso:

agua; arroz; carne; coca cola; zumo; pan; leche; manzana; fruta; paella; cerveza; café; té.

Cuarto curso:

coca cola; agua; arroz; café; pan; patata; leche; hamburguesa; manzana; zumo; cerveza; verdura; pollo; té.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

vino; ensalada; sopa; agua mineral; vino tinto; tomate; paella; marisco; patata; comida; fruta.

Segundo curso:

agua; carne; vino; paella; naranja; jamón; tomate; patata; huevo; tapa; sopa; refresco; verdura; tortilla; sangría.

Tercer curso:

vino; tomate; verdura; patata; pescado; sopa; naranja; comida; pollo; bebida; limón; jamón; cerdo; huevo; tapa; refresco.

Cuarto curso:

vino; carne; tomate; naranja; paella; fruta; sangría; tortilla.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

pasta, fideo, espagueti, tallarín; zumo, jugo; carne de ternera; té con leche; sprite; carne de cerdo; carne de cordero; zumo de naranja; pastel, tarta; galleta; cerveza; mantou (panecillo al vapor); chocolate; baozi (bollo relleno al vapor); hamburguesa.

Segundo curso:

pasta, fideo, espagueti, tallarín; hamburguesa; olla caliente; sprite; carne de ternera; té con leche; carne de cerdo; zumo de naranja; col, repollo; cerveza; carne de cordero; patatas fritas; baijiu (licor blanco chino); jiaozi (ravioli chino); baozi (bollo relleno al vapor); chocolate; galleta; pastel, tarta.

Tercer curso:

hamburguesa; té con leche; sprite; vino tinto; baijiu (licor blanco chino); pasta, fideo, espagueti, tallarín; patatas fritas; maíz; pastel, tarta; yogur; jianbing (tortita frita); olla caliente.

Cuarto curso:

pasta, fideo, espagueti, tallarín; carne de ternera; carne de cerdo; mantou

(panecillo al vapor); sprite; baozi (bollo relleno al vapor); zumo de naranja; té con leche; pescado; torta, pan plano; jiaozi (ravioli chino); carne de cordero; baijiu (licor blanco chino); vino tinto; pizza.

6.2.6. Objetos colocados en la mesa para la comida

06 Objetos colocados en la mesa para la comida				
	Primero	Segundo	Tercero	Cuarto
1	plato	plato	plato	plato
2	vaso	cuchillo	cuchillo	cuchillo
3	copa	vaso	copa	cuchara
4	botella	copa	vaso	tenedor
5	papel	botella	cuchara	palillo
6	taza	taza	botella	vaso
7	mesa	palillo	tenedor	copa
8	palillo	papel	mesa	taza
9	cuchara	tenedor	palillo	servilleta
10	tazón	cuchara	papel	mesa
11			taza	botella
12				papel

Tabla 6.42. Nociones con índice de disponibilidad superior a 0.1 en ELE en Objetos colocados en la mesa para la comida en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

plato; vaso; copa; botella; papel; taza; palillo; cuchara.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

cuchillo; tenedor; servilleta.

Segundo curso:

mesa; tazón; servilleta.

Tercer curso:

tazón; servilleta.

Cuarto curso:

tazón.

06 Objetos colocados en la mesa para la comida			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	plato	筷子	palillo
2	vaso	叉子	tenedor
3	copa	盘子	plato
4	botella	勺子	cuchara
5	papel	刀	cuchillo
6	taza	碗	cuenco
7	mesa	杯子	vaso, taza, tazón
8	palillo	碟子	platillo
9	cuchara	餐巾纸	servilleta de papel
10	tazón	高脚杯	copa
11		酒杯	vaso de licor, copa de vino
12		盆	cuenco grande, plato hondo

Tabla 6.43. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Objetos colocados en la mesa para la comida en el primer curso

06 Objetos colocados en la mesa para la comida			
Segundo curso			
	ELE	CLM	Homólogo en ELE
1	plato	筷子	palillo
2	cuchillo	叉子	tenedor
3	vaso	刀	cuchillo
4	copa	勺子	cuchara
5	botella	碗	cuenco
6	taza	盘子	plato
7	palillo	碟子	platillo
8	papel	杯子	vaso, taza
9	tenedor	桌布	mantel
10	cuchara	盆	cuenco grande, plato hondo
11		纸巾	pañuelo de papel
12		餐巾纸	servilleta de papel
13		茶杯	vaso de té, taza de té
14		高脚杯	copa

Tabla 6.44. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Objetos colocados en la mesa para la comida en el segundo curso

06 Objetos colocados en la mesa para la comida			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	plato	筷子	palillo
2	cuchillo	叉子	tenedor
3	copa	刀	cuchillo
4	vaso	勺子	cuchara
5	cuchara	盘子	plato
6	botella	碗	cuenco
7	tenedor	杯子	vaso, taza
8	mesa	纸巾	pañuelo de papel
9	palillo	碟子	platillo
10	papel	盆	cuenco grande, plato hondo
11	taza	餐巾	servilleta
12		高脚杯	copa
13		桌布	mantel
14		酒杯	vaso de licor, copa de vino

Tabla 6.45. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Objetos colocados en la mesa para la comida en el tercer curso

06 Objetos colocados en la mesa para la comida			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	plato	叉子	tenedor
2	cuchillo	筷子	palillo
3	cuchara	刀	cuchillo
4	tenedor	碗	cuenco
5	palillo	勺子	cuchara
6	vaso	盘子	plato
7	copa	碟子	platillo
8	taza	杯子	vaso, taza
9	servilleta	纸巾	pañuelo de papel
10	mesa	餐巾	servilleta
11	botella	桌布	mantel
12	papel	水杯	vaso de agua
13		高脚杯	copa
14		餐巾纸	servilleta de papel
15		酒杯	vaso de licor, copa de vino

Tabla 6.46. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Objetos colocados en la mesa para la comida en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

plato; vaso; copa; taza; palillo; cuchara; tazón.

Segundo curso:

plato; cuchillo; vaso; copa; taza; palillo; tenedor; cuchara.

Tercer curso:

plato; cuchillo; copa; vaso; cuchara; tenedor; palillo; taza.

Cuarto curso:

plato; cuchillo; cuchara; tenedor; palillo; vaso; copa; taza; servilleta.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

botella; papel; mesa.

Segundo curso:

botella; papel.

Tercer curso:

botella; mesa; papel.

Cuarto curso:

mesa; botella; papel.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

tenedor; cuchillo; cuenco; platillo; servilleta de papel; vaso de licor, copa de vino; cuenco grande, plato hondo.

Segundo curso:

cuenco; platillo; mantel; cuenco grande, plato hondo; pañuelo de papel; servilleta de papel; vaso de té, taza de té.

Tercer curso:

cuenco; pañuelo de papel; platillo; cuenco grande, plato hondo; servilleta; mantel; vaso de licor, copa de vino.

Cuarto curso:

cuenco; platillo; pañuelo de papel; mantel; vaso de agua; servilleta de papel; vaso de licor, copa de vino.

6.2.7. La cocina y sus utensilios

07 La cocina y sus utensilios				
	Primero	Segundo	Tercero	Cuarto
1	plato	plato	cuchillo	plato
2	cocina	cocina	plato	olla
3	vaso	cuchillo	fuego	cuchillo
4	copa	sartén	olla	sartén
5	agua	vaso	microondas	fuego
6	botella	olla	sartén	microondas
7	olla	copa	refrigerador	cuchara
8	mesa	botella	lavadora	nevera
9		fuego	agua	horno
10		sal	gas	tenedor
11		aceite	nevera	sal
12		agua	sal	agua
13			horno	
14			cocina	

Tabla 6.47. Nociones con índice de disponibilidad superior a 0.1 en ELE en La cocina y sus utensilios en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

plato; agua; olla.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

cuchillo; sartén; fuego; sal; aceite; microondas; refrigerador; lavadora; gas; nevera; horno; cuchara; tenedor.

Segundo curso:

mesa; microondas; refrigerador; lavadora; gas; nevera; horno; cuchara; tenedor.

Tercer curso:

vaso; copa; botella; mesa; aceite; cuchara; tenedor.

Cuarto curso:

cocina; vaso; copa; botella; mesa; fuego; aceite; refrigerador; lavadora; gas; tenedor.

07 La cocina y sus utensilios			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	plato	锅	olla
2	cocina	抽油烟机	campana extractora
3	vaso	微波炉	microondas
4	copa	碗	cuenco
5	agua	平底锅	sartén
6	botella	冰箱	frigorífico, nevera
7	olla	电饭煲	olla arrocera
8	mesa	刀	cuchillo
9		电磁炉	cocina de inducción
10		菜刀	cuchillo de cocina
11		锅铲	espátula
12		烤箱	horno
13		勺子	cuchara
14		盘子	plato
15		灶台	encimera de cocina
16		盆	cuenco grande, plato hondo
17		高压锅	olla a presión
18		筷子	palillos
19		瓢	cucharón
20		案板	tabla de cortar
21		煤气灶	fogón de gas
22		油	aceite

Tabla 6.48. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La cocina y sus utensilios en el primer curso

07 La cocina y sus utensilios			
Segundo curso			
	ELE	CLM	Homólogo en ELE
1	plato	锅	olla
2	cocina	抽油烟机	campana extractora
3	cuchillo	微波炉	microondas
4	sartén	菜刀	cuchillo de cocina
5	vaso	碗	cuenco
6	olla	平底锅	sartén
7	copa	电饭煲	olla arrocera
8	botella	盆	cuenco grande, plato hondo
9	fuego	烤箱	horno
10	sal	刀	cuchillo
11	aceite	冰箱	frigorífico, nevera
12	agua	锅铲	espátula
13		洗碗机	lavavajillas
14		瓢	cucharón
15		勺子	cuchara
16		高压锅	olla a presión
17		砧板	tabla de cortar
18		电磁炉	cocina de inducción
19		煤气灶	fogón de gas
20		盘子	plato
21		案板	tabla de cortar
22		蒸锅	olla vaporera
23		天然气	gas natural

Tabla 6.49. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La cocina y sus utensilios en el segundo curso

07 La cocina y sus utensilios			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	cuchillo	锅	olla
2	plato	冰箱	refrigerador, nevera
3	fuego	微波炉	microondas
4	olla	抽油烟机	campana extractora
5	microondas	碗	cuenco
6	sartén	刀	cuchillo
7	refrigerador	灶台	encimera de cocina
8	lavadora	电磁炉	cocina de inducción
9	agua	烤箱	horno
10	gas	平底锅	sartén
11	nevera	水龙头	grifo
12	sal	盘子	plato
13	horno	炉子	estufa
14	cocina	洗碗机	lavavajillas
15		锅铲	espátula
16		菜刀	cuchillo de cocina
17		叉子	tenedor
18		盆	cuenco grande, plato hondo
19		橱柜	alacena, armario de cocina
20		筷子	palillos
21		炒锅	wok
22		榨汁机	licuadora
23		高压锅	olla a presión
24		蒸锅	olla vaporera
25		勺子	cuchara
26		瓢	cucharón
27		砂锅	cazuela de barro
28		碗柜	alacena, armario de cocina
29		厨房	cocina

Tabla 6.50. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La cocina y sus utensilios en el tercer curso

07 La cocina y sus utensilios			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	plato	抽油烟机	campana extractora
2	olla	锅	olla
3	cuchillo	碗	cuenco
4	sartén	微波炉	microondas
5	fuego	刀	cuchillo
6	microondas	菜刀	cuchillo de cocina
7	cuchara	菜板	tabla de cortar
8	nevera	灶台	encimera de cocina
9	horno	锅铲	espátula
10	tenedor	平底锅	sartén
11	sal	冰箱	nevera
12	agua	盆	cuenco grande, plato hondo
13		炒锅	wok
14		洗碗机	lavavajillas
15		烤箱	horno
16		电饭煲	olla arrocera
17		高压锅	olla a presión
18		水龙头	grifo
19		煤气灶	fogón de gas
20		盘子	plato
21		盐	sal
22		汤勺	cucharón
23		橱柜	alacena, armario de cocina
24		电磁炉	cocina de inducción
25		洗洁精	líquido lavavajillas
26		水池	fregadero, pila de fregar
27		炉子	estufa
28		瓢	cucharón

Tabla 6.51. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La cocina y sus utensilios en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

plato; olla.

Segundo curso:

plato; cuchillo; sartén; olla.

Tercer curso:

cuchillo; plato; olla; microondas; sartén; refrigerador; nevera; horno; cocina.

Cuarto curso:

plato; olla; cuchillo; sartén; microondas; nevera; horno; sal.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

cocina; vaso; copa; agua; botella; mesa.

Segundo curso:

cocina; vaso; copa; botella; fuego; sal; aceite; agua.

Tercer curso:

fuego; lavadora; agua; gas; sal.

Cuarto curso:

fuego; cuchara; tenedor; agua.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

campana extractora; microondas; cuenco; sartén; frigorífico, nevera; olla arrocera; cuchillo; cocina de inducción; cuchillo de cocina; espátula; horno; cuchara; encimera de cocina; cuenco grande, plato hondo; olla a presión; palillos; cucharón; tabla de cortar; fogón de gas; aceite.

Segundo curso:

campana extractora; microondas; cuchillo de cocina; cuenco; olla arrocera; cuenco grande, plato hondo; horno; frigorífico, nevera; espátula; lavavajillas; cucharón; cuchara; olla a presión; tabla de cortar; cocina de inducción; fogón de gas; olla vaporera; gas natural.

Tercer curso:

campana extractora; cuenco; encimera de cocina; cocina de inducción; grifo; estufa; lavavajillas; espátula; cuchillo de cocina; tenedor; cuenco grande, plato hondo; alacena, armario de cocina; palillos; wok; licuadora; olla a presión; olla vaporera; cuchara; cucharón; cazuela de barro.

Cuarto curso:

campana extractora; cuenco; cuchillo de cocina; tabla de cortar; encimera de cocina; espátula; cuenco grande, plato hondo; wok; lavavajillas; olla arrocera; olla a presión; grifo; fogón de gas; cucharón; alacena, armario de cocina; cocina de inducción; líquido lavavajillas; fregadero, pila de fregar; estufa; cucharón.

6.2.8. La escuela: sus muebles y materiales

08 La escuela: sus muebles y materiales				
	Primero	Segundo	Tercero	Cuarto
1	mesa	mesa	mesa	mesa
2	libro	libro	silla	pizarra
3	silla	silla	ordenador	silla
4	biblioteca	profesor	libro	libro
5	lápiz	papel	lápiz	bolí(grafo)
6	profesor	pluma	bolí(grafo)	lápiz
7	escuela	lápiz	profesor	ordenador
8	universidad	ordenador	papel	papel
9	pluma	aula	aula	aula
10	clase	bolí(grafo)	clase	pluma
11	papel	biblioteca	biblioteca	profesor
12	estudiante	estudiante	pizarra	borrador
13	bolí(grafo)	escuela	pantalla	pantalla
14	aula	universidad	estudiante	biblioteca
15	ordenador	edificio	universidad	tiza
16	bachillerato	clase	móvil	computadora
17		comedor		

Tabla 6.52. Nociones con índice de disponibilidad superior a 0.1 en ELE en La escuela: sus muebles y materiales en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

mesa; libro; silla; biblioteca; lápiz; profesor; papel; bolí(grafo); aula; ordenador.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

edificio; comedor; pizarra; pantalla; móvil; borrador; tiza; computadora.

Segundo curso:

bachillerato; pizarra; pantalla; móvil; borrador; tiza; computadora.

Tercer curso:

escuela; pluma; bachillerato; edificio; comedor; borrador; tiza; computadora.

Cuarto curso:

escuela; universidad; clase; estudiante; bachillerato; edificio; comedor; móvil.

08 La escuela: sus muebles y materiales			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	mesa	黑板	pizarra
2	libro	粉笔	tiza
3	silla	椅子	silla
4	biblioteca	笔	pluma
5	lápiz	电脑	ordenador
6	profesor	投影仪	proyector
7	escuela	书	libro
8	universidad	橡皮	goma
9	pluma	本子	cuaderno, libreta
10	clase	铅笔	lápiz
11	papel	课桌	pupitre
12	estudiante	多媒体	multimedia
13	bolí(grafo)	钢笔	pluma estilográfica
14	aula	讲台	tarima
15	ordenador	桌子	mesa
16	bachillerato	教室	clase, aula
17		尺子	regla
18		教学楼	edificio docente
19		图书馆	biblioteca
20		板擦	borrador
21		圆珠笔	bolí(grafo)
22		操场	patio de recreo
23		纸	papel
24		食堂	comedor
25		宿舍	residencia universitaria
26		学生	estudiante
27		体育馆	gimnasio, estadio

Tabla 6.53. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La escuela: sus muebles y materiales en el primer curso

08 La escuela: sus muebles y materiales			
Segundo curso			
	ELE	CLM	Homólogo en ELE
1	mesa	黑板	pizarra
2	libro	椅子	silla
3	silla	电脑	ordenador
4	profesor	投影仪	proyector
5	papel	粉笔	tiza
6	pluma	桌子	mesa
7	lápiz	书	libro
8	ordenador	灯	lámpara
9	aula	橡皮	goma
10	bolí(grafo)	笔	pluma
11	biblioteca	钢笔	pluma estilográfica
12	estudiante	多媒体	multimedia
13	escuela	铅笔	lápiz
14	universidad	尺子	regla
15	edificio	板擦	borrador
16	clase	教学楼	edificio docente
17	comedor	课桌	pupitre
18		书桌	pupitre
19		本子	cuaderno, libreta
20		讲台	tarima
21		电视	televisor
22		圆珠笔	bolí(grafo)
23		纸	papel
24		书包	mochila

Tabla 6.54. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La escuela: sus muebles y materiales en el segundo curso

08 La escuela: sus muebles y materiales			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	mesa	黑板	pizarra
2	silla	椅子	silla
3	ordenador	粉笔	tiza
4	libro	笔	lápiz, bolí(grafo)
5	lápiz	投影仪	proyector
6	bolí(grafo)	电脑	ordenador
7	profesor	橡皮	goma
8	papel	多媒体	multimedia
9	aula	纸	papel
10	clase	教室	aula, clase
11	biblioteca	桌子	mesa
12	pizarra	课桌	pupitre
13	pantalla	钢笔	pluma estilográfica
14	estudiante	图书馆	biblioteca
15	universidad	修正带	cinta correctora
16	móvil	板擦	borrador
17		铅笔	lápiz
18		书桌	pupitre
19		圆珠笔	bolí(grafo)
20		空调	aire acondicionado
21		书包	mochila
22		教学楼	edificio docente
23		操场	patio de recreo
24		灯	lámpara
25		书	libro
26		文具盒	estuche escolar
27		笔记本	libreta de apuntes

Tabla 6.55. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La escuela: sus muebles y materiales en el tercer curso

08 La escuela: sus muebles y materiales			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	mesa	黑板	pizarra
2	pizarra	粉笔	tiza
3	silla	椅子	silla
4	libro	投影仪	proyector
5	bolí(grafo)	电脑	ordenador, computadora
6	lápiz	桌子	mesa
7	ordenador	笔	pluma
8	papel	板擦	borrador
9	aula	橡皮	goma
10	pluma	书	libro
11	profesor	多媒体	multimedia
12	borrador	灯	lámpara
13	pantalla	书桌	pupitre
14	biblioteca	铅笔	lápiz
15	tiza	尺子	regla
16	computadora	教学楼	edificio docente
17		本子	cuaderno, libreta
18		讲台	tarima
19		纸	papel
20		钢笔	pluma estilográfica
21		课桌	pupitre
22		操场	patio de recreo
23		圆珠笔	bolí(grafo)
24		食堂	comedor
25		教室	aula
26		屏幕	pantalla
27		电视	televisor
28		宿舍	residencia universitaria

Tabla 6.56. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La escuela: sus muebles y materiales en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

mesa; libro; silla; biblioteca; lápiz; pluma; clase; papel; estudiante; bolí(grafo); aula; ordenador.

Segundo curso:

mesa; libro; silla; papel; pluma; lápiz; ordenador; bolí(grafo).

Tercer curso:

mesa; silla; ordenador; libro; lápiz; bolí(grafo); papel; aula; clase; biblioteca; pizarra.

Cuarto curso:

mesa; pizarra; silla; libro; bolí(grafo); lápiz; ordenador; papel; aula; pluma;

borrador; pantalla; tiza; computadora.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

profesor; escuela; universidad; bachillerato.

Segundo curso:

profesor; aula; biblioteca; estudiante; escuela; universidad; edificio; clase; comedor.

Tercer curso:

profesor; pantalla; estudiante; universidad; móvil.

Cuarto curso:

profesor; biblioteca.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

pizarra; tiza; proyector; goma; cuaderno, libreta; pupitre; multimedia; pluma estilográfica; tarima; regla; edificio docente; borrador; patio de recreo; comedor; residencia universitaria; gimnasio, estadio.

Segundo curso:

pizarra; proyector; tiza; lámpara; goma; pluma estilográfica; multimedia; regla; borrador; edificio docente; pupitre; cuaderno, libreta; tarima; televisor; mochila.

Tercer curso:

tiza; proyector; goma; multimedia; pupitre; pluma estilográfica; cinta correctora; borrador; aire acondicionado; mochila; edificio docente; patio de recreo; lámpara; estuche escolar; libreta de apuntes.

Cuarto curso:

proyector; goma; multimedia; lámpara; pupitre; regla; edificio docente; cuaderno, libreta; tarima; pluma estilográfica; patio de recreo; comedor; pantalla; televisor; residencia universitaria.

6.2.9. Iluminación y calefacción

09 Iluminación y calefacción				
	Primero	Segundo	Tercero	Cuarto
1	lámpara	luz	luz	lámpara
2	sol	lámpara	lámpara	luz
3	luz	calefacción	calefacción	calefacción
4	calefacción	sol	sol	bombilla
5	frío	fuego	fuego	sol
6	calor	calentador	aire acondicionado	fuego
7	agua	aire acondicionado	gas	aire acondicionado
8		móvil	electricidad	calor
9			móvil	móvil
10			agua caliente	
11			agua	
12			carbón	

Tabla 6.57. Nociones con índice de disponibilidad superior a 0.1 en ELE en Iluminación y calefacción en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

lámpara; sol; luz; calefacción.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

fuego; calentador; aire acondicionado; móvil; gas; electricidad; agua caliente; carbón; bombilla.

Segundo curso:

frío; calor; agua; gas; electricidad; agua caliente; carbón; bombilla.

Tercer curso:

frío; calor; agua; calentador; bombilla.

Cuarto curso:

frío; agua; calentador; gas; electricidad; agua caliente; carbón.

09 Iluminación y calefacción			
Primer curso			
	ELE	CLM	Homólogo en ELE
1	lámpara	暖气	calefacción
2	sol	灯	lámpara, luz
3	luz	手电筒	linterna
4	calefacción	台灯	lámpara de mesa
5	frío	空调	aire acondicionado
6	calor	暖宝宝	parche térmico
7	agua	太阳	sol
8		地暖	suelo radiante
9		暖气片	radiador
10		电热毯	manta eléctrica
11		电	electricidad
12		手机	móvil
13		火炉	estufa
14		热水袋	bolsa de agua caliente
15		蜡烛	vela, candela
16		火	fuego
17		灯泡	bombilla

Tabla 6.58. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Iluminación y calefacción en el primer curso

09 Iluminación y calefacción Segundo curso			
	ELE	CLM	Homólogo en ELE
1	luz	暖气	calefacción
2	lámpara	台灯	lámpara de mesa
3	calefacción	手电筒	linterna
4	sol	空调	aire acondicionado
5	fuego	灯	luz, lámpara
6	calentador	吊灯	lámpara colgante
7	aire acondicionado	地暖	suelo radiante
8	móvil	蜡烛	vela, candela
9		电热毯	manta eléctrica
10		LED	LED
11		火炉	estufa
12		太阳	sol
13		热水袋	bolsa de agua caliente
14		电暖器	calefactor
15		暖宝宝	parche térmico
16		灯泡	bombilla
17		白炽灯	lámpara incandescente
18		灯管	tubo fluorescente
19		壁炉	hogar, fogón
20		日光灯	luminaria fluorescente
21		路灯	alumbrado público

Tabla 6.59. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Iluminación y calefacción en el segundo curso

09 Iluminación y calefacción Tercer curso			
	ELE	CLM	Homólogo en ELE
1	luz	暖气	calefacción
2	lámpara	灯	luz, lámpara
3	calefacción	台灯	lámpara de mesa
4	sol	空调	aire acondicionado
5	fuego	手电筒	linterna
6	aire acondicionado	暖气片	radiador
7	gas	吊灯	lámpara colgante
8	electricidad	光	luz
9	móvil	电	electricidad
10	agua caliente	太阳	sol
11	agua	蜡烛	vela, candela
12	carbón	小太阳	calefactor
13		暖宝宝	parche térmico
14		火炉	estufa
15		电暖器	calefactor
16		白炽灯	lámpara incandescente
17		地暖	suelo radiante
18		阳光	luz solar
19		灯泡	bombilla
20		太阳能	energía solar
21		日光灯	luminaria fluorescente
22		手机	móvil
23		壁炉	hogar, fogón

Tabla 6.60. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Iluminación y calefacción en el tercer curso

09 Iluminación y calefacción			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	lámpara	暖气	calefacción
2	luz	灯	lámpara, luz
3	calefacción	台灯	lámpara de mesa
4	bombilla	空调	aire acondicionado
5	sol	手电筒	linterna
6	fuego	灯泡	bombilla
7	aire acondicionado	暖气片	radiador
8	calor	白炽灯	lámpara incandescente
9	móvil	蜡烛	vela, candela
10		电	electricidad
11		地暖	suelo radiante
12		电暖器	calefactor
13		吊灯	lámpara colgante
14		太阳	sol
15		浴霸	bombilla de calentamiento para baño
16		LED	LED
17		热水袋	bolsa de agua caliente
18		火炉	estufa
19		暖手宝	calientamanos
20		火	fuego
21		电热毯	manta eléctrica

Tabla 6.61. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Iluminación y calefacción en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

lámpara; sol; luz; calefacción.

Segundo curso:

luz; lámpara; calefacción; sol; aire acondicionado.

Tercer curso:

luz; lámpara; calefacción; sol; aire acondicionado; electricidad; móvil.

Cuarto curso:

lámpara; luz; calefacción; bombilla; sol; fuego; aire acondicionado.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

frío; calor; agua.

Segundo curso:

fuego; calentador; móvil.

Tercer curso:

fuego; gas; agua caliente; agua; carbón.

Cuarto curso:

calor; móvil.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

linterna; lámpara de mesa; aire acondicionado; parche térmico; suelo radiante; radiador; manta eléctrica; electricidad; móvil; estufa; bolsa de agua caliente; vela, candela; fuego; bombilla.

Segundo curso:

lámpara de mesa; linterna; lámpara colgante; suelo radiante; vela, candela; manta eléctrica; LED; estufa; bolsa de agua caliente; calefactor; parche térmico; bombilla; lámpara incandescente; tubo fluorescente; hogar, fogón; luminaria fluorescente; farola, alumbrado público.

Tercer curso:

lámpara de mesa; linterna; radiador; lámpara colgante; vela, candela; calefactor; parche térmico; estufa; lámpara incandescente; suelo radiante; luz solar; bombilla; energía solar; luminaria fluorescente; hogar, fogón.

Cuarto curso:

lámpara de mesa; linterna; radiador; lámpara incandescente; vela, candela; electricidad; suelo radiante; calefactor; lámpara colgante; bombilla de calentamiento para baño; LED; bolsa de agua caliente; estufa; calentamanos; manta eléctrica.

6.2.10. La ciudad

10 La ciudad				
	Primero	Segundo	Tercero	Cuarto
1	ciudad	Madrid	coche	edificio
2	persona	Barcelona	edificio	coche
3	coche	coche	autobús	autobús
4	universidad	edificio	calle	calle
5	autobús	Beijing	metro	parque
6	Beijing	Buenos Aires	escuela	hospital
7	Lima	universidad	hospital	metro
8	Madrid	autobús	universidad	tren
9	Tianjin	París	persona	escuela
10	hospital	Ciudad de México	parque	Madrid
11	grande	Tianjin	Madrid	árbol
12	calle	Sevilla	policía	universidad
13	metro	calle	ciudad	bici(cleta)
14	casa	Nueva York	casa	Barcelona
15	colegio	escuela	gobierno	persona
16	escuela	casa	gente	contaminación
17	edificio	tienda	Barcelona	ayuntamiento
18	gente	árbol	tráfico	
19	centro	Valencia	Beijing	
20		hospital	tienda	

Tabla 6.62. Nociones con índice de disponibilidad superior a 0.1 en ELE en La ciudad en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

coche; universidad; autobús; Madrid; hospital; calle; escuela; edificio.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

Barcelona; Buenos Aires; París; Ciudad de México; Sevilla; Nueva York; tienda; árbol; Valencia; parque; policía; gobierno; tráfico; tren; bici(cleta); contaminación; ayuntamiento.

Segundo curso:

ciudad; persona; Lima; grande; metro; colegio; gente; centro; parque; policía; gobierno; tráfico; tren; bici(cleta); contaminación; ayuntamiento.

Tercer curso:

Lima; Tianjin; grande; colegio; centro; Buenos Aires; París; Ciudad de México; Sevilla; Nueva York; árbol; Valencia; tren; bici(cleta); contaminación;

ayuntamiento.

Cuarto curso:

ciudad; Beijing; Lima; Tianjin; grande; casa; colegio; gente; centro; Buenos Aires; París; Ciudad de México; Sevilla; Nueva York; tienda; Valencia; policía; gobierno; tráfico.

10 La ciudad			
	ELE	CLM	Homólogo en ELE
1	ciudad	北京	Beijing
2	persona	天津	Tianjin
3	coche	上海	Shanghai (Shanghái)
4	universidad	楼房	edificio
5	autobús	地铁	metro
6	Beijing	公交车	autobús
7	Lima	学校	escuela
8	Madrid	广州	Guangzhou (Cantón)
9	Tianjin	公园	parque
10	hospital	汽车	coche
11	grande	商场	centro comercial
12	calle	街道	calle
13	metro	马德里	Madrid
14	casa		
15	colegio		
16	escuela		
17	edificio		
18	gente		
19	centro		

Tabla 6.63. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La ciudad en el primer curso

10 La ciudad Segundo curso			
	ELE	CLM	Homólogo en ELE
1	Madrid	北京	Beijing
2	Barcelona	上海	Shanghai (Shanghái)
3	coche	天津	Tianjin
4	edificio	马德里	Madrid
5	Beijing	公园	parque
6	Buenos Aires	纽约	Nueva York
7	universidad	学校	escuela
8	autobús	医院	hospital
9	París	东京	Tokio
10	Ciudad de México	地铁	metro
11	Tianjin	伦敦	Londres
12	Sevilla	公交车	autobús
13	calle	巴塞罗那	Barcelona
14	Nueva York	街道	calle
15	escuela	交通	tráfico
16	casa	摩天大楼	rascacielos
17	tienda	商场	centro comercial
18	árbol		
19	Valencia		
20	hospital		

Tabla 6.64. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La ciudad en el segundo curso

10 La ciudad Tercer curso			
	ELE	CLM	Homólogo en ELE
1	coche	学校	escuela
2	edificio	地铁	metro
3	autobús	医院	hospital
4	calle	交通	tráfico
5	metro	街道	calle
6	escuela	汽车	coche
7	hospital	建筑	arquitectura, construcción
8	universidad	公交车	autobús
9	persona	北京	Beijing
10	parque	公园	parque
11	Madrid	上海	Shanghai (Shanghái)
12	policía	马路	carretera
13	ciudad	天津	Tianjin
14	casa	楼房	edificio
15	gobierno	火车	tren
16	gente	车	vehículo, automóvil
17	Barcelona	绿化	zona verde
18	tráfico		
19	Beijing		
20	tienda		

Tabla 6.65. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La ciudad en el tercer curso

10 La ciudad Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	edificio	地铁	metro
2	coche	北京	Beijing (Pekín)
3	autobús	公交车	autobús
4	calle	汽车	coche
5	parque	交通	tráfico
6	hospital	上海	Shanghai (Shanghái)
7	metro	医院	hospital
8	tren	学校	escuela
9	escuela	街道	calle
10	Madrid	公园	parque
11	árbol	火车站	estación de tren
12	universidad	商场	centro comercial
13	bici(cleta)	马路	carretera
14	Barcelona	拥堵	atasco, embotellamiento, congestión
15	persona	楼房	edificio
16	contaminación	天津	Tianjin
17	ayuntamiento	马德里	Madrid
18		污染	contaminación
19		市政府	ayuntamiento
20		超市	supermercado
21		广场	plaza

Tabla 6.66. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en La ciudad en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

coche; autobús; Beijing; Madrid; Tianjin; calle; metro; escuela; edificio.

Segundo curso:

Madrid; Barcelona; Beijing; autobús; Tianjin; calle; Nueva York; escuela; hospital.

Tercer curso:

coche; edificio; autobús; calle; metro; escuela; hospital; parque; tráfico; Beijing.

Cuarto curso:

edificio; coche; autobús; calle; parque; hospital; metro; escuela; Madrid; contaminación; ayuntamiento.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

ciudad; persona; universidad; Lima; hospital; grande; casa; colegio; gente; centro.

Segundo curso:

coche; edificio; Buenos Aires; universidad; París; Ciudad de México; Sevilla; casa; tienda; árbol; Valencia.

Tercer curso:

universidad; persona; Madrid; policía; ciudad; casa; gobierno; gente; Barcelona; tienda.

Cuarto curso:

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

Shanghai (Shanghái); Guangzhou (Cantón); parque; centro comercial.

Segundo curso:

Shanghai (Shanghái); parque; Tokio; metro; Londres; tráfico; rascacielos; centro comercial.

Tercer curso:

arquitectura, construcción; Shanghai (Shanghái); carretera; Tianjin; tren; vehículo, automóvil; zona verde.

Cuarto curso:

Beijing (Pekín); tráfico; Shanghai (Shanghái); estación de tren; centro comercial; carretera; atasco, embotellamiento, congestión; Tianjin; supermercado; plaza.

6.2.11. El campo

11 El campo				
	Primero	Segundo	Tercero	Cuarto
1	flor	río	campesino	campo
2	campo	árbol	campo	árbol
3	perro	campesino	tierra	campesino
4	pequeño	flor	río	tierra
5	casa	casa	árbol	río
6	verdura	tierra	montaña	flor
7	río	montaña	flor	animal
8	árbol	campo	animal	montaña
9	aire	animal	perro	aire
10		arroz	casa	perro
11		pueblo	planta	cerdo
12		perro	arroz	cielo
13		cultivar	lago	pobre
14		lago	cerdo	maíz
15			agricultura	
16			pobre	

Tabla 6.67. Nociones con índice de disponibilidad superior a 0.1 en ELE en El campo en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

flor; campo; perro; río; árbol.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

campesino; tierra; montaña; animal; arroz; pueblo; cultivar; lago; planta; cerdo; agricultura; pobre; cielo; maíz.

Segundo curso:

pequeño; verdura; aire; planta; cerdo; agricultura; pobre; cielo; maíz.

Tercer curso:

pequeño; verdura; aire; pueblo; cultivar; cielo; maíz.

Cuarto curso:

pequeño; casa; verdura; arroz; pueblo; cultivar; lago; planta; agricultura.

11 El campo Primer curso			
	ELE	CLM	Homólogo en ELE
1	flor	田	campo
2	campo	土地	tierra, terreno
3	perro	树	árbol
4	pequeño	农民	campesino
5	casa	河	río
6	verdura	田野	campo
7	río	小溪	arroyo
8	árbol	山	montaña, monte, colina
9	aire	草	hierba, yerba, pasto
10		平房	casa
11		羊	cabra, oveja
12		牛	toro, buey, vaca
13		鸡	gallo, gallina, pollo
14		猪	cerdo
15		花	flor
16		狗	perro

Tabla 6.68. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en El campo en el primer curso

11 El campo Segundo curso			
	ELE	CLM	Homólogo en ELE
1	río	农民	campesino
2	árbol	土地	tierra
3	campesino	河	río
4	flor	田野	campo
5	casa	田	campo
6	tierra	山	montaña
7	montaña	花	flor
8	campo	树	árbol
9	animal	牛	toro, buey, vaca
10	arroz	猪	cerdo
11	pueblo	水稻	arroz
12	perro	小麦	trigo
13	cultivar	狗	perro
14	lago	羊	cabra, oveja
15		鸡	gallo, gallina, pollo
16		鸭	pato
17		小溪	arroyo
18		小路	senda, sendero
19		草	hierba, yerba, pasto
20		平房	casa
21		玉米	maíz

Tabla 6.69. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en El campo en el segundo curso

11 El campo Tercer curso			
	ELE	CLM	Homólogo en ELE
1	campesino	山	montaña
2	campo	土地	tierra
3	tierra	田	campo
4	río	牛	toro, buey, vaca
5	árbol	猪	cerdo
6	montaña	羊	cabra, oveja
7	flor	树	árbol
8	animal	鸡	gallo, gallina, pollo
9	perro	河	río
10	casa	农民	campesino
11	planta	田野	campo
12	arroz	狗	perro
13	lago	空气	aire
14	cerdo	鸭	pato
15	agricultura	泥土	tierra, suelo
16	pobre	小路	senda, sendero
17		果园	huerto, huerta
18		水	agua
19		农村	campo, zona rural
20		庄稼	cultivo
21		田园	campo
22		花	flor
23		牲畜	ganado, animal doméstico
24		湖	lago
25		房屋	casa
26		牧场	prado
27		森林	bosque, selva
28		放牧	pastar, pastorear

Tabla 6.70. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en El campo en el tercer curso

11 El campo Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	campo	田	campo
2	árbol	土地	tierra
3	campesino	农民	campesino
4	tierra	田野	campo
5	río	平房	casa
6	flor	牛	toro, buey, vaca
7	animal	狗	perro
8	montaña	玉米	maíz
9	aire	鸡	gallo, gallina. pollo
10	perro	山	montaña
11	cerdo	拖拉机	tractor
12	cielo	羊	cabra, oveja
13	pobre	花	flor
14	maíz	池塘	charco, charca, estanque
15		树	árbol
16		鸭	pato
17		小麦	trigo
18		水稻	arroz
19		庄稼	cultivo
20		小溪	arroyo
21		房屋	casa

Tabla 6.71. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en El campo en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

flor; campo; perro; casa; río; árbol.

Segundo curso:

río; árbol; campesino; flor; casa; tierra; montaña; campo; arroz; perro.

Tercer curso:

campesino; campo; tierra; río; árbol; montaña; flor; perro; casa; lago; cerdo.

Cuarto curso:

campo; árbol; campesino; tierra; flor; montaña; perro; maíz.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

pequeño; verdura; aire.

Segundo curso:

animal; pueblo; cultivar; lago.

Tercer curso:

animal; planta; arroz; agricultura; pobre.

Cuarto curso:

río; animal; aire; cerdo; cielo; pobre.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

tierra, terreno; campesino; arroyo; montaña, monte, colina; hierba, yerba, pasto; cabra, oveja; toro, buey, vaca; gallo, gallina, pollo; cerdo.

Segundo curso:

toro, buey, vaca; cerdo; trigo; cabra, oveja; gallo, gallina, pollo; pato; arroyo; senda, sendero; hierba, yerba, pasto; maíz.

Tercer curso:

toro, buey, vaca; cabra, oveja; gallo, gallina, pollo; aire; pato; tierra, suelo; senda, sendero; huerto, huerta; agua; campo, zona rural; cultivo; ganado, animal doméstico; prado; bosque, selva; pastar, pastorear.

Cuarto curso:

casa; toro, buey, vaca; gallo, gallina, pollo; tractor; cabra, oveja; charco, charca, estanque; pato; trigo; arroz; cultivo; arroyo; casa.

6.2.12. Medios de transporte

	12 Medios de transporte			
	Primero	Segundo	Tercero	Cuarto
1	autobús	coche	coche	autobús
2	coche	autobús	autobús	bici(cleta)
3	metro	tren	bici(cleta)	coche
4	tren	avión	metro	avión
5	bici(cleta)	bici(cleta)	tren	tren
6	a pie	metro	avión	metro
7	avión	a pie	a pie	barco
8	taxi	taxi	barco	moto
9	moto	barco	taxi	taxi
10	carro	carro	moto	a pie
11		moto	carro	vehículo
12			camión	autocar
13			aeropuerto	camión
14			caballo	

Tabla 6.72. Nociones con índice de disponibilidad superior a 0.1 en ELE en Medios de transporte en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

autobús; coche; metro; tren; bici(cleta); a pie; avión; taxi; moto.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

barco; camión; aeropuerto; caballo; vehículo; autocar.

Segundo curso:

camión; aeropuerto; caballo; vehículo; autocar.

Tercer curso:

vehículo; autocar.

Cuarto curso:

carro; aeropuerto; caballo.

12 Medios de transporte Primer curso			
	ELE	CLM	Homólogo en ELE
1	autobús	火车	tren
2	coche	飞机	avión
3	metro	自行车	bici(cleta)
4	tren	汽车	coche, carro
5	bici(cleta)	地铁	metro
6	a pie	公交车	autobús
7	avión	高铁	tren de alta velocidad
8	taxi	轮船	buque
9	moto	摩托车	moto
10	carro	电动车	automóvil eléctrico
11		出租车	taxi
12		三轮车	triciclo
13		轿车	sedán, berlina
14		船	barco, nave, embarcación
15		公共汽车	autobús
16		共享单车	bicicleta compartida
17		马车	coche de caballos

Tabla 6.73. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Medios de transporte en el primer curso

12 Medios de transporte Segundo curso			
	ELE	CLM	Homólogo en ELE
1	coche	飞机	avión
2	autobús	自行车	bici(cleta)
3	tren	火车	tren
4	avión	地铁	metro
5	bici(cleta)	汽车	coche, carro
6	metro	公交车	autobús
7	a pie	高铁	tren de alta velocidad
8	taxi	摩托车	moto
9	barco	出租车	taxi
10	carro	马车	coche de caballos
11	moto	轮船	buque
12		三轮车	triciclo
13		船	barco
14		轿车	sedán, berlina
15		滑板	monopatín, patineta
16		电动车	automóvil eléctrico
17		私家车	coche privado
18		轻轨	tren ligero

Tabla 6.74. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Medios de transporte en el segundo curso

12 Medios de transporte Tercer curso			
	ELE	CLM	Homólogo en ELE
1	coche	火车	tren
2	autobús	汽车	coche, carro
3	bici(cleta)	飞机	avión
4	metro	地铁	metro
5	tren	自行车	bici(cleta)
6	avión	公交车	autobús
7	a pie	高铁	tren de alta velocidad
8	barco	摩托车	moto
9	taxi	船	barco
10	moto	出租车	taxi
11	carro	公共汽车	autobús
12	camión	共享单车	bicicleta compartida
13	aeropuerto	私家车	coche privado
14	caballo	步行	a pie
15		三轮车	triciclo
16		大巴	autobús
17		轿车	sedán, berlina

Tabla 6.75. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Medios de transporte en el tercer curso

12 Medios de transporte Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	autobús	自行车	bici(cleta)
2	bici(cleta)	汽车	coche
3	coche	火车	tren
4	avión	飞机	avión
5	tren	公交车	autobús
6	metro	地铁	metro
7	barco	摩托车	moto
8	moto	高铁	tren de alta velocidad
9	taxi	出租车	taxi
10	a pie	电动车	automóvil eléctrico
11	vehículo	共享单车	bicicleta compartida
12	autocar	轮船	buque
13	camión	船	barco
14		轻轨	tren ligero
15		三轮车	triciclo
16		电瓶车	bicicleta eléctrica
17		货车	camión
18		步行	a pie

Tabla 6.76. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Medios de transporte en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

autobús; coche; metro; tren; bici(cleta); avión; taxi; moto; carro.

Segundo curso:

coche; autobús; tren; avión; bici(cleta); metro; taxi; barco; carro; moto.

Tercer curso:

coche; autobús; bici(cleta); metro; tren; avión; a pie; barco; taxi; moto; carro.

Cuarto curso:

autobús; bici(cleta); coche; avión; tren; metro; barco; moto; taxi; a pie; camión.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

a pie.

Segundo curso:

a pie.

Tercer curso:

camión; aeropuerto; caballo.

Cuarto curso:

vehículo; autocar.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

tren de alta velocidad; buque; automóvil eléctrico; triciclo; sedán, berlina; barco, nave, embarcación, bote; bicicleta compartida; coche de caballos.

Segundo curso:

tren de alta velocidad; coche de caballos; buque; triciclo; sedán, berlina; monopatín, patineta; automóvil eléctrico; coche privado; tren ligero.

Tercer curso:

tren de alta velocidad; bicicleta compartida; coche privado; triciclo; sedán, berlina.

Cuarto curso:

tren de alta velocidad; automóvil eléctrico; bicicleta compartida; buque; tren ligero; triciclo; bicicleta eléctrica.

6.2.13. Trabajos del campo y del jardín

13 Trabajos del campo y del jardín				
	Primero	Segundo	Tercero	Cuarto
1	flor	cultivar	árbol	tierra
2	agua	flor	campesino	flor
3	trabajar	árbol	agua	árbol
4	árbol	campesino	flor	campesino
5	tierra	agua	regar	cultivar
6	cansado	regar	cultivar	campo
7	jardín	arroz	tierra	regar
8	campo	tierra	campo	agua
9	trabajo	plantar	plantar	trabajar
10	verdura	campo	planta	jardín
11		jardín	jardín	cosecha
12		fruta	arroz	
13			hierba	
14			sol	
15			cortar	

Tabla 6.77. Nociones con índice de disponibilidad superior a 0.1 en ELE en Trabajos del campo y del jardín en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

flor; agua; árbol; tierra; jardín; campo.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

cultivar; campesino; regar; arroz; plantar; fruta; planta; hierba; sol; cortar; cosecha.

Segundo curso:

trabajar; cansado; trabajo; verdura; planta; hierba; sol; cortar; cosecha.

Tercer curso:

trabajar; cansado; trabajo; verdura; fruta; cosecha.

Cuarto curso:

cansado; trabajo; verdura; arroz; plantar; fruta; planta; hierba; sol; cortar.

13 Trabajos del campo y del jardín Primer curso			
	ELE	CLM	Homólogo en ELE
1	flor	浇水	regar, riego
2	agua	锄头	azada
3	trabajar	施肥	poner abono
4	árbol	播种	sembrar, siembra
5	tierra	农民	campesino
6	cansado	除草	arrancar hierbas, cortar hierbas
7	jardín	树	árbol
8	campo	收割	cosechar, cosecha
9	trabajo	修剪	recortar, podar
10	verdura	花	flor
11		松土	remover la tierra

Tabla 6.78. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Trabajos del campo y del jardín en el primer curso

13 Trabajos del campo y del jardín Segundo curso			
	ELE	CLM	Homólogo en ELE
1	cultivar	浇水	regar
2	flor	施肥	poner abono
3	árbol	播种	sembrar, siembra
4	campesino	除草	arrancar hierbas, cortar hierbas
5	agua	修剪	recortar, podar
6	regar	收割	cosechar, cosecha
7	arroz	翻土	remover la tierra
8	tierra	种树	plantar árboles
9	plantar	灌溉	regar
10	campo	除虫	matar insectos
11	jardín	插秧	trasplantar plántulas de arroz
12	fruta	种花	plantar flores
13		喷农药	tirar plaguicida
14		耕地	trabajar la tierra
15		松土	remover la tierra
16		种地	trabajar la tierra
17		锄草	arrancar hierbas, cortar hierbas

Tabla 6.79. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Trabajos del campo y del jardín en el segundo curso

13 Trabajos del campo y del jardín Tercer curso			
	ELE	CLM	Homólogo en ELE
1	árbol	播种	sembrar, siembra
2	campesino	浇水	regar
3	agua	施肥	poner abono
4	flor	修剪	cortar
5	regar	收割	cosechar, cosecha
6	cultivar	除草	arrancar hierbas, cortar hierbas
7	tierra	树	árbol
8	campo	种菜	plantar hortalizas, plantar verduras
9	plantar	翻土	remover la tierra
10	planta	花园	jardín
11	jardín	锄头	azada
12	arroz	农民	campesino
13	hierba	耕地	trabajar la tierra
14	sol	插秧	trasplantar plántulas de arroz
15	cortar	松土	remover la tierra
16		园丁	jardinero
17		花	flor
18		喷农药	tirar plaguicida

Tabla 6.80. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Trabajos del campo y del jardín en el tercer curso

13 Trabajos del campo y del jardín			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	tierra	浇水	regar
2	flor	锄头	azada
3	árbol	播种	sembrar, siembra
4	campesino	施肥	poner abono
5	cultivar	农药	plaguicida
6	campo	犁	arado
7	regar	插秧	trasplantar plántulas de arroz
8	agua	犁地	arar la tierra
9	trabajar	修剪	recortar, podar
10	jardín	耕地	trabajar la tierra
11	cosecha	剪刀	tijera
12		农民	campesino
13		除草	arrancar hierbas, cortar hierbas
14		喷农药	tirar plaguicida
15		种子	semilla
16		牛	toro, buey, vaca
17		灌溉	regar
18		树	árbol
19		采摘	recoger, recogida
20		收割	cosecha
21		花	flor
22		拖拉机	tractor

Tabla 6.81. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Trabajos del campo y del jardín en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

flor; árbol.

Segundo curso:

regar.

Tercer curso:

árbol; campesino; flor; regar; jardín; cortar.

Cuarto curso:

flor; árbol; campesino; regar; trabajar; cosecha.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

agua; trabajar; tierra; cansado; jardín; campo; trabajo; verdura.

Segundo curso:

cultivar; flor; árbol; campesino; agua; arroz; tierra; plantar; campo; jardín; fruta.

Tercer curso:

agua; cultivar; tierra; campo; plantar; planta; arroz; hierba; sol.

Cuarto curso:

tierra; cultivar; campo; agua; jardín.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

regar, riego; azada; poner abono; sembrar, siembra; campesino; arrancar hierbas, cortar hierbas; cosechar, cosecha; recortar, podar; remover la tierra.

Segundo curso:

poner abono; sembrar, siembra; arrancar hierbas, cortar hierbas; recortar, podar; cosechar, cosecha; remover la tierra; plantar árboles; matar insectos; trasplantar plántulas de arroz; plantar flores; tirar plaguicida; trabajar la tierra.

Tercer curso:

sembrar, siembra; poner abono; cosechar, cosecha; arrancar hierbas, cortar hierbas; plantar hortalizas, plantar verduras; remover la tierra; azada; trabajar la tierra; trasplantar plántulas de arroz; jardinero; tirar plaguicida.

Cuarto curso:

azada; sembrar, siembra; poner abono; plaguicida; arado; trasplantar plántulas de arroz; arar la tierra; recortar, podar; trabajar la tierra; tijera; arrancar hierbas, cortar hierbas; tirar plaguicida; semilla; toro, buey, vaca; recoger, recogida; tractor.

6.2.14. Los animales

14 Los animales				
	Primero	Segundo	Tercero	Cuarto
1	perro	perro	perro	perro
2	gato	gato	gato	gato
3	pez	tigre	tigre	tigre
4	tigre	león	león	león
5	caballo	mono	mono	cerdo
6	cerdo	pez	pez	conejo
7	ratón	oso	cerdo	mono
8	pollo	ratón	serpiente	pez
9	dragón	zorro	pájaro	caballo
10	buey	caballo	caballo	elefante
11	gallo	pescado	conejo	serpiente
12	león	pollo	vaca	zorro
13		cuervo	pato	vaca
14		cerdo	gallo	panda
15		pájaro	pollo	lobo
16		lobo	gallina	gallo
17		elefante	pescado	ratón
18		toro	elefante	dragón
19		panda	oso	pájaro
20		vaca	toro	oveja
21			ave	lince
22			panda	ciervo
23			oveja	gallina
24			zorro	cabra
25				toro
26				jirafa
27				oso
28				ave
29				rana
30				pato

Tabla 6.82. Nociones con índice de disponibilidad superior a 0.1 en ELE en Los animales en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

perro; gato; pez; tigre; caballo; cerdo; león.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

mono; oso; zorro; pescado; cuervo; pájaro; lobo; elefante; toro; panda; vaca; serpiente; conejo; pato; gallina; ave; oveja; lince; ciervo; cabra; jirafa; rana.

Segundo curso:

dragón; buey; gallo; serpiente; conejo; pato; gallina; ave; oveja; lince; ciervo; cabra; jirafa; rana.

Tercer curso:

ratón; dragón; buey; cuervo; lobo; lince; ciervo; cabra; jirafa; rana.

Cuarto curso:

pollo; buey; pescado; cuervo.

14 Los animales Primer curso			
	ELE	CLM	Homólogo en ELE
1	perro	狗	perro
2	gato	猫	gato
3	pez	老虎	tigre
4	tigre	牛	buey
5	caballo	兔子	conejo, liebre
6	cerdo	羊	cabra, oveja
7	ratón	鸡	pollo, gallo
8	pollo	猪	cerdo
9	dragón	马	caballo
10	buey	老鼠	ratón
11	gallo	蛇	serpiente
12	león	狮子	león
13		鸭	pato
14		猴子	mono
15		鱼	pez
16		大象	elefante
17		熊猫	panda, oso panda
18		鸟	pájaro, ave
19		熊	oso
20		龙	dragón
21		狼	lobo
22		企鹅	pingüino

Tabla 6.83. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los animales en el primer curso

14 Los animales Segundo curso			
	ELE	CLM	Homólogo en ELE
1	perro	狗	perro
2	gato	猫	gato
3	tigre	老虎	tigre
4	león	牛	vaca
5	mono	猪	cerdo
6	pez	兔子	conejo, liebre
7	oso	羊	cabra, oveja
8	ratón	鸡	pollo
9	zorro	老鼠	ratón
10	caballo	蛇	serpiente
11	pescado	马	caballo
12	pollo	狮子	león
13	cuervo	猴子	mono
14	cerdo	鸭	pato
15	pájaro	鱼	pez, pescado
16	lobo	熊	oso
17	elefante	龙	dragón
18	toro	鸟	pájaro
19	panda	大象	elefante
20	vaca	豹	leopardo
21		熊猫	panda
22		驴	asno, burro
23		狼	lobo
24		乌龟	tortuga
25		海豚	delfín
26		长颈鹿	jirafa
27		狐狸	zorro
28		鹅	ganso

Tabla 6.84. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los animales en el segundo curso

14 Los animales Tercer curso			
	ELE	CLM	Homólogo en ELE
1	perro	狗	perro
2	gato	猫	gato
3	tigre	猪	cerdo
4	león	牛	vaca, toro
5	mono	羊	oveja
6	pez	老虎	tigre
7	cerdo	鸡	gallo, pollo, gallina
8	serpiente	蛇	serpiente
9	pájaro	兔子	conejo
10	caballo	老鼠	ratón, rata
11	conejo	马	caballo
12	vaca	鱼	pez, pescado
13	pato	猴子	mono
14	gallo	鸭	pato
15	pollo	狮子	león
16	gallina	龙	dragón
17	pescado	熊猫	panda
18	elefante	大象	elefante
19	oso	熊	oso
20	toro	鹅	ganso
21	ave	狐狸	zorro
22	panda	狼	lobo
23	oveja	鲸鱼	ballena
24	zorro	长颈鹿	jirafa

Tabla 6.85. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los animales en el tercer curso

14 Los animales Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	perro	狗	perro
2	gato	猫	gato
3	tigre	牛	vaca, toro
4	león	羊	oveja, cabra
5	cerdo	猪	cerdo
6	conejo	马	caballo
7	mono	兔子	conejo
8	pez	老虎	tigre
9	caballo	鸡	gallo, gallina
10	elefante	蛇	serpiente
11	serpiente	猴子	mono
12	zorro	老鼠	ratón
13	vaca	鱼	pez
14	panda	熊	oso
15	lobo	狮子	león
16	gallo	大象	elefante
17	ratón	熊猫	panda
18	dragón	鸭	pato
19	pájaro	龙	dragón
20	oveja	狼	lobo
21	lince	河马	hipopótamo
22	ciervo	狐狸	zorro
23	gallina	海豚	delfín
24	cabra	猩猩	chimpancé, gorila
25	toro	鸟	pájaro, ave
26	jirafa	鲸鱼	ballena
27	oso	企鹅	pingüino
28	ave	长颈鹿	jirafa
29	rana		
30	pato		

Tabla 6.86. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Los animales en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

perro; gato; pez; tigre; caballo; cerdo; ratón; pollo; dragón; buey; gallo; león.

Segundo curso:

perro; gato; tigre; león; mono; pez; oso; ratón; zorro; caballo; pescado; pollo; cerdo; pájaro; lobo; elefante; toro; panda; vaca.

Tercer curso:

perro; gato; tigre; león; mono; pez; cerdo; serpiente; caballo; conejo; vaca; pato; gallo; pollo; gallina; pescado; elefante; oso; toro; panda; oveja; zorro.

Cuarto curso:

perro; gato; tigre; león; cerdo; conejo; mono; pez; caballo; elefante; serpiente; zorro; vaca; panda; lobo; gallo; ratón; dragón; pájaro; oveja; gallina; cabra; toro; jirafa; oso; ave; pato.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

-

Segundo curso:

cuervo.

Tercer curso:

pájaro; ave.

Cuarto curso:

lince; ciervo; rana.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

conejo, liebre; cabra, oveja; serpiente; pato; mono; elefante; panda, oso panda; pájaro, ave; oso; lobo; pingüino.

Segundo curso:

conejo, liebre; cabra, oveja; serpiente; pato; dragón; elefante; leopardo; asno, burro; tortuga; delfín; jirafa; ganso.

Tercer curso:

ratón, rata; dragón; ganso; lobo; ballena; jirafa.

Cuarto curso:

hipopótamo; delfín; chimpancé, gorila; ballena; pingüino.

6.2.15. Juegos y distracciones

15 Juegos y distracciones				
	Primero	Segundo	Tercero	Cuarto
1	deporte	fútbol	correr	fútbol
2	jugar	baloncesto	fútbol	baloncesto
3	fútbol	correr	nadar	correr
4	móvil	nadar	videojuego	nadar
5	juego	tenis	deporte	tenis
6	tenis	voleibol	juego	bádminton
7	tele(visión)	jugar	carta	ping-pong
8	correr	deporte	baloncesto	videojuego
9	cantar	bailar	voleibol	voleibol
10	contento	cantar	tenis	carta
11		cine	gimnasio	cantar
12		móvil	saltar	natación
13		juego	ping-pong	bailar
14		pasear	pasear	juego
15		saltar	bailar	gimnasio
16			cantar	discoteca
17			película	
18			ordenador	
19			jugar	
20			bádminton	

Tabla 6.87. Nociones con Índice de disponibilidad superior a 0.1 en ELE en Juegos y distracciones en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

fútbol; juego; tenis; correr; cantar.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

baloncesto; nadar; voleibol; bailar; cine; pasear; saltar; videojuego; carta; gimnasio; ping-pong; película; ordenador; bádminton; natación; discoteca.

Segundo curso:

jugar; tele(visión); contento; videojuego; carta; gimnasio; ping-pong; película; ordenador; bádminton; natación; discoteca.

Tercer curso:

móvil; tele(visión); contento; cine; natación; discoteca.

Cuarto curso:

deporte; jugar; móvil; tele(visión); contento; cine; pasear; saltar; película;

ordenador.

15 Juegos y distracciones Primer curso			
	ELE	CLM	Homólogo en ELE
1	deporte	跑步	correr
2	jugar	篮球	baloncesto
3	fútbol	羽毛球	bádminton
4	móvil	网球	tenis
5	juego	游泳	nadar, natación
6	tenis	足球	fútbol
7	tele(visión)	王者荣耀	Rey de Gloria (videojuego)
8	correr	排球	voleibol
9	cantar	跳绳	salto a la comba
10	contento	乒乓球	ping-pong, tenis de mesa
11		跳舞	bailar
12		唱歌	cantar
13		电影	película, cine
14		台球	billar
15		散步	pasear, paseo
16		纸牌	cartas, naipes
17		瑜伽	yoga

Tabla 6.88. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Juegos y distracciones en el primer curso

15 Juegos y distracciones Segundo curso			
	ELE	CLM	Homólogo en ELE
1	fútbol	篮球	baloncesto
2	baloncesto	足球	fútbol
3	correr	游泳	nadar
4	nadar	网球	tenis
5	tenis	乒乓球	ping-pong, tenis de mesa
6	voleibol	跑步	correr
7	jugar	羽毛球	bádminton
8	deporte	排球	voleibol
9	bailar	击剑	esgrima
10	cantar	纸牌	cartas, naipes
11	cine	跳绳	salto a la comba
12	móvil	跳舞	bailar
13	juego	台球	billar
14	pasear	麻将	mah-jong
15	saltar	电影	cine
16		垒球	sóftbol
17		王者荣耀	Rey de Gloria (videojuego)
18		唱歌	cantar
19		逛街	ir de compras
20		滑冰	patinar, patinaje
21		电子游戏	videojuego
22		体操	gimnasia

Tabla 6.89. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Juegos y distracciones en el segundo curso

15 Juegos y distracciones			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	correr	羽毛球	bádminton
2	fútbol	跑步	correr
3	nadar	足球	fútbol
4	videojuego	游泳	nadar
5	deporte	乒乓球	ping-pong
6	juego	篮球	baloncesto
7	carta	排球	voleibol
8	baloncesto	网球	tenis
9	voleibol	王者荣耀	Rey de Gloria (videojuego)
10	tenis	跳绳	salto a la comba
11	gimnasio	跳高	salto de altura
12	saltar	健身	gimnasio
13	ping-pong	瑜伽	yoga
14	pasear	跳远	salto de longitud, salto largo
15	bailar	击剑	esgrima
16	cantar	跳舞	bailar
17	película	电影	película
18	ordenador	电子竞技	deporte electrónico
19	jugar	电脑	ordenador
20	bádminton	太极拳	taichí
21		纸牌	carta
22		游戏	juego, jugar

Tabla 6.90. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Juegos y distracciones en el tercer curso

15 Juegos y distracciones			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	fútbol	游泳	nadar, natación
2	baloncesto	跑步	correr
3	correr	篮球	baloncesto
4	nadar	羽毛球	bádminton
5	tenis	足球	fútbol
6	bádminton	乒乓球	ping-pong
7	ping-pong	排球	voleibol
8	videojuego	麻将	mah-jong
9	voleibol	跳舞	bailar
10	carta	电影	película, cine
11	cantar	网球	tenis
12	natación	瑜伽	yoga
13	bailar	扑克牌	baraja francesa
14	juego	唱歌	cantar
15	gimnasio	王者荣耀	Rey de Gloria (videojuego)
16	discoteca	KTV	KTV
17		跳绳	salto a la comba
18		电子游戏	videojuego
19		蹦极	puenting
20		纸牌	carta

Tabla 6.91. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Juegos y distracciones en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

fútbol; tenis; correr; cantar.

Segundo curso:

fútbol; baloncesto; correr; nadar; tenis; voleibol; bailar; cantar; cine.

Tercer curso:

correr; fútbol; nadar; juego; carta; baloncesto; voleibol; tenis; gimnasio; ping-pong; bailar; película; ordenador; jugar; bádminton.

Cuarto curso:

fútbol; baloncesto; correr; nadar; tenis; bádminton; ping-pong; videojuego; voleibol; carta; cantar; natación; bailar.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

deporte; jugar; móvil; juego; tele(visión); contento.

Segundo curso:

jugar; deporte; móvil; juego; pasear; saltar.

Tercer curso:

videojuego; deporte; saltar; pasear; cantar.

Cuarto curso:

juego; gimnasio; discoteca.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

baloncesto; bádminton; nadar, natación; Rey de Gloria (videojuego); voleibol; salto a la comba; ping-pong, tenis de mesa; bailar; película, cine; billar; pasear, paseo; cartas, naipes; yoga.

Segundo curso:

ping-pong, tenis de mesa; bádminton; esgrima; cartas, naipes; salto a la comba; billar; mah-jong; sóftbol; Rey de Gloria (videojuego); ir de compras; patinar, patinaje; videojuego; gimnasia.

Tercer curso:

Rey de Gloria (videojuego); salto a la comba; salto de altura; yoga; salto de longitud, salto largo; esgrima; deporte electrónico; taichí.

Cuarto curso:

mah-jong; película, cine; yoga; baraja francesa; Rey de Gloria (videojuego); KTV; salto a la comba; puenting.

6.2.16. Profesiones y oficios

16 Profesiones y oficios				
	Primero	Segundo	Tercero	Cuarto
1	profesor	profesor	profesor	profesor
2	estudiante	médico	estudiante	estudiante
3	médico	estudiante	médico	médico
4	camarero	policía	policía	cocinero
5	alumno	enfermero	doctor	abogado
6	policía	camarero	trabajador	alumno
7	enfermero	alumno	actor	enfermero
8	bibliotecario	obrero	enfermero	policía
9	conductor	bombero	camarero	doctor
10	secretario	abogado	cantante	conductor
11	taxista	cantante	abogado	cantante
12		jefe	obrero	actor
13		actor	campesino	secretario
14		secretario	cocinero	presidente
15		campesino	jefe	jefe
16		gerente	actriz	ingeniero
17			gerente	trabajador
18			vendedor	campesino
19			secretario	maestro
20			alumno	director
21			traductor	bailarín
22				escritor

Tabla 6.92. Nociones con índice de disponibilidad superior a 0.1 en ELE en Profesiones y oficios en los cuatro cursos universitarios

a1) Vocablos en ELE coincidentes en los cuatro cursos:

profesor; estudiante; médico; alumno; policía; enfermero; secretario.

a2) Vocablos en ELE ausentes en cada curso (en comparación con los demás cursos):

Primer curso:

obrero; bombero; abogado; cantante; jefe; actor; campesino; gerente; doctor; trabajador; cocinero; actriz; vendedor; traductor; presidente; ingeniero; maestro; director; bailarín; escritor.

Segundo curso:

bibliotecario; conductor; taxista; doctor; trabajador; cocinero; actriz; vendedor; traductor; presidente; ingeniero; maestro; director; bailarín; escritor.

Tercer curso:

bibliotecario; conductor; taxista; bombero; presidente; ingeniero; maestro;

director; bailarín; escritor.

Cuarto curso:

camarero; bibliotecario; taxista; obrero; bombero; gerente; actriz; vendedor; traductor.

16 Profesiones y oficios Primer curso			
	ELE	CLM	Homólogo en ELE
1	profesor	老师	profesor
2	estudiante	医生	médico
3	médico	学生	estudiante, alumno
4	camarero	翻译	traductor, intérprete
5	alumno	护士	enfermero
6	policía	律师	abogado
7	enfermero	清洁工	limpiador, barrendero
8	bibliotecario	警察	policía
9	conductor	教师	profesor
10	secretario	演员	actor, actriz
11	taxista	司机	conductor
12		经理	gerente
13		厨师	cocinero
14		歌手	cantante
15		服务员	camarero
16		工程师	ingeniero
17		工人	obrero, trabajador
18		会计	contable
19		法官	juez
20		公务员	funcionario
21		记者	periodista, reportero, corresponsal
22		消防员	bombero

Tabla 6.93. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Profesiones y oficios en el primer curso

16 Profesiones y oficios			
Segundo curso			
	ELE	CLM	Homólogo en ELE
1	profesor	老师	profesor
2	médico	医生	médico
3	estudiante	学生	estudiante, alumno
4	policía	律师	abogado
5	enfermero	警察	policía
6	camarero	护士	enfermero
7	alumno	翻译	traductor, intérprete
8	obrero	教师	profesor
9	bombero	公务员	funcionario
10	abogado	清洁工	limpiador, barrendero
11	cantante	司机	conductor, chófer
12	jefe	服务员	camarero
13	actor	工人	obrero
14	secretario	演员	actor
15	campesino	工程师	ingeniero
16	gerente	会计	contable
17		消防员	bombero
18		商人	comerciante
19		厨师	cocinero
20		农民	campesino
21		记者	periodista, reportero, corresponsal
22		军人	soldado
23		歌手	cantante
24		经理	gerente
25		空姐	azafata
26		保安	seguridad

Tabla 6.94. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Profesiones y oficios en el segundo curso

16 Profesiones y oficios			
Tercer curso			
	ELE	CLM	Homólogo en ELE
1	profesor	医生	médico, doctor
2	estudiante	老师	profesor
3	médico	学生	estudiante, alumno
4	policía	护士	enfermero
5	doctor	教师	profesor
6	trabajador	翻译	traductor
7	actor	商人	comerciante
8	enfermero	厨师	cocinero
9	camarero	律师	abogado
10	cantante	清洁工	limpiador, barrendero
11	abogado	工人	trabajador, obrero
12	obrero	警察	policía
13	campesino	经理	gerente
14	cocinero	公务员	funcionario
15	jefe	司机	conductor, chófer
16	actriz	工程师	ingeniero
17	gerente	服务员	camarero
18	vendedor	消防员	bombero
19	secretario	收银员	cajero
20	alumno	建筑师	arquitecto
21	traductor	记者	periodista, reportero, corresponsal
22		军人	soldado
23		演员	actor, actriz
24		法官	juez
25		秘书	secretario
26		老板	jefe
27		外交官	diplomático
28		作家	escritor

Tabla 6.95. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Profesiones y oficios en el tercer curso

16 Profesiones y oficios			
Cuarto curso			
	ELE	CLM	Homólogo en ELE
1	profesor	老师	profesor, maestro
2	estudiante	医生	médico, doctor
3	médico	教师	profesor, maestro
4	cocinero	学生	estudiante, alumno
5	abogado	律师	abogado
6	alumno	厨师	cocinero
7	enfermero	销售	vendedor, dependiente
8	policía	护士	enfermero
9	doctor	会计	contable
10	conductor	公务员	funcionario
11	cantante	经理	gerente
12	actor	演员	actor
13	secretario	歌手	cantante
14	presidente	翻译	traductor, intérprete
15	jefe	清洁工	limpiador, barrendero
16	ingeniero	警察	policía
17	trabajador	服务员	camarero
18	campesino	作家	escritor
19	maestro	司机	conductor
20	director	秘书	secretario
21	bailarín	工人	trabajador
22	escritor	消防员	bombero
23		售货员	vendedor, dependiente

Tabla 6.96. Nociones con índice de disponibilidad superior a 0.1 en ELE y en CLM en Profesiones y oficios en el cuarto curso

b1) Conceptos coincidentes en las dos lenguas:

Primer curso:

profesor; estudiante; médico; camarero; alumno; policía; enfermero; conductor.

Segundo curso:

profesor; médico; estudiante; policía; enfermero; camarero; alumno; obrero; bombero; abogado; cantante; actor; campesino; gerente.

Tercer curso:

profesor; estudiante; médico; policía; doctor; trabajador; actor; enfermero; camarero; abogado; obrero; cocinero; jefe; actriz; gerente; secretario; alumno; traductor.

Cuarto curso:

profesor; estudiante; médico; cocinero; abogado; alumno; enfermero; policía; doctor; conductor; cantante; actor; secretario; trabajador; maestro;

escritor.

b2) Vocablos únicamente presentes en ELE:

Primer curso:

bibliotecario; secretario; taxista.

Segundo curso:

jefe; secretario.

Tercer curso:

cantante; campesino; vendedor.

Cuarto curso:

presidente; jefe; ingeniero; campesino; director; bailarín.

b3) Vocablos ausentes en ELE (en comparación con los vocablos en CLM):

Primer curso:

traductor, intérprete; abogado; limpiador, barrendero; actor, actriz; gerente; cocinero; cantante; ingeniero; obrero, trabajador; contable; juez; funcionario; periodista, reportero, corresponsal; bombero.

Segundo curso:

traductor, intérprete; profesor; funcionario; limpiador, barrendero; conductor, chófer; ingeniero; contable; comerciante; cocinero; periodista, reportero, corresponsal; soldado; azafata; seguridad.

Tercer curso:

comerciante; limpiador, barrendero; funcionario; conductor, chófer; ingeniero; bombero; cajero; arquitecto; periodista, reportero, corresponsal; soldado; juez; diplomático; escritor.

Cuarto curso:

vendedor, dependiente; contable; funcionario; gerente; traductor, intérprete; limpiador, barrendero; camarero; bombero.

Como resultado de los análisis cualitativos, llegamos a ver, por un lado, la coincidencia de palabras en español que producen los estudiantes de diferentes niveles de estudio y, por otro lado, cuáles son las palabras que los alumnos han podido actualizar mientras vamos de un curso a otro. Esto da

información didáctica sobre qué nuevos vocablos españoles hay que enseñar a los estudiantes teniendo en cuenta su nivel de estudio.

Además, en lo que respecta a la diferencia entre las dos lenguas tanto global como según el nivel de estudio, hemos podido conocer la coincidencia de conceptos que se les ocurren a los informantes en español y en chino. Y, sobre todo, hemos detectado cuáles son las nociones en las que los alumnos han pensado pero que no son capaces de reproducir en español. Las traducciones al español de estas últimas —las presentadas a lo largo de este capítulo— pueden servir como complemento en la enseñanza del vocabulario español con el fin de disminuir la carencia léxica de la lengua extranjera.

7. CONCLUSIONES

En el marco de la línea de investigación de la disponibilidad léxica, se han llevado a cabo gran cantidad de estudios en el mundo hispánico. La presente tesis forma parte de ese amplio abanico de investigaciones, aportando el léxico disponible de los estudiantes de español como lengua extranjera en Tianjin, China. La originalidad de esta tesis radica en haber recaudado y analizado el léxico disponible de los informantes tanto en español como lengua extranjera (ELE) como en chino como lengua materna (CLM). De modo que nuestro estudio tiene una orientación didáctica que estriba en comparar los resultados léxicos en las dos lenguas y conocer la deficiencia léxica en ELE de los sujetos que integran la muestra en relación con el comportamiento léxico que ellos mismos manifiestan en CLM.

En cuanto a los diccionarios de disponibilidad léxica en las dos lenguas, se han elaborado y se adjuntan como apéndices al final de la tesis. Para el léxico disponible de ambas lenguas, contemplamos los mismos factores lingüísticos —los dieciséis centros de interés clásicos— y las mismas variables extralingüísticas —nivel de estudio, sexo, procedencia, nivel de inglés y tipo de centro—. A continuación, vamos a concluir los resultados en ELE y en CLM que hemos obtenido a lo largo de los apartados anteriores de esta tesis, con respecto a dichos factores y variables.

I. Análisis cuantitativo global

En lo que se refiere al análisis cuantitativo, observamos, en primer lugar, los índices globales, es decir, los índices que revelan los datos léxicos en su conjunto sin discriminar entre los centros de interés. Tanto en relación con las palabras como con los vocablos, los resultados cuantitativos de CLM superan a los de ELE, tal como era de esperar.

1) Productividad léxica

Sobre el total de palabras, aunque el número de informantes en la muestra de ELE (288) es mayor que aquel de CLM (172), obtenemos dos totales de palabras poco diferenciados (44525 en ELE y 44191 en CLM). Esto repercute en el promedio de palabras por centro de interés (2782.81 en ELE y 2762.06 en CLM). Al eliminar la influencia provocada por la diferencia en el número de informantes, vemos que el promedio de palabras por informante es 154.6 en ELE y 256,94 en CLM. Al dividir este promedio por los centros de interés, hallamos que la producción léxica por informante y por centro de interés es de 9.66 palabras en las encuestas de ELE y 16.06 en las de CLM. De este modo, nos damos cuenta de que los alumnos escriben, como media, 6.4 palabras menos en un centro de interés de la encuesta en ELE que en uno de la encuesta en CLM. No cabe duda de que los estudiantes manifiestan una productividad léxica bastante menor en ELE que en CLM, toda vez que la producción léxica de los alumnos en ELE solo supone alrededor del 60% de su producción en CLM.

2) Riqueza léxica

El total de vocablos en las dos lenguas están próximos entre sí (5650 en ELE y 6145 en CLM). El volumen de los vocablos en ELE supone el 91.94% de aquel en CLM. Así pues, por centro de interés, obtenemos un promedio de riqueza léxica de 353.13 vocablos en ELE y 384.06 en CLM. Al relacionar la cifra de los vocablos con la de las palabras, conseguimos dos porcentajes de vocablos sobre palabras cercanos (12.69% en ELE y 13.90% en CLM), los cuales indican cuántos vocablos se registran por cada cien palabras. Vemos que la riqueza léxica señala un déficit leve en los resultados obtenidos en ELE con respecto a los en CLM.

Con el objetivo de eliminar la influencia provocada por el comportamiento de los informantes atípicos, recurrimos a cuatro índices más sobre los vocablos. Al poner el límite de vocablos en el 75% de la frecuencia acumulada,

detectamos que el número de vocablos en ELE es inferior en comparación con el obtenido en CLM (850 en ELE y 1203 en CLM). De manera semejante, la inferioridad estadística también se encuentra al poner el límite en el 0.02 del índice de disponibilidad (791 en ELE y 1162 en CLM). Se trata de dos criterios de corte para hacer comparaciones de vocablos entre distintos subconjuntos. Por otra parte, el número de vocablos altamente disponibles —con un I.D. superior a 0.1— (223 en ELE y 341 en CLM) y el número de vocablos altamente frecuentes —mencionados por más del 25% de los informantes— (139 en ELE y 242 en CLM) nos revelan otra vez el déficit en la riqueza léxica de los estudiantes en ELE con respecto a la en CLM.

II. Análisis cuantitativo por centro de interés

Una vez expuesto esto, reseñamos, en segundo lugar, las conclusiones sacadas sobre los índices por centro de interés. En resumen, los índices —sean relativos a las palabras o a los vocablos— de los distintos centros difieren entre sí en cada lengua, pero, el mayor o menor grado de los índices —el rango según las palabras o los vocablos— de los distintos centros se manifiesta de manera parecida en ELE y en CLM. Ahora lo explicamos.

1) Productividad léxica

Los resultados evidencian que los centros de interés, como factores lingüísticos que influyen en el léxico disponible, se relacionan con una producción léxica muy diferente en cada caso. La mayor distancia en ELE se da entre *Alimentos y bebidas* (4355) e *Iluminación y calefacción* (1613), una diferencia de 2742 palabras; mientras que la mayor distancia en CLM entre el centro de interés más productivo y el menos es de 1809 palabras, y se da entre *Los animales* (3765) y *Objetos colocados en la mesa para la comida* (1956). Las áreas nocionales se diferencian en un mayor grado en ELE que en CLM con respecto a la producción de palabras.

Se detecta una coincidencia entre ELE y CLM en el mayor o menor grado

de productividad léxica de las áreas nocionales: salvo *Ropa y complementos*, los centros de interés que superan la media del total de palabras son los mismos en las dos lenguas. Concretamente, por encima de la media están los centros más productivos: *Partes del cuerpo humano, Alimentos y bebidas, La escuela: sus materiales y muebles, La ciudad, El campo, Los animales y Profesiones y oficios* en las dos lenguas, más *Ropa y complementos* en CLM. Por el contrario, quedan por debajo de la media los centros menos productivos: *Partes de la casa (sin muebles), Los muebles de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Iluminación y calefacción, Medios de transporte, Trabajos del campo y del jardín y Juegos y distracciones* en las dos lenguas, más *Ropa y complementos* en ELE.

De acuerdo con el total de palabras, los centros de interés se ordenan, de mayor a menor producción léxica, tal como se presentan a continuación:

En ELE	En CLM
<i>Alimentos y bebidas</i>	<i>Los animales</i>
<i>La ciudad</i>	<i>Alimentos y bebidas</i>
<i>Partes del cuerpo humano</i>	<i>Partes del cuerpo humano</i>
<i>Profesiones y oficios</i>	<i>Profesiones y oficios</i>
<i>La escuela: muebles y materiales</i>	<i>La ciudad</i>
<i>Los animales</i>	<i>Ropa y complementos</i>
<i>El campo</i>	<i>La escuela: muebles y materiales</i>
<i>Juegos y distracciones</i>	<i>El campo</i>
<i>Medios de transporte</i>	<i>Juegos y distracciones</i>
<i>Los muebles de la casa</i>	<i>Los muebles de la casa</i>
<i>Ropa y complementos</i>	<i>Medios de transporte</i>
<i>Partes de la casa (sin muebles)</i>	<i>La cocina y sus utensilios</i>
<i>Trabajos del campo y del jardín</i>	<i>Partes de la casa (sin muebles)</i>
<i>La cocina y sus utensilios</i>	<i>Trabajos del campo y del jardín</i>
<i>Objetos en la mesa para la comida</i>	<i>Iluminación y calefacción</i>
<i>Iluminación y calefacción</i>	<i>Objetos en la mesa para la comida</i>

Observamos que *La ciudad*, *Los animales* y *Ropa y complementos* presentan una mayor diferencia entre sus rangos en ELE y en CLM, es decir, son las áreas ante las que los informantes se comportan de la manera más distinta en una lengua que en la otra en cuanto al rango según la producción de palabras. El resto de centros son bastante coincidentes en el orden según sus rangos de producción léxica en ELE y en CLM.

Al observar el promedio de palabras, que elimina la influencia provocada por el número de informantes, vemos que los resultados de producción en ELE son superados por los obtenidos en CLM en todos los centros de interés sin excepción, tal como era de esperar. A pesar de ello, llamamos la atención sobre la gran deficiencia de los datos en ELE: el área semántica que alcanza el máximo promedio en ELE (*Alimentos y bebidas*, 15.12) tiene un resultado inferior a la décima área en CLM (*Los muebles de la casa*, 15.28), y el centro de interés con la menor cantidad de producción en CLM (*Objetos colocados en la mesa para la comida*, 11.37) muestra un resultado por encima del centro que ocupa la quinta posición en ELE (*La escuela: sus muebles y materiales*, 11.13).

Para tener una idea de cuán grande es la deficiencia en la productividad léxica en ELE en comparación con la arrojada en CLM, cotejamos los promedios de palabras de las dos lenguas en cada área nocional. Observamos que la mayor deficiencia se da en *Los animales*, donde los estudiantes escriben, en promedio, 11.35 palabras menos en ELE que en CLM. En síntesis, *Los animales*, *Ropa y complementos* y *La cocina y sus utensilios* son las áreas ncionales en las que la producción léxica de los estudiantes en ELE se aleja demasiado de su producción en CLM; por tanto, son los centros donde los informantes muestran una mayor deficiencia al respecto en ELE.

2) Riqueza léxica

Al igual que lo hallado en el análisis de la productividad léxica, los centros de interés, como factores lingüísticos, también influyen en el grado de riqueza

léxica. Dicho de otra manera, el total de vocablos de un centro difiere mucho del total de otro. La mayor diferencia al respecto en ELE y en CLM se detecta entre los mismos dos centros de interés: *La ciudad* (700 en ELE y 789 en CLM) y *Objetos colocados en la mesa para la comida* (165 en ELE y 185 en CLM). Dicha divergencia en ELE —535 vocablos— supone un 9.47% del conjunto del vocabulario en ELE, mientras esa cifra en CLM —604 vocablos— corresponde al 9.81% del conjunto en CLM. Así pues, con respecto a la riqueza léxica, los centros de interés difieren en un grado parecido en ELE y en CLM.

Se encuentra una total coincidencia entre ELE y CLM en cuanto a los centros que suponen una mayor riqueza léxica. Es decir, las áreas temáticas que tienen un número total de vocablos por encima de la media son las mismas en las dos lenguas: *Alimentos y bebidas*, *La ciudad*, *El campo*, *Trabajos del campo y del jardín*, *Juegos y distracciones* y *Profesiones y oficios*. Por consiguiente, los centros menos ricos en el vocabulario —los que tienen un número de vocablos inferior a la media— también son los mismos en las dos lenguas: *Partes del cuerpo humano*, *Ropa y complementos*, *Partes de la casa (sin muebles)*, *Los muebles de la casa*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios*, *La escuela: sus muebles y materiales*, *Iluminación y calefacción*, *Medios de transporte* y *Los animales*.

De acuerdo con el total de vocablos, los centros de interés se ordenan, de mayor a menor riqueza léxica, tal como se presentan a continuación:

En ELE	En CLM
<i>La ciudad</i>	<i>La ciudad</i>
<i>El campo</i>	<i>El campo</i>
<i>Trabajos del campo y del jardín</i>	<i>Alimentos y bebidas</i>
<i>Profesiones y oficios</i>	<i>Profesiones y oficios</i>
<i>Juegos y distracciones</i>	<i>Juegos y distracciones</i>
<i>Alimentos y bebidas</i>	<i>Trabajos del campo y del jardín</i>
<i>La escuela: muebles y materiales</i>	<i>Ropa y complementos</i>
<i>Medios de transporte</i>	<i>Los animales</i>

<i>Ropa y complementos</i>	<i>La escuela: muebles y materiales</i>
<i>Los animales</i>	<i>Medios de transporte</i>
<i>Partes del cuerpo humano</i>	<i>Los muebles de la casa</i>
<i>Iluminación y calefacción</i>	<i>Partes del cuerpo humano</i>
<i>La cocina y sus utensilios</i>	<i>Iluminación y calefacción</i>
<i>Partes de la casa (sin muebles)</i>	<i>La cocina y sus utensilios</i>
<i>Los muebles de la casa</i>	<i>Partes de la casa (sin muebles)</i>
<i>Objetos en la mesa para la comida</i>	<i>Objetos en la mesa para la comida</i>

Observamos que *Los muebles de la casa* presenta la mayor diferencia entre sus rangos en ELE y en CLM, es decir, es el área ante la que los informantes se comportan de la manera más distinta en una lengua que en la otra en cuanto al rango según el número de vocablos. El resto de centros coinciden bastante en el orden según sus rangos de riqueza léxica en ELE y en CLM.

Al igual que en el análisis de los índices globales, para el análisis por centro de interés, también recurrimos a cuatro índices más sobre los vocablos, con el fin de eliminar la influencia provocada por el comportamiento de los informantes atípicos.

Según el número de vocablos hasta el 75% de la frecuencia acumulada, los centros que tienen la cifra superior a la media en ELE y en CLM son los mismos, se da una coincidencia exacta: *Alimentos y bebidas, La ciudad, El campo, Trabajos del campo y del jardín, Juegos y distracciones y Profesiones y oficios*. De modo que las áreas nocionales que muestran un resultado por debajo de la media en ambas lenguas son *Partes del cuerpo humano, Ropa y complementos, Partes de la casa (sin muebles), Los muebles de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, La escuela: sus muebles y materiales, Iluminación y calefacción, Medios de transporte y Los animales*.

Con respecto al rango que los centros alcanzan según el número de

vocablos hasta el 75% de la frecuencia acumulada, las dos lenguas coinciden en la mitad de las áreas —es decir, la mitad de los centros muestran el mismo rango en ELE y en CLM—, mientras que, para el resto de campos semánticos, el rango que un campo alcanza en una lengua es distinto, pero no se aleja mucho del rango que alcanza en la otra lengua. En ambas lenguas, el campo léxico que da cabida al mayor número de vocablos es *La ciudad*, seguido por *El campo*. Por el contrario, el área semántica que contiene la menor cantidad de vocablos en ambas lenguas es *Objetos colocados en la mesa para la comida*, seguido por *Partes de la casa (sin muebles)*, *Partes del cuerpo humano*, *Los muebles de la casa* y *Medios de transporte*.

Llama la atención que, según el número de vocablos hasta el 75% de la frecuencia acumulada, en *Trabajos del campo y del jardín*, el número de vocablos en ELE aventaja el número en CLM, aunque se trata de una ventaja pequeña (111 en ELE y 95 en CLM). El resto de centros evidencian, como siempre, la superioridad de CLM sobre ELE.

Según el número de vocablos hasta el 0.02 del índice de disponibilidad, los resultados de las dos lenguas no muestran mucha coincidencia, ni en la división de los centros por la media, ni en el rango de las áreas. Ello resulta comprensible, porque el índice de disponibilidad considera, además de la frecuencia, la posición de las palabras en las listas. Los estudiantes pueden coincidir fácilmente en los vocablos —lo que determina la frecuencia de un vocablo y el número de vocablos frecuentes—, pero difícilmente coinciden en la posición de los vocablos.

Los campos léxicos que albergan más vocablos hasta el 0.02 del índice de disponibilidad—en referencia a la media— en ELE son, de mayor a menor grado, *La ciudad*, *Alimentos y bebidas*, *Profesiones y oficios*, *El campo*, *Juegos y distracciones* y *La escuela: sus muebles y materiales*. En CLM, los campos léxicos de esta índole son *Alimentos y bebidas*, *La ciudad*, *El campo*, *Juegos y profesiones*, *Profesiones y oficios*, *Ropa y complementos*, *La cocina y sus utensilios* y *Trabajos del campo y del jardín*.

Además, tal como se imagina, en todos los centros de interés, la cantidad de los vocablos hasta el 0.02 del índice de disponibilidad en ELE se encuentra en desventaja con respecto a esa cantidad en CLM.

Según el número de vocablos con índice de disponibilidad igual o mayor que 0.1 (vocablos altamente disponibles), los campos temáticos que cuenta con más de veinte vocablos altamente disponibles en ELE son dos: *Alimentos y bebidas* (26) y *Los animales* (20). En cambio, en CLM, hay diez centros de interés de esta índole: *Partes del cuerpo humano* (31), *Alimentos y bebidas* (29), *La escuela: sus muebles y materiales* (27), *Ropa y complementos* (24), *Los animales* (24), *Los muebles de la casa* (23), *La cocina y sus utensilios* (23), *Profesiones y oficios* (23), *Partes de la casa (sin muebles)* (20) e *Iluminación y calefacción* (20). Destacamos que *Iluminación y calefacción* en ELE (7) es el único campo léxico en el que se registran menos de diez vocablos altamente disponibles.

Llama la atención que, en *La ciudad*, el número de vocablos altamente disponibles en ELE aventaja el número obtenido en CLM, aunque se trata de una ventaja mínima (15 en ELE y 14 en CLM). El resto de centros evidencia, como siempre, la superioridad de CLM sobre ELE.

Según el número de vocablos mencionados por más del 25% de los informantes (vocablos altamente frecuentes), los campos temáticos que cuentan con más de quince vocablos altamente frecuentes en ELE son dos: *Alimentos y bebidas* (18) y *Partes del cuerpo humano* (16). En cambio, en CLM, hay ocho centros de interés de esta índole: *Partes del cuerpo humano* (26), *Los animales* (25), *Ropa y complementos* (20), *La escuela: sus muebles y materiales* (19), *Los muebles de la casa* (19), *Profesiones y oficios* (18), *La cocina y sus utensilios* (17) y *Alimentos y bebidas* (15). Cabe mencionar que *Iluminación y calefacción* y *Trabajos del campo y del jardín* en ELE (4 en ambos centros) son los únicos campos léxicos en los que se registran menos de cinco vocablos altamente frecuentes.

3) Densidad léxica e índice de cohesión

Después de analizar, por separado, las palabras y los vocablos, nos fijamos en la densidad léxica y el índice de cohesión, los que relacionan las palabras con los vocablos para medir la homogeneidad de las respuestas dadas por los informantes en cada centro de interés.

De acuerdo con la densidad léxica y el índice de cohesión, los centros de interés se ordenan, de mayor a menor grado de coincidencia de las respuestas, tal como se presentan a continuación:

En ELE	En CLM
<i>Partes del cuerpo humano</i>	<i>Partes del cuerpo humano</i>
<i>Los muebles de la casa</i>	<i>Los animales</i>
<i>Alimentos y bebidas</i>	<i>Objetos en la mesa para la comida</i>
<i>Objetos en la mesa para la comida</i>	<i>La cocina y sus utensilios</i>
<i>Los animales</i>	<i>Los muebles de la casa</i>
<i>Partes de la casa (sin muebles)</i>	<i>Partes de la casa (sin muebles)</i>
<i>La escuela: muebles y materiales</i>	<i>La escuela: muebles y materiales</i>
<i>Profesiones y oficios</i>	<i>Iluminación y calefacción</i>
<i>Medios de transporte</i>	<i>Ropa y complementos</i>
<i>La cocina y sus utensilios</i>	<i>Medios de transporte</i>
<i>Ropa y complementos</i>	<i>Profesiones y oficios</i>
<i>Juegos y distracciones</i>	<i>Juegos y distracciones</i>
<i>Iluminación y calefacción</i>	<i>Trabajos del campo y del jardín</i>
<i>La ciudad</i>	<i>Alimentos y bebidas</i>
<i>El campo</i>	<i>El campo</i>
<i>Trabajos del campo y del jardín</i>	<i>La ciudad</i>

El centro de interés más compacto —centro en el que las respuestas por distintos informantes son más homogéneas— en las dos lenguas es el mismo: *Partes del cuerpo humano*. Por el contrario, el área temática más difusa —área

en la que las respuestas por distintos informantes son más heterogéneas— en ELE difiere de aquella en CLM. En ELE, *Trabajos del campo y del jardín* es la más difusa, mientras que, en CLM, la más difusa es el área *La ciudad*.

A pesar de que los campos léxicos manifiestan en las dos lenguas distinto grado en la escala de más compacto a más difuso, observamos que, salvo *Alimentos y bebidas*, *La cocina y sus utensilios* e *Iluminación y calefacción*, existe una tendencia más o menos parecida entre los rangos que los centros alcanzan en ELE y en CLM: los centros más compactos en ambas lenguas son *Partes del cuerpo humano*, *Los muebles de la casa*, *Objetos colocados en la mesa para la comida* y *Los animales*; mientras que los centros más difusos en ambas lenguas son *Trabajos del campo y del jardín*, *El campo*, *La ciudad* y *Juegos y distracciones*.

Al comparar la densidad léxica o el índice de cohesión de cada centro en ELE y en CLM, vemos que, únicamente en *Alimentos y bebidas*, el resultado en ELE aventaja el alcanzado en CLM —más compacto en ELE que en CLM—. Esto se puede deber a que, en su país de origen, los estudiantes tienen contacto con una gran variedad de alimentos y bebidas y son capaces de aludirlos todos en su lengua materna; pero, en lengua extranjera, es posible que el alumnado adquiera en las aulas los mismos vocablos, es decir, solamente los referentes a aquellos alimentos y bebidas básicos y habituales de la vida cotidiana.

III. Análisis cuantitativo de las variables extralingüísticas

De aquí en adelante, reseñamos, en tercer lugar, las conclusiones que obtenemos después de analizar la producción léxica —en concreto, el promedio de palabras por informante— de los distintos estratos en los que hemos dividido las variables extralingüísticas. Con ello, llegamos a saber la incidencia cuantitativa de las cinco variables.

1) Nivel de estudio

La estratificación según el nivel de estudio se corresponde con los cuatro cursos universitarios en los que están estudiando los informantes en el momento de la encuesta: primero, segundo, tercero y cuarto cursos.

a) En ELE

Alimentos y bebidas, La ciudad, Partes del cuerpo humano y Profesiones y oficios se cuentan entre los campos semánticos más productivos en ELE en los cuatro niveles de estudio. Aparte de dichos centros, los alumnos del primer curso también disponen de mejor producción léxica en ELE en *La escuela: sus muebles y materiales*; los del segundo curso, en *El campo*; mientras que los del tercer y cuarto curso, en *Los animales*.

Por un lado, suponemos que los estudiantes de primer curso, sin haber aprendido muchos vocablos de español por la limitación que tiene su fase de estudio, lógicamente dominan mejor las palabras relacionadas con su entorno habitual de estudio —*La escuela: sus muebles y materiales*—, porque es donde toman las clases de ELE y, por tanto, es la circunstancia comunicativa en la que utilizan con más probabilidad el español. Por otro lado, vemos que, en lo que se refiere a la productividad léxica, *Los animales* es el campo asociativo que asciende con mayor rapidez en ELE a medida que incrementa el nivel de estudio de los estudiantes. Adelantamos que *Los animales* es el centro de interés que alcanza la mayor productividad léxica en CLM en todos los cursos. Por eso, dicho ascenso de *Los animales* en ELE es del todo lógico, en el sentido de que, a medida que aumenta el nivel de estudio, el grado de productividad de este centro en ELE se acerca cada vez más al grado de productividad que los alumnos tienen en su lengua materna en este mismo centro.

Objetos colocados en la mesa para la comida, Iluminación y calefacción y La cocina y sus utensilios se cuentan entre los centros menos productivos en ELE en los cuatro niveles de estudio. Como diferencias, el primer curso

también muestra una menor producción léxica en ELE en *Trabajos del campo y del jardín* y *El campo*; el segundo curso, en *Trabajos del campo y del jardín* y *Partes de la casa (sin muebles)*; el tercer curso, en *Trabajos del campo y del jardín* y *Los muebles de la casa*; y el cuarto curso, en *Partes de la casa (sin muebles)* y *Los muebles de la casa*.

Al observar la productividad léxica de los cuatro cursos en cada centro de interés, concluimos que la productividad léxica en ELE de los estudiantes crece a medida que incrementa su nivel de estudio —tanto del primero al segundo como del segundo al tercer curso—. No obstante, el nivel de estudio deja de ejercer influencias en la productividad léxica en ELE de los alumnos, una vez que estos entran en el nivel de estudio más alto —el cuarto curso—, ya que la producción se mantiene estable en los últimos dos cursos. Suponemos que se debe a que los estudiantes de esa última fase de estudio no avanzan en el vocabulario porque tienen muchas menos clases programadas de especialidad en comparación con las fases anteriores y habitualmente se ocupan de otras tareas ya no académicas, como las prácticas externas y entrevistas de trabajo.

b) En CLM

Los animales, Alimentos y bebidas, Partes del cuerpo humano y Profesiones y oficios se cuentan entre los campos semánticos más productivos en CLM en todos los niveles de estudio. Además, los alumnos de primer curso también disponen de mejor producción léxica en CLM en *Ropa y complementos*; los de segundo y tercer curso, en *El campo*; mientras los de cuarto curso, en *La ciudad*.

Objetos colocados en la mesa para la comida, Iluminación y calefacción, Trabajos del campo y del jardín y Partes de la casa (sin muebles) son los centros menos productivos en CLM de los cuatro niveles de estudio. Como diferencias, el primer curso muestra debilidad en la producción léxica en CLM en *Medios de transporte*; el tercer curso, en *Los muebles de la casa*; mientras

el segundo y el cuarto curso, en *La cocina y sus utensilios*.

Comparando estos resultados con los previamente extraídos en ELE según el nivel de estudio, vemos que, en todos los niveles de estudio, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son los centros de interés con un alto grado de productividad léxica tanto en ELE como en CLM. Mientras tanto, *Objetos colocados en la mesa para la comida e Iluminación y calefacción* son, tanto en ELE como en CLM, los dos centros con una menor productividad léxica en los cuatro niveles de estudio.

Deducimos que la producción léxica que los alumnos tienen en ELE viene fuertemente afectada por la que ellos tienen en CLM, ya que las áreas más y menos productivas en CLM resultan ser las más y menos productivas también en ELE. Suponemos que, independientemente de qué palabras aprendan en el aula de ELE, los estudiantes muestran la inclinación a dominar mejor las palabras españolas de aquellos campos léxicos que ya manejan muy bien en su lengua materna, porque el léxico disponible en CLM les impulsa a aprender el léxico homólogo en ELE. De igual modo, los alumnos tienden a tener menor productividad en ELE en aquellos campos semánticos en los que tienen menos competencia léxica en CLM.

Si observamos la productividad léxica en CLM de los cuatro cursos en cada centro de interés, llegamos a ver la misma tendencia demostrada en ELE. La producción léxica en CLM de los estudiantes se incrementa del primer curso al segundo, y del segundo al tercero. La incidencia positiva del nivel de estudio se demuestra tanto en el léxico disponible en ELE como en el que tienen en CLM. Sin embargo, al igual que en ELE, el estrato del cuarto curso constituye una excepción, ya que muestra en CLM un retroceso con respecto al curso anterior.

c) Entre ELE y CLM

La producción léxica en ELE es inferior a aquella alcanzada en CLM en

todos los centros de interés y en todos los niveles de estudio, lo cual es del todo comprensible.

Destacamos que *Ropa y complementos* y *Los animales* son los centros en los que los alumnos de todos los niveles de estudio muestran considerable debilidad de producción en ELE, tomando como referencia su producción léxica en CLM. A este respecto, hay unos campos léxicos que implican una alta deficiencia en ELE en tres de los estratos: el segundo, tercero y cuarto curso también en *La cocina y sus utensilios*; el primero, segundo y tercer curso, en *Iluminación y calefacción*; el primero, segundo y tercer curso, en *El campo*; y el primero, segundo y cuarto curso, en *Juegos y distracciones*. Como casos únicos, solo el cuarto curso presenta debilidad en *Partes de la casa (sin muebles)* y *Los muebles de la casa*; mientras solo el tercer curso, en *Objetos colocados en la mesa para la comida*.

Observamos que la diferencia que existe entre los promedios de palabras en ELE y en CLM va disminuyendo a medida que incrementa el nivel de estudio de los informantes. Es decir, en referencia a la productividad léxica en CLM, los informantes de niveles de estudio altos señalan mejores resultados en ELE que los de niveles bajos. Así pues, concluimos que los estudiantes generan una producción léxica en ELE cada vez más aproximada a la que tienen en CLM, a la par que van ascendiendo de nivel de estudio.

2) Sexo

Bajo la variable sexo, comparamos los resultados obtenidos en los varones y los obtenidos en las mujeres.

a) En ELE

Los chicos y las chicas coinciden totalmente en los campos semánticos con una mayor producción léxica en ELE: *Alimento y bebidas*, *La ciudad*, *Partes del cuerpo humano*, *La escuela: sus muebles y materiales* y *Profesiones*

y oficios.

Igualmente, en cuanto a las áreas nocionales menos productivas en ELE, los chicos y las chicas coinciden totalmente en los campos semánticos con una menor producción léxica: *Iluminación y calefacción, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Trabajos del campo y del jardín y Partes de la casa (sin muebles)*.

En realidad, los varones y las mujeres coinciden en la mayoría de los centros con respecto al rango de producción léxica en ELE. Deducimos que el sexo no ejerce incidencia notoria en relación al mayor o menor grado de productividad léxica en ELE de los centros de interés.

Al comparar la producción léxica en ELE de los chicos y de las chicas en cada centro de interés, vemos que las mujeres muestran mayor productividad léxica en ELE que los varones en más centros de interés, pero con una ventaja que apenas supera una palabra. No creemos que el sexo tenga una influencia significativa en la productividad léxica en ELE de nuestros informantes.

b) En CLM

Los animales, Alimentos y bebidas, Partes del cuerpo humano y Profesiones y oficios son los campos semánticos más productivos en CLM tanto de los chicos como de las chicas. Además, *La ciudad* está entre los campos asociativos más productivos en CLM de los varones; mientras que *Ropa y complementos* está entre los más productivos de las mujeres.

Objetos colocados en la mesa para la comida, Iluminación y calefacción, Partes de la casa (sin muebles) y Trabajos del campo y del jardín son los centros con una menor producción léxica en CLM tanto de los chicos como de las chicas. Como diferencias, los chicos muestran resultados deficientes en ELE en *Los muebles de la casa*, mientras que las chicas lo hacen en *La cocina y sus utensilios*.

Comparando estos resultados con los previamente extraídos en ELE según el sexo, vemos que, en ambos sexos, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son los centros de interés comunes en ELE y en CLM con un alto grado de productividad léxica. Al mismo tiempo, las cuatro áreas nocionales menos productivas en CLM, comunes en ambos sexos, también se hallan entre sus áreas menos productivas en ELE. Comprobamos cómo la productividad en CLM de las áreas semánticas afecta a la productividad en ELE, sin verse afectada por la variable sexo.

Si comparamos la producción léxica en CLM de los chicos y de las chicas en cada centro de interés, podemos ver que no existe una superioridad de un sexo sobre el otro. En los siete campos léxicos donde la divergencia de producción léxica entre los dos sexos está por encima de una palabra, los chicos manifiestan mejores resultados que las chicas en cinco campos. Por eso, aunque los varones disponen, de esta manera, de mayor productividad léxica en CLM que las mujeres, la superioridad resulta muy leve.

c) Entre ELE y CLM

Como puede imaginarse, el promedio de palabras en ELE es inferior al obtenido en CLM en todos los centros de interés y en ambos sexos.

Los chicos y las chicas coinciden en seis centros de interés que señalan una gran divergencia de producción léxica entre las dos lenguas: *Partes del cuerpo humano*, *Ropa y complementos*, *La cocina y sus utensilios*, *Los animales*, *Juegos y distracciones* y *Profesiones y oficios*. Es decir, ambos sexos muestran una considerable debilidad en estos seis centros en ELE, tomando como referencia su producción léxica en CLM. A este respecto, también indicamos que solo los varones tienen alta deficiencia en ELE en *Alimentos y bebidas* y *La ciudad*. Únicamente en el estrato de las mujeres, son *Los muebles de la casa*, *Iluminación y calefacción* y *El campo* en los que la deficiencia de la producción en ELE resulta considerable con respecto a la

producción en CLM.

El grado de distancia entre la producción léxica en ELE y en CLM es más o menos el mismo en los chicos que en las chicas. El sexo tiene una influencia muy pequeña con relación a la diferencia de productividad léxica entre las dos lenguas. Únicamente en *La ciudad*, la productividad léxica en ELE dista de la productividad en CLM en los varones en un grado mayor que el que observamos en las mujeres. Es decir, solo *La ciudad* favorece a las chicas, en cuanto a la productividad en ELE con respecto a CLM.

3) Procedencia

La variable de la procedencia contempla el lugar de origen de los estudiantes según de cuál de las cuatro zonas socioeconómicas del país procedan: la zona del este, la del nordeste, la del centro y la del oeste. El orden de las zonas, de la más desarrollada económicamente a la menos desarrollada, es el siguiente: este, nordeste, centro y oeste. Pretendemos averiguar si la disponibilidad léxica de los alumnos tienen que ver con el grado de desarrollo de su zona de procedencia.

a) En ELE

Alimentos y bebidas, La ciudad, Partes del cuerpo humano y Profesiones y oficios se cuentan entre los campos semánticos más productivos en ELE en las cuatro zonas. Aparte de ellos, los alumnos del este, centro y oeste también disponen de mejor producción léxica en ELE en *La escuela: sus muebles y materiales*; mientras que los informantes del nordeste lo hacen en *Los animales*.

Iluminación y calefacción, Objetos colocados en la mesa para la comida y La cocina y sus utensilios son los campos menos productivos en ELE en todas las zonas. Además, los estratos del este y oeste muestran una menor producción léxica en ELE en *Trabajos del campo y del jardín y Partes de la*

casa (sin muebles); los alumnos del nordeste, en *Partes de la casa (sin muebles)*, *Los muebles de la casa* y *Medios de transporte*; y los del centro, en *Trabajos del campo y del jardín* y *Ropa y complementos*.

Al observar la productividad léxica en ELE de las tres zonas de procedencia en cada centro de interés (excluimos la zona del nordeste por sus pocos integrantes en la muestra), vemos la incidencia de la procedencia en el léxico disponible: el estrato del centro muestra la mayor productividad léxica en ELE, seguido por el estrato del oeste, y el del este es el que tiene el menor índice. Así pues, la productividad léxica en ELE se ve aventajada en la zona de grado de desarrollo medio.

Sin embargo, *Alimentos y bebidas* y *Trabajos del campo y del jardín* se desmarcan de la tendencia comentada. Creemos que el vocabulario de dichas dos áreas nocionales está fuertemente afectado por la zona geográfica. Es decir, generalmente la geografía puede determinar, por un lado, qué tipo de comidas la gente consume por hábito o por gusto y, por otro lado, qué tipo de vegetales crecen en el territorio y en qué modalidades la gente trabaja con lo que planta. A diferencia de *Alimentos y bebidas* y *Trabajos del campo y del jardín*, el vocabulario de los demás centros de interés se mantiene estable sea cuál sea la zona geográfica. Por eso, en los demás campos léxicos, observamos más claramente la incidencia de la zona de procedencia en la producción léxica de los estudiantes, pero, en *Alimentos y bebidas* y *Trabajos del campo y del jardín*, vemos más bien la incidencia de la zona de procedencia en el vocabulario de los centros de interés.

b) En CLM

Los animales, *Alimentos y bebidas*, *Partes del cuerpo humano* y *Profesiones y oficios* son las áreas temáticas más productivas en CLM en todas las zonas de procedencia. Además, los estratos del este y oeste cuentan con *La ciudad* entre sus áreas con una mayor producción léxica en CLM. Solamente la zona del nordeste incluye entre sus campos más productivos en

CLM *Juegos y distracciones* y, solamente la zona del centro, *Ropa y complementos*.

Objetos colocados en la mesa para la comida, Iluminación y calefacción, Partes de la casa (sin muebles) y Trabajos del campo y del jardín se presentan en todas las zonas de procedencia con un menor promedio de palabras por informante en CLM. *La cocina y sus utensilios* se halla, además, en los estratos del este, nordeste y centro, siendo otro campo léxico menos productivo en CLM. Como caso único, solo la zona del oeste tiene *Medios de transporte* entre sus áreas nocionales con una menor producción léxica en CLM.

Comparando estos resultados con los previamente extraídos en ELE según la procedencia, *Alimentos y bebidas, Partes del cuerpo humano y Profesiones y oficios* son los centros de interés comunes en ELE y en CLM en cuanto a tener un alto grado de productividad léxica. De igual modo, *Objetos colocados en la mesa para la comida e Iluminación y calefacción* son dos centros con una menor productividad léxica en las cuatro zonas, tanto en ELE como en CLM. Comprobamos otra vez cómo las áreas semánticas más productivas en CLM pueden ser también las más productivas en ELE.

Al comparar la producción léxica en CLM de las tres zonas de procedencia en cada centro de interés (igual excluimos la zona del nordeste), sabemos que, en la inmensa mayoría de las áreas nocionales, el estrato del oeste muestra la mayor productividad léxica en CLM, seguido por el estrato del centro, y el del este es el que tiene la menor productividad. Así que la productividad léxica en CLM aventaja en la zona de grado de desarrollo bajo.

Esta deducción coincide parcialmente con la que hemos sacado analizando ELE: en las dos lenguas, los estudiantes de la zona de grado de desarrollo alto son los que tienen la menor productividad léxica. Pero, los estratos de las zonas de grado de desarrollo medio y bajo alternan sus posiciones en las dos lenguas: en ELE la zona de grado de desarrollo medio ostenta la mayor productividad léxica, mientras que en CLM la mayor

productividad la tiene la zona de grado de desarrollo bajo.

c) Entre ELE y CLM

La producción léxica en ELE es inferior a la que hay en CLM en todos los centros de interés y en todas las zonas de procedencia, con excepción de *La ciudad* del estrato del nordeste, donde el promedio en ELE supera dicho índice en CLM con 3.71 palabras más.

Hay cuatro centros en los que los alumnos de todas las zonas muestran una considerable debilidad de producción en ELE, tomando como referencia su producción léxica en CLM: *Ropa y complementos*, *La cocina y sus utensilios*, *Los animales* y *Profesiones y oficios*. Además, hay varios campos léxicos que implican deficiencia en ELE con respecto a la producción en CLM en tres de los estratos: los estratos del este, centro y oeste, en *Partes del cuerpo humano*; las zonas del nordeste, centro y oeste, en *Iluminación y calefacción*; y las zonas del este, nordeste y oeste, en *Juegos y distracciones*. Como casos únicos, solo los alumnos del nordeste presentan debilidad en *Alimentos y bebidas*, *Objetos colocados en la mesa para la comida* y *Medios de transporte*; mientras solo la zona del oeste, en *El campo*.

La diferencia que existe entre los promedios de palabras en ELE y en CLM (sin contar con el estrato del nordeste) favorece al estrato de la zona del centro —zona que obtiene la mayor productividad léxica en ELE—. Los informantes del centro tienen una productividad léxica en ELE que se aproxima más a su productividad en CLM, seguidos por los del este y después, los del oeste. Como conclusión, con respecto a la diferencia de productividad léxica entre las dos lenguas, la zona de grado de desarrollo medio tiene la mayor ventaja.

4) Nivel de inglés

Bajo esta variable, vamos a examinar los resultados obtenidos en

informantes con cuatro niveles de inglés distintos: el nivel básico, el intermedio, el avanzado y el superior.

a) En ELE

Alimentos y bebidas, La ciudad y Partes del cuerpo humano son los centros más productivos en ELE en los cuatro estratos según el nivel de inglés. Los niveles básico e intermedio también revelan una mejor producción léxica en ELE en *La escuela: sus muebles y materiales* y *Profesiones y oficios*; el nivel avanzado, en *Profesiones y oficios* y *Los animales*; y el nivel superior, en *La escuela: sus muebles y materiales* y *Los animales*.

Iluminación y calefacción, Objetos colocados en la mesa para la comida y La cocina y sus utensilios son los centros menos productivos en ELE en todos los niveles de inglés. El estrato del nivel básico cuenta entre sus campos semánticos menos productivos en ELE *El campo y Trabajos del campo y del jardín*; el nivel intermedio, *Ropa y complementos y Trabajos del campo y del jardín*; el nivel avanzado, *Partes de la casa (sin muebles)* y *Trabajos del campo y del jardín*; mientras el nivel superior, *Partes de la casa (sin muebles)* y *Los muebles de la casa*.

Al comparar la productividad léxica en ELE de alumnos de los cuatro niveles de inglés en cada área nocional, llegamos a ver que la producción léxica en ELE de los estudiantes incrementa desde el nivel de inglés básico al nivel intermedio, y del intermedio al avanzado. Pero, los resultados del estrato superior no siempre —sí en la mayoría de los centros— manifiestan ventaja sobre el estrato de nivel avanzado. Por tanto, la productividad léxica en ELE de los estudiantes incrementa a medida que aumenta su nivel de inglés.

b) En CLM

Los animales, Alimentos y bebidas y *Profesiones y oficios* son las áreas temáticas más productivas en CLM en estudiantes de todos los niveles de

inglés. Además, el estrato de nivel básico cuenta con una mayor producción léxica en CLM en *Partes del cuerpo humano* y *Ropa y complementos*; el de nivel intermedio, en *Partes del cuerpo humano* y *La ciudad*; el de nivel avanzado, en *Partes del cuerpo humano* y *El campo*; y el de nivel superior, en *La ciudad* y *El campo*.

Partes de la casa (sin muebles), *Objetos colocados en la mesa para la comida* y *Iluminación y calefacción* tienen en todos los niveles de inglés un menor promedio de palabras por informante en CLM. Los alumnos del nivel básico y avanzado muestran una menor producción en CLM también en *La cocina y sus utensilios* y *Trabajos del campo y del jardín*; los del nivel intermedio, en *Medios de transporte* y *Trabajos del campo y del jardín*; y los del nivel superior, en *Los muebles de la casa* y *La cocina y sus utensilios*.

Comparando estos resultados con los previamente extraídos en ELE según el nivel de inglés, solo *Alimentos y bebidas* es el centro de interés común en ELE y en CLM con un alto grado de productividad léxica; mientras que *Objetos colocados en la mesa para la comida* y *Iluminación y calefacción* son los centros con una menor productividad léxica en todos los niveles de inglés tanto en ELE como en CLM. Otra vez vemos que la productividad en CLM de las áreas semánticas afecta de alguna manera a su productividad en ELE.

Observando la producción léxica en CLM de informantes de los cuatro niveles de inglés en cada centro de interés, advertimos que la productividad léxica en CLM de los estudiantes incrementa a medida que aumenta su nivel de inglés. Esta observación coincide con la obtenida en los resultados en ELE.

c) Entre ELE y CLM

La producción léxica en ELE es inferior a la que hay en CLM en todos los centros de interés y en todos los niveles de inglés, con excepción de *Alimentos y bebidas* del estrato del nivel superior, donde el promedio en ELE supera dicho índice en CLM con 1.5 palabras más.

Los animales es el centro de interés en el que los alumnos de todos los niveles de inglés muestran una debilidad considerable en ELE, tomando como referencia su producción léxica en CLM. Además, hay tres campos léxicos que implican deficiencia en ELE con respecto a la producción en CLM en tres de los estratos: los estratos del nivel básico, intermedio y avanzado, en *Ropa y complementos*; los estratos del nivel intermedio, avanzado y superior, en *Juegos y distracciones*; y los estratos del nivel básico, intermedio y superior, en *Profesiones y oficios*. Como casos únicos, solo los alumnos de nivel básico de inglés presentan debilidad en *Alimentos y bebidas*; solo los de nivel avanzado, en *Los muebles de la casa*; y solo los de nivel superior, en *Objetos colocados en la mesa para la comida y Trabajos del campo y del jardín*.

La diferencia que existe entre la producción léxica en ELE y en CLM va disminuyendo de los estudiantes de niveles de inglés bajos a los de niveles de inglés altos. Dicho de otra manera, los estudiantes generan una producción léxica en ELE cada vez más aproximada a la que tienen en CLM, a medida que van ascendiendo de nivel de inglés.

5) Tipo de centro

Bajo la variable del tipo de centro, comparamos los resultados obtenidos en dos centros públicos y en uno privado.

a) En ELE

Los dos tipos de centro coinciden en cuatro centros de interés con una mayor productividad léxica en ELE: *Alimentos y bebidas*, *La ciudad*, *Partes del cuerpo humano* y *Profesiones y oficios*. Los centros públicos también cuentan con *Los animales* entre sus campos semánticos más productivos en ELE; y el centro privado, con *La escuela: sus muebles y materiales*.

En cuanto a los campos semánticos menos productivos en ELE, los

estratos de centro público y privado coinciden en cuatro: *Iluminación y calefacción*, *Objetos colocados en la mesa para la comida*, *La cocina y sus utensilios* y *Trabajos del campo y del jardín*. Además, los centros públicos cuentan con *Ropa y complementos* entre sus campos nocionales menos productivos en ELE; y el centro privado, con *Partes de la casa (sin muebles)*.

Al comparar la producción léxica en ELE de los dos tipos de centro en cada área nocional, llegamos a ver que, en la inmensa mayoría de los campos semánticos, los centros públicos muestran mayor productividad léxica en ELE que el centro privado. En cuatro áreas nocionales, la ventaja de los centros públicos sobre el privado supone más de una palabra: *Partes del cuerpo humano* (3.02), *Alimentos y bebidas* (2.76), *Los animales* (2.51) y *El campo* (1.02). Creemos que dichos resultados según el tipo de centro favorecen el estrato de los centros públicos.

b) En CLM

Los centros públicos y el centro privado coinciden totalmente en los primeros cinco campos semánticos con una mayor producción léxica en CLM: *Los animales*, *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad*.

En los dos tipos de centro, las áreas nocionales menos productivas en CLM son las mismas, aunque no tienen las mismas posiciones en el rango: *Objetos colocados en la mesa para la comida*, seguido por *Iluminación y calefacción*, *Trabajos del campo y del jardín*, *Partes de la casa (sin muebles)* y *La cocina y sus utensilios*.

Comparando estos resultados con los previamente extraídos en ELE según el tipo de centro, vemos que, en ambos tipos de centro, *Alimentos y bebidas*, *Partes del cuerpo humano*, *Profesiones y oficios* y *La ciudad* son los centros de interés comunes en ELE y en CLM con un alto grado de productividad léxica. Mientas tanto, *Objetos colocados en la mesa para la*

comida, Iluminación y calefacción, Trabajos del campo y del jardín y La cocina y sus utensilios se hallan entre las áreas menos productivas en las dos lenguas. Comprobamos otra vez que las áreas semánticas más y menos productivas en CLM resultan ser también las más y menos productivas en ELE.

Si comparamos la producción léxica en CLM de los dos tipos de centro en cada centro de interés, nos damos cuenta de que, sistemáticamente, los alumnos de los centros públicos disponen de mayor productividad léxica en CLM que los del centro privado. Al igual que en el análisis de ELE, estos resultados en CLM también favorecen el estrato de los centros públicos.

c) Entre ELE y CLM

Otra vez vemos que el promedio de palabras en ELE es inferior al alcanzado en CLM, en todos los centros de interés de ambos tipos de centro.

Existen cinco centros de interés en los que tanto los estudiantes de los centros públicos como los del privado señalan gran divergencia de producción léxica entre las dos lenguas: *Partes del cuerpo humano, Ropa y complementos, La cocina y sus utensilios, Iluminación y calefacción y Los animales*. Esto quiere decir que los estudiantes de ambos tipos de centro muestran una debilidad considerable en estos cinco centros en ELE, tomando como referencia su producción léxica en CLM. También señalamos que solo el estrato de los centros públicos tiene mucha deficiencia en ELE en *El campo, Juegos y distracciones y Profesiones y oficios*; mientras únicamente en el estrato del centro privado, son *Los muebles de la casa y Alimentos y bebidas* en los que la deficiencia de la producción en ELE resulta considerable con respecto a la producción en CLM.

Los alumnos de los centros públicos muestran una mayor distancia de producción léxica entre ELE y CLM que los alumnos del centro privado. Revisando los datos, suponemos que la menor distancia entre las dos lenguas que manifiesta el estrato del centro privado no se traduce en que dicho estrato

cuenta con una ventaja de producción léxica en ELE. En cambio, debido a que los estudiantes del centro privado tienen una productividad léxica demasiado baja en CLM, muestran una ventaja relativa en ELE al tomarse como referencia su productividad léxica en CLM.

IV. Análisis cualitativo general

En el capítulo del análisis cualitativo, nos ocupamos de examinar las convergencias y divergencias conceptuales que existen entre las listas de palabras en ELE y en CLM de cada centro de interés. Para ello, comparamos los vocablos disponibles de los alumnos en ELE y en CLM bajo cada área nocional en busca de los homólogos conceptuales de los vocablos de una lengua en la otra. De este modo, llegamos a ver cuáles son 1) los vocablos que los estudiantes son capaces de actualizar en ELE, aun cuando los conceptos de dichos vocablos no se les ocurran en CLM, 2) los vocablos que solo aparecen en las listas de ELE y que no son aportados en CLM, y —lo más importante— 3) los vocablos que los estudiantes pueden evocar en CLM cuyos conceptos están ausentes en las listas de palabras en ELE.

A lo largo de los apartados anteriores, hablamos de la inferioridad de los resultados en ELE en referencia a los obtenidos en CLM. Pues bien, esta deficiencia puede compensarse, si los alumnos aprenden los homólogos en español de las palabras que solamente son capaces de actualizar en CLM — los vocablos referidos en el punto 3 de arriba—. De ahí viene la aplicación práctica de nuestro análisis cualitativo a la didáctica de ELE en el aula. Por eso, en el apartado del análisis cualitativo general, tras un cotejo pormenorizado entre el léxico altamente disponible (los vocablos con un índice de disponibilidad superior a 0.1) de las dos lenguas, presentamos, bajo cada centro de interés, las palabras que pueden contribuir al fortalecimiento del vocabulario en ELE de los informantes con respecto a su conocimiento léxico en CLM.

V. Análisis cualitativo de la variable “nivel de estudio”

Para el análisis cualitativo de las variables, nos reducimos en esta tesis a analizar las incidencias cualitativas de la variable “nivel de estudio”, teniendo en cuenta la orientación didáctica de nuestro estudio, toda vez que los resultados podrían aplicarse directamente a los estudiantes de diferentes cursos universitarios.

Con el objetivo de conseguir tal meta, por un lado, comparamos los vocablos altamente disponibles en ELE de los cuatro cursos. Así, obtenemos 1) los vocablos en ELE coincidentes en los cursos y 2) los vocablos en ELE ausentes en cada curso, pero presentes en alguno o algunos otros cursos. Estos últimos ya constituyen la deficiencia léxica de los alumnos de un curso con respecto a los de otros cursos.

Por otro lado, cotejamos los vocablos altamente disponibles en ELE y en CLM de cada uno de los cuatro cursos, de modo que descubrimos 1) los conceptos coincidentes en las dos lenguas de cada curso, 2) los conceptos únicamente presentes en ELE de cada curso, y 3) los conceptos presentes en CLM, pero ausentes en ELE de cada curso. Los homólogos en español de estos últimos vocablos únicamente presentes en CLM constituyen el vocabulario que los estudiantes deberían aprender en el aula para compensar su deficiencia léxica en ELE con respecto a su conocimiento léxico en CLM.

Estas son las principales conclusiones del trabajo de investigación *Disponibilidad léxica de estudiantes de ELE en Tianjin, China: estudio comparado entre lengua extranjera y lengua materna*. Así como no hay libro sin erratas, puesto que los libros son humanos, al finalizar el presente estudio, somos conscientes de que podría contener cuestiones que generen dudas, sobre todo, por las limitaciones humanas, que no perjudican el resultado final de la tesis. Seguro que las posibles dudas nos van a motivar en el futuro para que profundicemos aún más el tema.

Sin duda alguna, la presente tesis deja abiertos muchos caminos. La enorme cantidad de datos léxicos en las dos lenguas adjuntos en los apéndices nos proporcionan materiales para análisis de todo tipo. Por ejemplo, desde el enfoque de la relación entre la lengua y la cultura, valdría la pena analizar las nociones altamente disponibles en ambas lenguas para hallar disparidades culturales interesantes. Las demás variables extralingüísticas contempladas en nuestro estudio —sexo, procedencia, nivel de inglés y tipo de centro— también son dignas de ser analizadas cualitativamente. Dados los datos que se han recopilado y procesado bajo una metodología más o menos homogénea, otro tema que se podría abordar es una comparación entre distintos conjuntos de informantes que compartan perfiles parecidos. Sería interesante analizar las diferencias de disponibilidad léxica en estudiantes de ELE de países de origen diferentes. Por otra parte, una comparación entre el léxico disponible de estudiantes de ELE y el de otros nativos que manejan el español como lengua materna daría más sugerencias sobre la planificación de la enseñanza del vocabulario en el aula de ELE. Queríamos destacar que aquí no acaba nuestra labor como investigadores, sino todo lo contrario, que este es el principio.

8. BIBLIOGRAFÍA

- Aabidi, L. (2018). Evolución del léxico disponible en español como lengua extranjera (ELE) de los alumnos marroquíes de la enseñanza media. *Dirāsāt Hispānicas: Revista Tunecina de Estudios Hispánicos*, (5), 7-19.
- Aabidi, L. (2019). La disponibilidad léxica en español como lengua extranjera: dos décadas de investigación científica. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (28), 1-25.
- Ahumada Lara, I. (2006). *El léxico disponible de los estudiantes preuniversitarios de la provincia de Jaén*. Jaén: Publicaciones de la Universidad de Jaén.
- Alba, O. (1995a). Anglicismos léxicos en el español dominicano: análisis cuantitativo. En *El español dominicano dentro del contexto americano* (pp. 11-38). Santo Domingo: La Trinitaria.
- Alba, O. (1995b). *El léxico disponible de la República Dominicana*. Santiago de los Caballeros: Pontificia Universidad Católica Madre y Maestra.
- Alba, O. (1998). Variable léxica y dialectología hispánica. *La Torre: Revista de la Universidad de Puerto Rico*, 3(7-8), 299-316.
- Alba, O. (1999). Densidad de anglicismos en el léxico disponible de la República Dominicana. En *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina Las Palmas de Gran Canaria* (Vol. 2, pp. 853-866). Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- Alvar Ezquerro, M. (2005). La frecuencia léxica y su utilidad en la enseñanza del español como lengua extranjera. En *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad: Actas del XV Congreso Internacional de ASELE* (pp. 19-39). Sevilla.

- Areta Lara, M. (2009). *El léxico disponible de los estudiantes preuniversitarios navarros. Estudio sociolingüístico*. Pamplona: Universidad de Navarra.
- Arnal Purroy, M. L. (2004). *Léxico disponible de Aragón*. Zaragoza: Libros Pórtico.
- Arnal Purroy, M. L. (2008a). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Zaragoza: Institución Fernando el Católico.
- Arnal Purroy, M. L. (2008b). Los dialectalismos en el léxico disponible de los jóvenes aragoneses. En *Estudios sobre disponibilidad léxica en los jóvenes aragoneses* (pp. 17-49). Zaragoza: Institución "Fernando el Católico".
- Arnal Purroy, M. L. (2008c). Niveles socioculturales y léxico dialectal en el vocabulario disponible de Aragón. En *Discurso y sociedad II: nuevas contribuciones al estudio de la lengua en contexto social* (pp. 457-470). Castellón de la Plana: Universitat Jaume I, Servei de Comunicació i Publicacions.
- Ávila Martín, M. del C. (2010). Estadística y lingüística de corpus: implicaciones pedagógicas en la enseñanza y el aprendizaje del léxico. *Cauce. Revista internacional de Filología, Comunicación y sus Didácticas*, (33), 163-175.
- Ávila Muñoz, A. M. (1999). *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Málaga: Universidad de Málaga.
- Ávila Muñoz, A. M. (2006). *Léxico disponible de los estudiantes preuniversitarios de Málaga*. Málaga: Universidad de Málaga.
- Ávila Muñoz, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos. *Journal of Spanish Language Teaching*, 3. <https://doi.org/10.1080/23247797.2016.1163038>

- Ávila Muñoz, A. M. (2017). La utilidad del vocabulario dialectal en el aula de lenguas extranjeras: Propuesta de selección léxica basada en coronas concéntricas. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (25).
- Ayora Esteban, M. del C. (2004). Deficiencias en la adquisición del léxico como problema específico de las lenguas en contacto. En *Actas del V Congreso de Lingüística General: León 5-8 de marzo de 2002* (Vol. 1, pp. 281-293).
- Ayora Esteban, M. del C. (2006). *Disponibilidad léxica en Ceuta: aspectos sociolingüísticos*. Cádiz: Universidad de Cádiz.
- Azurmendi Ayerbe, M. J. (1982). *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la Comarca de San Sebastián*. Madrid: Universidad Complutense de Madrid.
- Bartol Hernández, J. A. (2001). Reflexiones sobre la disponibilidad léxica. En *Nuevas aportaciones al estudio de la lengua española: investigaciones filológicas* (pp. 221-235).
- Bartol Hernández, J. A. (2002). Léxico disponible de la provincia de Soria: Primeros datos. *Archivo de filología aragonesa*, (59), 1075-1104.
- Bartol Hernández, J. A. (2003). Léxico disponible y norma lingüística. En *Lengua, variación y contexto: estudios dedicados a Humberto López Morales* (Vol. 1, pp. 127-143).
- Bartol Hernández, J. A. (2004). *Léxico disponible de Soria: estudio y diccionarios*. Burgos: Fundación Instituto Castellano y Leonés de la Lengua.
- Bartol Hernández, J. A. (2006). La disponibilidad léxica. *Revista española de lingüística*, 36(1), 379-384.

- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En *De moneda nunca usada: Estudios dedicados a José M^a Enguita Utrilla* (pp. 85-107). Zaragoza: Insituto Fernando El Católico, CSIC.
- Bartol Hernández, J. A., & Borrego Nieto, J. (2019). El léxico disponible de la ropa en Cádiz, Valladolid y el conjunto de España: norma y variación. *Ogigia: Revista electrónica de estudios hispánicos*, (25), 57-82.
- Bellón Fernández, J. J. (2011). *Léxico disponible de la provincia de Córdoba*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- Bellón Fernández, J. J., Samper Hernández, M., & Samper Padilla, J. A. (2003). El proyecto de estudio de la disponibilidad léxica en español. En *Pautas y pistas en el estudio del léxico hispano (americano)* (pp. 27-140). Madrid: Iberoamericana: Verbuert.
- Benítez Pérez, P. (1991). Reseña a Azurmendi Ayerbe, María José, Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la Comarca de San Sebastián. *Revista española de lingüística*, 21(2), 355-357.
- Benítez Pérez, P. (1992a). Disponibilidad léxica en la zona metropolitana de Madrid. *Boletín de la Academia Puertorriqueña de la Lengua Española*, 1(1), 71-102.
- Benítez Pérez, P. (1992b). Listas abiertas y listas cerradas de palabras en disponibilidad léxica. En *Actas del VIII Congreso Nacional de la Asociación Española de Lingüística Aplicada* (pp. 117-127). Vigo: AESLA.
- Benítez Pérez, P. (1994). Léxico real / léxico irreal en los manuales de español para extranjeros. En *Actas del II Congreso Nacional de ASELE Español para extranjeros: didáctica e investigación* (pp. 325-333). Málaga: ASELE.

- Benítez Pérez, P., & Zebrowski, J. (1993). El léxico español en los manuales polacos. En *El español como lengua extranjera, de la teoría al aula: actas del tercer Congreso Nacional de ASELE* (pp. 223-230). Málaga.
- Borrego Nieto, J., & Fernández Juncal, M. del C. (2002). Léxico disponible: aplicaciones a los estudios dialectales. En *IV Congreso de Lingüística General: Cádiz del 3 al 6 de abril 2000* (Vol. 2, pp. 297-306). Cádiz: Universidad de Cádiz.
- Buchanan, M. A. (1929). *A graded spanish word book*. Toronto: University of Toronto Press.
- Calero Fernández, M. Á., & Chen, L. (2021). Léxico disponible en estudiantes de ELE en Tianjin por centros de interés frente a otras sintopías euroasiáticas. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 63-79). Valencia: Tirant lo Blanch.
- Calero Fernández, M. Á., & Chen, L. (en prensa). “Disponibilidad léxica y planificación de la enseñanza del vocabulario en estudiantes de ELE sinohablantes”. *Tejuelo*, (34).
- Calero Fernández, M. Á., & Serrano Zapata, M. (2017). Transferencias léxicas del catalán al español de Lleida. *Linred: Lingüística en la Red*, (15).
- Calero Fernández, M. Á., & Serrano Zapata, M. (2019). Incidencia del factor sexo en el léxico disponible de una comunidad bilingüe. *Ogigia: Revista electrónica de estudios hispánicos*, (25), 83-107.
- Callealta Barroso, F. J., & Gallego Gallego, D. J. (2016). Medidas de disponibilidad léxica: comparabilidad y normalización. *Boletín de filología (Universidad de Chile)*, 51(1), 39-92.

- Cañizal Arévalo, A. V. (1987). *Disponibilidad léxica en escolares de Primaria terminada. Análisis de seis centros de interés*. Tesis de licenciatura inédita. México D.F.: UNAM.
- Carcedo González, A. (1998a). Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE. *RILCE: Revista de filología hispánica*, 14(2), 205-224.
- Carcedo González, A. (1998b). Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica. *Lingüística*, (10), 5-68.
- Carcedo González, A. (1999). Análisis de errores léxicos del español en la interlengua de los finlandeses. En *Español como lengua extranjera, enfoque comunicativo y gramática: actas del IX congreso internacional de ASELE* (pp. 465-472). Santiago de Compostela.
- Carcedo González, A. (2000). *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Turun Yliopisto.
- Carcedo González, A. (2001). *Léxico disponible de Asturias*. Turku: Publicaciones del Departamento de Lengua Española de la Universidad de Turku.
- Carrasco Arroyo, S. (2005). *Aproximación a la estadística desde las ciencias sociales*. Valencia: Universidad de Valencia.
- Casanova Ávalos, M. (2006). La disponibilidad léxica en la Comunidad Valenciana. En *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social* (pp. 737-752). Castellón de la Plana: Universitat Jaume I, Servei de Comunicació i Publicacions.
- Casanova Ávalos, M. (2008). Anglicismos en el léxico disponible de Castellón. En *Discurso y sociedad II: nuevas contribuciones al estudio de la lengua*

en contexto social (pp. 471-500). Castellón de la Plana: Universitat Jaume I, Servei de Comunicació i Publicacions.

Casanova Ávalos, M. (2013). Valencianismos en el léxico disponible de Castellón. En *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas* (Vol. 8, pp. 531-543). Berlín: De Gruyter.

Casanova Ávalos, M. (2017). *Léxico disponible de Castellón. Estudio y diccionarios*. Castellón de la Plana: Universitat Jaume I.

Castillo Fadic, M. N., & Santos Díaz, I. C. (2021). Dos miradas sobre el cuerpo humano en el marco de la literacidad en salud: profesionales de la salud y personas con diabetes mellitus 2. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 273-289). Valencia: Tirant lo Blanch.

Chen, L., & Calero Fernández, M. Á. (2021). Disponibilidad léxica de los estudiantes de español en Tianjin en comparación con otras sintopías en el contexto de aprendizaje de lenguas extranjeras: Datos generales. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 45-62). Valencia: Tirant lo Blanch.

Cruz Alonso, R. (2012). *El léxico disponible de Castilla y León*. Salamanca: Universidad de Salamanca.

de la Fuente, M. H. (2015). *Estudio de disponibilidad léxica de los hablantes de herencia alumnos de CEHI (Madrid)*. Madrid: Universidad Nebrija.

de la Granja, M. Á., & López Meirama, B. (2021). La presencia del español en el léxico disponible del gallego. El centro de interés el Cuerpo Humano. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 115-145). Valencia: Tirant lo Blanch.

de la Maya Retamar, G., & López Pérez, M. (2021). Léxico disponible en inglés de maestros en formación: incidencia del conocimiento de lenguas. En M.

S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 99-114). Valencia: Tirant lo Blanch.

del Barrio de la Rosa, F. (2016). Algunas observaciones sobre la disponibilidad léxica en estudiantes itálofonos de español. En *Geométrica explosión Estudios de lengua y literatura en homenaje a René Lenarduzzi* (pp. 127-144). Venezia: Ca' Foscari.

del Barrio de la Rosa, F. (2017). Los estudiantes itálofonos de ELE y los estudios de disponibilidad léxica en español. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, (22), 52-57.

Dimitrijević, N. R. (1969). *Lexical availability: A new aspect of the lexical availability of secondary school children*. Heidelberg: Julius Groos Verlag.

Dimitrijević, N. R. (1981). A comparative study of the lexical availability of monolingual and bilingual schoolchildren. *Studia Anglica Posnaniensia*, (13), 109-130.

Echeverría, M. S., Herrera, M. O., Moreno, P., & Pradenas, F. (1987). Disponibilidad léxica en Educación Media. *RLA: Revista de lingüística teórica y aplicada*, (25), 55-116.

Echeverría, M. S., Vargas, R., Urzúa, P., & Ferreira, R. (2008). DispoGrafo: una nueva herramienta computacional para el análisis de relaciones semánticas en el léxico disponible. *RLA: Revista de lingüística teórica y aplicada*, (46), 81-92.

Etxebarria Arostegui, M. (1996). Disponibilidad léxica en escolares del País Vasco. *Revista española de lingüística*, 26(2), 301-326.

Etxebarria Arostegui, M. (1997). Disponibilidad léxica y caracterización sociolingüística: variación y diversidad. En *Estudios de lingüística general: conferencias y trabajos presentados en el II Congreso Nacional de Lingüística General* (Vol. 2, pp. 65-84). Granada: Método Ediciones.

- Etxebarria Arostegui, M. (1999). Disponibilidad léxica y enseñanza de la lengua materna en el área metropolitana del Gran Bilbao. En *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina Las Palmas de Gran Canaria* (Vol. 2, pp. 1479-1494). Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones, 1999.
- Etxebarria Arostegui, M. (2012). Disponibilidad léxica en el castellano de los escolares del País Vasco: variación y modelos de enseñanza bilingüe. En *El castellano del País Vasco* (pp. 65-86). Bilbao: Universidad del País Vasco-Euskal Herriko Unibertsitatea.
- Ezquerria, R. (1974). Los diccionarios de frecuencia en español. En *Boletín de la Asociación Europea de Profesores de Español Año VI. Núm. 10. Marzo, 1974* (pp. 3-27). Madrid: Centro Virtual Cervantes.
- Fernández dos Santos, M. (2014). Disponibilidad léxica en alumnos de español lengua extranjera del Distrito de Oporto (Portugal). En *El español como lengua extranjera en Portugal: retos de la enseñanza de lenguas cercanas* (pp. 92-104).
- Fernández Juncal, C. (2008). *Léxico disponible de Burgos*. Burgos: Instituto Castellano y Leonés de la Lengua.
- Fernández Juncal, C. (2013). *Léxico disponible de Cantabria: estudio sociológico*. Salamanca: Ediciones Universidad de Salamanca.
- Fernández Juncal, C., & Hernández Muñoz, N. (2021). Antropónimos disponibles: consolidación de modelos socionomásticos. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 247-271). Valencia: Tirant lo Blanch.
- Fernández Merino Gutiérrez, P. V. (2011). Disponibilidad léxica de inmigrantes: propuesta para una necesidad. *Lengua y migración*, 3(2), 83-106.

- Fernández Orío, S., & Jiménez Catalán, R. M. (2015). Lexical availability of EFL learners at the end of Spanish secondary education: The effect of language program and prompt. *ES: Revista de filología inglesa*, (36), 103-127.
- Fernández Smith, G., Rico Martín, A. M., Molina García, M. J., & Jiménez Jiménez, M. Á. (2008). *Léxico disponible de Melilla: estudio sociolingüístico y repertorios léxicos*. Madrid: Arco Libros.
- Ferreira Cabrera, A., Salcedo Lagos, P., & del Valle Leo, M. (2014). Estudio de disponibilidad léxica en el ámbito de las matemáticas. *Estudios filológicos*, (54), 69-84.
- Fuster Pérez, N. (2013). *El léxico disponible de jóvenes puertorriqueños: Reflejo sociocultural de dos épocas*. León: Universidad de León.
- Gallego Gallego, D. J. (2014). Léxico disponible de 82 inmigrantes estudiantes de español en la ciudad de Alcalá de Henares. *Lengua y migración*, 6(2), 95-123.
- Galoso Camacho, M. V. (1997a). El léxico disponible en el nivel preuniversitario (provincia de Zamora). *Interlingüística*, (7), 75-80.
- Galoso Camacho, M. V. (1997b). Tratamiento y perspectivas del léxico disponible: el caso de la provincia de Zamora. *REALE: revista de estudios de adquisición de la lengua española*, (8), 23-34.
- Galoso Camacho, M. V. (1999). Hacia una definición de léxico disponible: posibles aplicaciones. En *Lingüística para el siglo XXI: III Congreso organizado por el Departamento de Lengua Española* (pp. 711-718). Salamanca: Ediciones Universidad de Salamanca.
- Galoso Camacho, M. V. (2001). *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Salamanca: Universidad de Salamanca.

- Galoso Camacho, M. V. (2002a). *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Salamanca: Ediciones Universidad de Salamanca.
- Galoso Camacho, M. V. (2002b). ¿Para qué sirven los estudios de léxico disponible? En *IV Congreso de Lingüística General: Cádiz del 3 al 6 de abril 2000* (Vol. 3, pp. 1143-1152). Cádiz: Universidad de Cádiz.
- Galoso Camacho, M. V. (2003a). *El léxico disponible de Ávila, Salamanca y Zamora*. Burgos: Instituto Castellano y Leonés de la Lengua.
- Galoso Camacho, M. V. (2003b). El léxico disponible de los estudiantes preuniversitarios en Salamanca. *Salamanca: revista de estudios*, (50), 201-224.
- Galoso Camacho, M. V., & Prado Aragonés, J. (2005). La estructura estadística del léxico disponible de informantes nativos e informantes de ELE. En *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad: Actas del XV Congreso Internacional de ASELE* (pp. 370-375). Sevilla.
- Gamazo Carretero, E. (2014). Estereotipos en el léxico disponible de universitarios portugueses. En *El español como lengua extranjera en Portugal: retos de la enseñanza de lenguas cercanas* (pp. 28-41).
- García Casero, M. J. (2013). *El léxico disponible en estudiantes de 4º de Educación Secundaria Obligatoria en Santander*.
- García Domínguez, M. J., Marrero Pulido, V., Pérez Martín, J. A., & Piñero Piñero, G. (1994). Estudio de la disponibilidad léxica en Gran Canaria: la variable geográfica y el tipo de educación. *REALE: Revista de Estudios de Adquisición de la Lengua Española*, (2), 65-72.

- García Marcos, F. J., & López Morales, H. (1995). Disponibilidad léxica en Andalucía: proyecto de investigación. *REALE: revista de estudios de adquisición de la lengua española*, (3), 67-78.
- García Marcos, F. J., Mateo García, M. V., & Fuentes González, A. D. (1994). Indices de mortandad del léxico cañero en la costa granadina. En *1492: lo dulce a la conquista de Europa: actas del Cuarto Seminario Internacional sobre la caña de azúcar, Motril, 21-25 septiembre de 1992* (pp. 219-254).
- García Megía, A. (2003). *La disponibilidad léxica en la ciudad de Almería entre los nueve y los doce años*. Almería: Universidad de Almería.
- Garzón Segura, A. M., & Penagos López, L. A. (2016). Disponibilidad léxica en estudiantes de primer semestre de pregrado de una institución universitaria de Villavicencio, Colombia. *Forma y función*, 29(2), 63-84.
- Gimeno Menéndez, F. (2002). Alberto Carcedo González, Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición), Turku: *Annales Universitatis Turkuensis, Humaniora. Revista de Filología de la Universidad de La Laguna*, (20), 354-358.
- Gladkiĭ, A. V., & Mel'čuk, I. A. (1972). *Introducción a la lingüística matemática*. Barcelona: Planeta.
- Gómez Devís, M. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Valencia: Universitat de València.
- Gómez Devís, M. B. (2010). Aportes de la disponibilidad léxica a la investigación sobre fenómenos de contacto lingüístico. En *Interacció comunicativa i ensenyament de llengües* (pp. 233-240). Universitat de València, 2010.

- Gómez Devís, M. B., & Serrano Zapata, M. (2021). Variación léxica en el catalán occidental a partir del léxico disponible de estudiantes preuniversitarios: la influencia del español. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 147-170). Valencia: Tirant lo Blanch.
- González Fernández, J. (2013). La disponibilidad léxica de los estudiantes turcos de español como lengua extranjera. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (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 a la enseñanza de Lenguas*, 8(16), 41-53.
- González Martínez, A. (2002). *La disponibilidad léxica de los alumnos preuniversitarios de la provincia de Cádiz*. Cádiz: Universidad de Cádiz.
- González Martínez, A., & Orellana Ramírez, P. (2000). Cotejo de los léxicos disponibles de Cádiz y Zamora. *Documentos del Español Actual (DEA)*, 2, 139-160.
- Gómez Molina, J. R. (2021). La disponibilidad léxica y sus aplicaciones cognitivas. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 224-226). Valencia: Tirant lo Blanch.
- Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1964). *L'élaboration du français fondamental (1er degré): étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. París: Didier.
- Gutiérrez Cabria, S. (1994). *Filosofía de la estadística*. Valencia: Universitat de València.
- Henmon, V. A. C. (1924). *A french word book based on a count of 400,000 running words*. Wisconsin: Madison, Wis.

- Hernández Muñoz, N. (2004). *El léxico disponible de los estudiantes conquenses*. Salamanca: Ediciones Universidad de Salamanca.
- Hernández Muñoz, N. (2006). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*. Salamanca: Ediciones Universidad de Salamanca.
- Hernández Muñoz, N. (2009). Variación léxica y zonas dialectales de Castilla-La Mancha. *Revista de Filología Española*, LXXXIX(2), 279-300.
- Hernández Muñoz, N. (2015a). ¿Es el conocimiento morfológico un mecanismo determinante en la recuperación del léxico disponible? En *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General* (pp. 259-267). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- Hernández Muñoz, N. (2015b). 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., & Bartol Hernández, J. A. (2006). Proyecto del léxico disponible en España. En *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social* (pp. 725-736).
- Hernández Muñoz, N., & Tomé Cornejo, C. (2017). Léxico disponible en primera y segunda lengua: bases cognitivas. En *Palabras Vocabulario Léxico: La lexicología aplicada a la didáctica y a la diacronía* (pp. 99-122). Venezia: Edizioni Ca' Foscari.
- Herranz Llácer, C. V. (2018a). Comparación del léxico disponible. Español de futuros docentes que estudian en modalidad bilingüe y no bilingüe. *Miscelánea Comillas: Revista de Ciencias Humanas y Sociales*, 76(149), 529-553.

- Herranz Llácer, C. V. (2018b). Disponibilidad léxica de los futuros profesores de Educación Infantil y Primaria. *Revista electrónica interuniversitaria de formación del profesorado*, 21(1), 143-159.
- Herranz Llácer, C. V. (2018c). El nivel socioeconómico y el léxico disponible de futuros docentes de educación infantil y educación primaria. *E-AESLA. REVISTA DIGITAL*, 4, 294-304.
- Herranz Llácer, C. V., & Marcos Calvo, M. A. (2019). Análisis del léxico disponible español de extranjeros que estudian los grados de educación. *Ogigia: Revista electrónica de estudios hispánicos*, (26), 5-30.
- Herreros Marcilia, M. (2015a). La disponibilidad léxica como propuesta de investigación docente: Análisis en alumnos inmigrantes no hispanohablantes. En *La enseñanza de ELE centrada en el alumno* (pp. 459-468). Getafe: ASELE.
- Herreros Marcilla, M. (2015b). Los centros de carácter cultural en los estudios de disponibilidad léxica: análisis y una propuesta. En *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General* (pp. 279-290). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- Hidalgo Gallardo, M. (2017a). Evolución diacrónica del léxico disponible de estudiantes sinohablantes de ELE. *SinoELE*, (16), 1-16.
- Hidalgo Gallardo, M. (2017b). La disponibilidad léxica como herramienta de evaluación interventiva del vocabulario en ELE: aplicación en un contexto universitario sinohablante. *SinoELE*, (16), 43-63.
- Hidalgo Gallardo, M. (2017c). *La disponibilidad léxica como método de detección del vocabulario y de su selección en manuales. Aplicación en una muestra de estudiantes sinohablantes de ELE*. Jaén: Universidad de Jaén.

- Hidalgo Gallardo, M. (2017d). Sobre la disponibilidad léxica en ELE. Revisión de la literatura. *Boletín ASELE*, (56), 83-94.
- Hidalgo Gallardo, M. (2019a). El perfil léxico del manual de ELE *Español Moderno 1* desde la óptica de la disponibilidad léxica: posibles mejoras para la selección del vocabulario. *Ogigia: Revista electrónica de estudios hispánicos*, (25), 233-257.
- Hidalgo Gallardo, M. (2019b). *Factores y fuentes que inciden en el léxico disponible de estudiantes sinohablantes de ELE*. ASELE.
- Hoz, V. G. (1953). *Vocabulario usual, vocabulario común y vocabulario fundamental: determinación y análisis de sus factores*. Madrid: Instituto San José de Calasanz.
- Juilland, A. G., & Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words*. Paris: Den Haag.
- Justo Hernández, H. G. (1987). *Disponibilidad léxica en colores*. Tesis de licenciatura inédita. México D.F.: UNAM.
- Kaeding, F. W. (1898). *Häufigkeitwörterbuch der deutschen Sprache: Festgestellt durch einen Arbeitsausschuss der deutschen Stenographiesysteme*. Berlín: Selbstverlag des Herausgebers; E.S. Mittler & Sohn.
- Keniston, H. (1920). Common words in spanish. *Hispania*, 3(2), 85-96.
- Keniston, H. (1929). *Spanish idiom list: selected on the basis of range and frequency of occurrence*. New York: MacMillan.
- Lin, J. (2012). El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (14).

- Llorente Pinto, M. del R. (2005). Léxico disponible y léxico dialectal en la provincia de Ávila. En *Palabras, norma, discurso: en memoria de Fernando Lázaro Carreter* (pp. 681-694). Salamanca: Universidad de Salamanca, 2005.
- López Chávez, J. (1992). Alcances panhispánicos del léxico disponible. *Lingüística*, (4), 26-124.
- López Chávez, J. (1994). Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica. *REALE: revista de estudios de adquisición de la lengua española*, (1), 67-84.
- López Chávez, J., & Pérez Durán, M. A. (2014). Los estudios de la disponibilidad léxica en México desde el año 2000. *Tonos digital: Revista de estudios filológicos*, (26).
- López Chávez, J., & Rodríguez Fonseca, L. (1992). Un modelo para la programación de la enseñanza del vocabulario. En *Actas del I Congreso Internacional sobre la Enseñanza del Español*. Madrid: CEMIP.
- López Chávez, J., & Strassburger Frías, C. (1991). Un modelo para el cálculo del índice de disponibilidad léxica individual. En *La enseñanza del español como lengua materna* (pp. 91-112). Río Piedras: Universidad de Puerto Rico.
- López Chávez, J., & Strassburger Frías, C. (2000). El diseño de una fórmula para obtener un índice de disponibilidad léxica confiable. *Anuario de Letras*, 38, 227-251.
- López González, A. M. (2009). La disponibilidad léxica en las secciones bilingües de español en Polonia. En *El profesor de español LE-L2: Actas del XIX Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE)* (Vol. 2, pp. 567-582). Cáceres: Universidad de Extremadura, Servicio de Publicaciones.

- López González, A. M. (2014). Desarrollo de los estudios de disponibilidad léxica en español lengua extranjera (ELE). En *La enseñanza del Español como LE/L2 en el siglo XXI* (pp. 397-408). Madrid: ASELE.
- López González, A. M. (2016). Las asociaciones léxicas en el léxico disponible en lengua materna y en lengua extranjera. *Tonos digital: Revista de estudios filológicos*, (31).
- López Meirama, B. (2008). *Léxico disponible en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela.
- López Meirama, B. (2011). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela.
- López Morales, H. (1973). *Disponibilidad léxica en escolares de San Juan*. San Juan: Universidad de Puerto Rico (manuscrito).
- López Morales, H. (1978). Frecuencia léxica, disponibilidad y programación curricular. *Boletín de la Academia Puertorriqueña de la Lengua Española*, 6(1), 73-86.
- López Morales, H. (1979). Disponibilidad léxica y estratificación socioeconómica. En *Dialectología y Sociolingüística. Temas puertorriqueños*. (pp. 173-181). Madrid: Hispanova de Ediciones.
- López Morales, H. (1986). *La enseñanza de la lengua materna: lingüística para maestros de español*. Madrid: Playor.
- López Morales, H. (1994). *Métodos de investigación lingüística*. Salamanca: Colegio de España.
- López Morales, H. (1995). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología*, 35(1), 245-259.

- López Morales, H. (1999a). Anglicismo en el léxico disponible de Puerto Rico. En *El Caribe hispánico: perspectivas lingüísticas actuales. Homenaje a Manuel Álvarez Nazario* (pp. 147-170). Frankfurt-Madrid: Vervuert-Iberoamericana.
- López Morales, H. (1999b). *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.
- López Morales, H., & Trigo Ibáñez, E. (2019). La disponibilidad léxica: tendencias actuales y perspectivas de futuro. *Ogigia: Revista electrónica de estudios hispánicos*, (25), 7-10.
- López Rivero, E. (2008). *Estudio de disponibilidad léxica en 43 estudiantes de ELE*. Madrid: Universidad Nebrija.
- Lorán Santos, R. & López Morales, H. (1983). *Nouveau calcul de l'indice de disponibilité*. Puerto Rico: MS.
- Mackey, W. F., Savard, J. G., & Ardouin, P. (1971). *Le vocabulaire disponible du français. Tome I: Le vocabulaire concret usuel des enfants français et acadiens: étude témoin. Tome II: Vocabulaire disponible des enfants acadiens*. Paris-Bruxelles-Montréal: Didier.
- Magnúsdóttir, S. (2012). *Disponibilidad léxica en alumnos de español como lengua extranjera. Estudio sobre el léxico disponible en alumnos de ELE en la secundaria en Islandia*. Reikiavik: Háskóli Íslands.
- Mangado Cruz, M., & Areta Lara, M. (2008). Procesamiento informático de datos para la elaboración de diccionarios de disponibilidad léxica. En *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística* (pp. 479-494). Pamplona: Universidad de Navarra.
- Manning, C. D., & Schütze, H. (2000). *Foundations of statistical natural language processing*. London: The MIT Press.

- Marcos Calvo, M. A., & Herranz Llácer, C. V. (2021). Porcentaje de disponibilidad léxica retenida: análisis del vocabulario de futuros maestros. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 81-97). Valencia: Tirant lo Blanch.
- Martínez Olmos, E. (2008). *Disponibilidad léxica en las comunidades de habla alicantinas*. Alicante: Universidad de Alicante.
- Martínez Segura, G. E. (2012). *Disponibilidad léxica de las/los estudiantes de II de bachillerato académico, sección «A» del Instituto 1º de Mayo de 1954. Una propuesta para la enseñanza del vocabulario*. Alicante: Biblioteca Virtual Miguel de Cervantes.
- Massa, E. (2016). Stories of words, words as stories. Some lexico-statistically based reflections on the meaning unit in spoken language. *Linguistik Online*, 75(1).
- Mateo García, M. V. (1997). Léxico-estadística y lingüística aplicada. En *Estudios de lingüística general: conferencias [y] trabajos presentados en el II Congreso Nacional de Lingüística General: Granada, 25 al 27 de marzo de 1996* (Vol. 2, pp. 253-262). Granada: Método Ediciones.
- Mateo García, M. V. (1998). *Disponibilidad léxica en el C.O.U. almeriense: estudio de estratificación social*. Almería: Universidad de Almería.
- Matéus Ferro, G. E., & Santiago Galvis, Á. W. (2006). Disponibilidad léxica en estudiantes bogotanos. *Folios: revista de la Facultad de Humanidades*, (24), 3-26.
- Mena Osorio, M. (1986). *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Tesis de maestría inédita. Concepción: Universidad de Concepción.

- Mendoza Puertas, J. D. (2018). El léxico disponible de 82 estudiantes coreanos de español como lengua extranjera. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (26).
- Miguel Ojeda, M., Díaz Camacho, J. E., Apodaca Victoria, C., & Trujillo Landa, I. (2004). *Metodología de diseño estadístico*. Xalapa: Universidad Veracruzana.
- Moliler, M. (1998). *Diccionario de uso del español (DUE)* Madrid: Gredos.
- Morales, A. (1986). *Léxico básico del español de Puerto Rico*. San Juan de Puerto Rico: Academia Puertorriqueña de la Lengua Española.
- 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 a la enseñanza de Lenguas*, 6(11), 53-61.
- Müller, C. (1973). *Estadística lingüística*. Madrid: Editorial Gredos, S.A.
- Murillo Rojas, M. (1993). Disponibilidad léxica en los preescolares: estudio de cinco campos semánticos, *Káñina*, XVII (2), 117-127.
- Murillo Rojas, M. (1994). Comidas y bebidas: estudio de disponibilidad léxica en preescolares, *Káñina*, XVIII (2), 117-133.
- Murillo Rojas, M. (1999). Crecimiento de la disponibilidad léxica: niños de preescolar y primer ciclo de la educación básica costarricense. *Filología y Lingüística*, XXV(2), 187-203.
- Nalesso, G. (2018). El desarrollo de la competencia léxica de estudiantes italianos universitarios de ELE. *Orillas - Rivista d'Isanistica*, 381-394.
- Navarro Marrero, Y. (2016). Léxico disponible de inmigrantes en Castilla y León. Estudio y diccionarios. *Lengua y migración*, 8(1), 111-116.

- Njock, P. E. (1978). *L'univers familial de l'enfant africain*. Quebec: Centre International de Recherche sur le Bilinguisme. Universidad de Laval.
- Ó Huallacháin, C. (1966). *Buntús gaeilge*. Baile Átha Cliath (Dublin): Oifig Dhiolta Foilseachán Rialtais (Government Publications Office).
- Ortolano Ríos, B. (2005). Estudios de disponibilidad léxica sobre una muestra de alumnos de Ayamonte (Huelva). *Tonos digital: Revista de estudios filológicos*, (9).
- Pacheco, C. R., Cabrera, J. S., & López, I. G. (2017). Incidencia de la variable “sexo” en la disponibilidad léxica de los estudiantes de preuniversitario de Pinar del Río, Cuba. *Íkala*, 22(2), 237-253.
- Palapanidi, K. (2012). La aplicación de la disponibilidad léxica a la didáctica del léxico de LE. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, 6(11), 62-70.
- Palapanidi, K. (2017). La disponibilidad léxica de aprendientes griegos de español como lengua extranjera en diferentes niveles lingüísticos. En *Estudios y homenajes hispanoamericanos IV*. Madrid: Ediciones Clásicas, S.A.
- Palapanidi, K., & Agustín Llach, M. P. (2017). Análisis comparativo de la disponibilidad léxica de aprendientes griegos de español como lengua extranjera con la de estudiantes nativos. En *Estudios y homenajes hispanoamericanos V* (pp. 297-305). Madrid: Ediciones Clásicas, S.A.
- Paredes García, F. (1999). La ortografía en las encuestas de disponibilidad léxica. *REALE: revista de estudios de adquisición de la lengua española*, (11), 75-98.
- Paredes García, F. (2001). Disponibilidad léxica de alumnos de enseñanza secundaria de Alcalá y su comarca: resultados generales. En *La lingüística aplicada a finales de siglo XX: ensayos y propuestas* (Vol. 2, pp. 721-728).

- Paredes García, F. (2012). Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, 6(11), 70-94.
- 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 a la enseñanza de Lenguas*, 8(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: Lingüística en la Red*, (13).
- Paredes García, F. (2017). La comparabilidad de los trabajos de disponibilidad léxica. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, (22), 71-77.
- Paredes García, F., & Gallego Gallego, D. J. (2019). Procedimientos neológicos en el léxico disponible de español como lengua materna y como lengua extranjera. *Ogigia: Revista electrónica de estudios hispánicos*, (25), 109-138.
- Pastor Milán, M. Á., & Sánchez García, F. J. (2008). *El léxico disponible de Granada y su provincia*. Granada: Universidad de Granada.
- Payrató Giménez, L. (1998). *De profesión, lingüista: panorama de la lingüística aplicada*. Barcelona: Ariel.
- Pedroni Torres, T. (2015). *El léxico disponible de los profesores de ELE en formación en la ciudad de São Paulo (Brasil)*. Salamanca: Universidad de Salamanca.
- Pérez Juste, R., García Llamas, J. L., Gil Pascual, J. A., & Galán González, A. (2009). *Estadística aplicada a la educación*. Madrid: Pearson Educación.

- 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»*. Madrid: Universidad Internacional Menéndez Pelayo.
- Pfeffer, J. A. (1964). *Basic (spoken) german word list: Grundstufe*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Prado Aragonés, J., & Galloso Camacho, M. V. (2005). *Léxico disponible de Huelva: nivel preuniversitario*. Huelva: Universidad de Huelva.
- Prado Aragonés, J., Galloso Camacho, M. V., & Conceição, M. C. (2010). *La Disponibilidad léxica en situación de contacto de lenguas en las zonas limítrofes de Andalucía y Extremadura (España) y Algarve y Alentejo (Portugal)*. Huelva: Universidad de Huelva.
- Real Academia Española. (2014). *Diccionario de la Lengua Española*. Madrid: Espasa, 23.^a edición. <https://dle.rae.es>.
- Ríos González, G. (2011). *Características del lenguaje de los jóvenes costarricenses desde la disponibilidad léxica*. Salamanca: Ediciones Universidad de Salamanca.
- Rodríguez Bou, I. (1952). *Recuento de vocabulario español*. San Juan: Universidad de Puerto Rico.
- Rodríguez Muñoz, F. J., & Muñoz Hernández, I. O. (2009). De la disponibilidad a la didáctica léxica. *Tejuelo: Didáctica de la Lengua y la Literatura. Educación*, (4).
- Rodríguez Muñoz, F. J., & Muñoz Hernández, I. O. (2011). Disponibilidad léxica sobre palabras específicas en estudiantes de educación secundaria de Almería. *Espiral. Cuadernos del profesorado*, 4(8), 22-31.
- Rojas, D., Zambrano, C., & Salcedo Lagos, P. (2017). Metodología de Análisis de Disponibilidad Léxica en Alumnos de Pedagogía a través de la

Comparación Jerárquica de Lexicones. *Formación Universitaria*, 10(4), 3-14.

Rubio Lastra, M. (2018). Disponibilidad léxica de 52 estudiantes taiwaneses universitarios de ELE A1. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (27).

Rubio Sánchez, R. (2015). Estudio de disponibilidad léxica en aprendices italianos de español: análisis cuantitativo. En *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General* (pp. 430-441). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.

Ruiz Basto, A. (1987). *Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades Plantel Naucalpan*. Tesis de licenciatura inédita. México D.F.: UNAM.

Ruiz Muñoz, D. (2004). *Manual de estadística*. Sevilla: Eumed.net.

Salcedo Lagos, P., Ferreira Cabrera, A., & del Valle Leo, M. (2014). *La disponibilidad léxica matemática en estudiantes de enseñanza media de la ciudad de Concepción de Chile*. Concepción: Universidad de Concepción.

Samper Hernández, M. (2001). Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de edición de los materiales. En *Nuevas aportaciones al estudio de la lengua española: investigaciones filológicas* (pp. 277-285). Salamanca: Luso-Española de Ediciones.

Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.

Samper Hernández, M. (2003a). Análisis comparativo entre dos léxicos disponibles en ELE. En *Léxico, fraseología y falsos amigos* (pp. 93-114).

- Samper Hernández, M. (2003b). *Evolución de la disponibilidad léxica en estudiantes grancanarios de enseñanza primaria y secundaria*. Salamanca: Universidad de Salamanca.
- Samper Hernández, M. (2005a). Dialectalismos en el léxico disponible de escolares grancanarios. En *Palabras, norma, discurso: en memoria de Fernando Lázaro Carreter* (pp. 1065-1078). Salamanca : Universidad de Salamanca, 2005.
- Samper Hernández, M. (2005b). La disponibilidad léxica de escolares grancanarios: resultados generales de la investigación. En *Con quien tanto quería : estudios en homenaje a María del Prado Escobar Bonilla* (pp. 577-588).
- Samper Hernández, M. (2014). The relationship of language proficiency to the lexical availability of learners of spanish. En R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (Vol. 17, pp. 103-124). Springer. https://doi.org/10.1007/978-94-007-7158-1_7
- Samper Padilla, J. A. (1998). Criterios de edición del léxico disponible: sugerencias. *Lingüística*, (10), 311-333.
- Samper Padilla, J. A. (1999). Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria. En *Estudios de lingüística hispánica: homenaje a María Vaquero* (pp. 550-573).
- Samper Padilla, J. A. (2021). Disponibilidad léxica y sociolingüística. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 171-204). Valencia: Tirant lo Blanch.
- Samper Padilla, J. A., & Hernández Cabrera, C. E. (1997). El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística. En *Contribuciones al estudio de la lingüística hispánica : homenaje al profesor Ramón Trujillo* (Vol. 2, pp. 229-240). [Barcelona] : Montesinos, 1997.

- Samper Padilla, J. A., & Hernández Cabrera, C. E. (2003). Los dialectalismos en el léxico disponible de Gran Canaria. Análisis de un centro de interés. En *Lengua, variación y contexto: estudios dedicados a Humberto López Morales* (Vol. 1, pp. 339-353).
- Samper Padilla, J. A., Hernández Cabrera, C. E., & Bellón Fernández, J. J. (2003). Léxico disponible y variación dialectal: datos de Canarias y Córdoba. En *Actas del I Congreso Internacional sobre el español de Canarias* (pp. 1077-1099). Islas Canarias: Academia Canaria de la Lengua.
- Samper Padilla, J. A., Hernández Cabrera, C. E., & Samper Hernández, M. (2019). Disponibilidad léxica en niños grancanarios de 6 años. Comparación con México y Costa Rica. *Ogigia: Revista electrónica de estudios hispánicos*, (25), 139-163.
- Samper Padilla, J. A., & Samper Hernández, M. (2006). Aportaciones recientes de los estudios de disponibilidad léxica. *Lynx: Panorámica de estudios lingüísticos*, (5), 5-95.
- Sánchez Gómez, C. (2005). Naturaleza gramatical del léxico disponible en informantes del español como lengua extranjera. *Interlingüística*, (16), 977-986.
- Sánchez-Saus Laserna, M. (2009). La variable «nivel de español» en el léxico disponible de los estudiantes de español como lengua extranjera. *Pragmalingüística*, (17), 140-153.
- Sánchez-Saus Laserna, M. (2010). Estudio de la influencia de tres variables en el léxico disponible de los estudiantes de español como lengua extranjera. *IX Congreso De Lingüística General*.
- Sánchez-Saus Laserna, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas*. Sevilla: Editorial Universidad de Sevilla.

- Sánchez-Saus Laserna, M. (2017). Los orígenes de la disponibilidad léxica: El Français Fondamental - antecedentes, elaboración y recepción. *Beitrag zur Geschichte der Sprachwissenschaft*, 27, 73-100.
- Sánchez-Saus Laserna, M. (2021). Reflexiones en torno a la disponibilidad léxica en ELE: cuestiones historiográficas, problemas metodológicos y retos de futuro. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 19-44). Valencia: Tirant lo Blanch.
- Sandu, B. M. (2012). La disponibilidad léxica en alumnos rumanos de ELE: incidencia de la variable «sexo/género» y su correlación con el «nivel escolar». *Lingua Americana*, (31), 61-85.
- Sandu, B. M. (2013). Romanian students' lexical availability in Spanish as a foreign language. *Estudios interlingüísticos*, (1), 121-133.
- Sandu, B. M. (2014). *Léxico disponible de alumnos rumanos que aprenden español como lengua extranjera en centros escolares bilingües rumano-españoles*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- Santiago Guervós, J. de. (2008). *Léxico disponible de Segovia: estudio y diccionarios*. Salamanca: Instituto de la Lengua Castellano y Leonés.
- Santos Díaz, I. C. (2015). La influencia del conocimiento de varias lenguas extranjeras en el léxico disponible. En *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General* (pp. 477-484). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- Santos Díaz, I. C. (2016). La influencia del centro educativo en el léxico disponible en lengua extranjera. En *Educación intercultural: metodología de aprendizaje en contextos bilingües* (pp. 107-111). Granada : Atrio, 2016.

- Santos Díaz, I. C. (2017). Selección del léxico disponible: propuesta metodológica con fines didácticos. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (27), 122-139.
- Santos Palmou, X. (2016). La selección del vocabulario en ELE: estado de la cuestión y nuevas metodologías. En *El español como lengua extranjera en Portugal II: retos de la enseñanza de lenguas cercanas* (pp. 164-178).
- Saralegui Platero, C., & Tabernero Sala, C. (2008). Aportación al proyecto panhispánico de léxico disponible: Navarra. En *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL) [Recurso electrónico]* (pp. 745-762). Pamplona: Universidad de Navarra.
- Serfati, M. (2015a). Cálculo de disponibilidad léxica en ELE de universitarios marroquíes de Agadir: el programa LexiDisp. *Actas del I Congreso de español como lengua extranjera del Magreb del Instituto Cervantes de Rabat*, 111-117.
- Serfati, M. (2015b). La ropa, alimentos y bebidas en el léxico disponible de estudiantes marroquíes de nivel universitario. En *Actas del I Congreso Internacional de la AEPE (Asociación Europea de Profesores de Español)* (pp. 490-501). Burgos: Universidad Isabel I de Castilla.
- Serfati, M. (2016). La disponibilidad léxica de estudiantes marroquíes de nivel universitario: resultados generales. *Philologica canariensis*, (22), 105-116.
- Serfati, M. (2017a). Disponibilidad léxica y la didáctica de ELE en Marruecos (niveles de secundaria y enseñanza superior en la región Souss-Massa). *Actas del II Congreso de español como lengua extranjera del Magreb del Instituto Cervantes de Casablanca*, 83-89.
- Serfati, M. (2017b). Incidencia cuantitativa de la variable «curso académico» en el léxico disponible de estudiantes marroquíes de español como lengua extranjera (Nivel universitario). *Aljamía. Revista de la Consejería de Educación en Marruecos*, (28), 35-54.

- Serfati, M. (2017c). Incidencia cuantitativa del factor «lengua materna» en la disponibilidad léxica de estudiantes marroquíes de Español como Lengua Extranjera (Nivel universitario). *Estudios interlingüísticos*, (5), 121-145.
- Serfati, M. (2018). El léxico disponible en estudiantes marroquíes de ELE (nivel universitario): incidencia de la variable 'sexo' y su correlación con el 'curso académico'. *Beoiberística*, 2(1), 153-180.
- Serrano Zapata, M. (2003). Disponibilidad léxica en la provincia de Lleida: aspectos metodológicos. *Interlingüística*, (14), 919-928.
- Serrano Zapata, M. (2004). Aspectos sociolingüísticos del léxico disponible castellano de los preuniversitarios leridanos. *Pragmalinguística*, (12), 147-168.
- Serrano Zapata, M. (2006). Consecuencias del contacto de lenguas en Lérida: interferencias detectadas en las encuestas de disponibilidad léxica. En *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social* (pp. 811-829).
- Serrano Zapata, M. (2014). *Disponibilidad léxica en la provincia de Lleida: estudio comparado de dos lenguas en contacto*. Lleida: Universidad de Lleida.
- Sifrar Kalan, M. (2009). Disponibilidad léxica en español como lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Finlandia. *Verba hispanica: anuario del Departamento de la Lengua y Literatura Españolas de la Facultad de Filosofía y Letras de la Universidad de Ljubljana*, (17), 165-182.
- Sifrar Kalan, M. (2012). Análisis comparativo de la disponibilidad léxica en español como lengua extranjera (ELE) y lengua materna (ELM). *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (15).

- Sifrar Kalan, M. (2014). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. *Studia Romanica Posnaniensia*, 41(1), 63-85. <https://doi.org/10.14746/strop.2014.411.005>
- Stevenson, A. (2010). *Oxford Dictionary of English*. Oxford: Oxford University Press. <https://doi.org/10.1093/acref/9780199571123.001.0001>
- Stevenson, A., & Lindberg, C. A. (2010). *New Oxford American Dictionary*. Oxford: Oxford University Press. <https://doi.org/10.1093/acref/9780195392883.001.0001>
- Terrádez Gurrea, M. (2000). Frecuencias léxicas y análisis estadístico. En *¿Cómo se comenta un texto coloquial?* (pp. 111-124). Barcelona: Editorial Ariel, S. A.
- Terrádez Gurrea, M. (2001). *Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo*. Valencia: Universitat de Valencia.
- 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*. Salamanca: Universidad de Salamanca.
- Trigo Ibáñez, E. (2011). *Dialectología y cultura: el léxico disponible de los preuniversitarios sevillanos*. Valencia: Aduana Vieja.
- Trigo Ibáñez, E., & Santos Díaz, I. C. (2021). Influencia del tipo de centro educativo (público o privado) sobre el léxico disponible de los preuniversitarios sevillanos. En M. S. Zapata & M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 227-246). Valencia: Tirant lo Blanch.
- Valencia Espinoza, A. (1997). Disponibilidad léxica. Muestreo y estadísticos. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, (2), 197-226.

- Valencia Espinoza, A., & Echeverría, M. S. (1999a). *Disponibilidad léxica en estudiantes chilenos*. Santiago de Chile: Universidad de Chile.
- Valencia Espinoza, A., & Echeverría, M. S. (1999b). El factor geográfico en la disponibilidad léxica chilena. En *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina Las Palmas de Gran Canaria* (Vol. 2, pp. 1705-1712). Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones.
- Vander Beke, G. E. (1932). *French word book*. New York: Macmillan.
- Vargas Sandoval, P. E. (1988). La disponibilidad léxica: un diagnóstico al hablante medio de la V Región en seis centros de interés. *Nueva Revista del Pacífico*, (33), 115-123.
- Velarde Inzunza, J. de J. (2014). *Análisis de léxico disponible en hablantes escolares de Culiacán, Sinaloa*. Culiacán: Universidad Autónoma de Sinaloa.
- Vélez Ibarrola, R., Ramos Méndez, E., Hernández Morales, V., Carmena Yañez, E., & Navarro Fernández, J. (2004). *Métodos estadísticos en ciencias sociales*. Madrid: Ediciones Académicas.
- Verdeses Mirabal, R. T. (2012). Disponibilidad léxica de los estudiantes hispanos de Redwood City, California. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, 6(11).
- Victory, J. B. (1971). *A study of lexical availability among monolingual-bilingual speakers of Spanish and English*. Houston: Rice University.
- Vox. (2009). *Diccionario General de la Lengua Española*. Barcelona: Vox Bibliograf.