



Università
Ca' Foscari
Venezia

Corso di Laurea magistrale
in Scienze del Linguaggio

Tesi di Laurea

**Disponibilidad léxica de los estudiantes italianos de ELE en
contexto inmersivo: análisis de tres centros de interés**

Relatore

Prof. Florencio del Barrio de la Rosa

Correlatrice interna

Prof.ssa Elena Dal Maso

Correlatrice esterna

Prof.ssa Marisela del Carmen Pérez Rodríguez

Laureanda

Irene Pantarotto

Matricola 878750

Anno Accademico

2023 / 2024

AGRADECIMIENTOS

Antes de empezar quiero expresar mi más sincero agradecimiento a todas las personas sin las cuales este proyecto nunca se habría podido realizar.

A mi director, el profesor del Barrio, por proponerme, hace más de un año, investigar en esta disciplina, por haberme impulsado y guiado durante todo el proceso y por haber dado respuestas a todas mis preguntas.

A la profesora Marisela, mi codirectora de la Universidad de Oviedo, por la disponibilidad demostrada en seguirme en este trabajo, por su orientación y por la confianza que ha depositado en mí.

A la profesora Dal Maso, mi codirectora de la Universidad Ca' Foscari, por la disponibilidad en acogerme en las etapas finales de este camino.

A mis padres y a mi hermano, a vosotros os debo todo. Gracias por estar siempre orgullosos de mí y de mis resultados. A mis tíos y a toda mi familia, gracias por animarme y quererme tanto.

A Giacomo, por estar a mi lado desde siempre. Gracias por apoyarme cada día y empujarme a ser atrevida y valerosa en mis decisiones, aunque implicaran poner kilómetros y kilómetros de distancia entre nosotros.

A Ana por su apoyo incondicional y escucha constante en estos años. Gracias por haberme soportado, en todos los sentidos.

A Pier por el soporte en mis últimos años venecianos y en este periodo español. Por haberme animado siempre y haber creído en mí más que yo misma.

A La Sara, porque desde aquel mensaje sobre el examen de inglés, nació una verdadera amistad.

A Marta e Irene, por las aventuras en nuestra casita. Gracias por quererme y haberme hecho sonreír y ver la luz, aun cuando todo me parecía oscuro.

A Riccardo y Annabella, mis compañeros de Erasmus. Entrasteis en mi vida en septiembre y gracias a vosotros Oviedo ha sido casa. Vuestra Princesa os lo agradece infinitamente.

A todos mis amigos italianos y españoles que me han acompañado y sostenido en este camino.

Last but not least, gracias a los catorce estudiantes italianos que participaron en las encuestas. Sin vuestra ayuda y colaboración todo esto no habría sido posible.

ÍNDICE

INTRODUCCIÓN	7
ESTADO DE LA CUESTIÓN	10
1.1. Orígenes de la disponibilidad léxica	10
1.2. Los herederos del <i>Français Fondamental</i>	17
1.2.1. <i>Estudios en el mundo hispánico</i>	19
1.2.2. <i>Estudios en ELE</i>	27
DESCRIPCIÓN DE LA ENCUESTA	32
2.1. La muestra.....	33
2.1.1. <i>Los informantes</i>	33
2.1.2. <i>Las variables seleccionadas</i>	34
2.2. El cuestionario sociológico	37
2.3. La prueba de disponibilidad léxica.....	42
2.3.1. <i>Los centros de interés</i>	42
2.3.2. <i>La edición de los datos</i>	47
ANÁLISIS CUANTITATIVO DE LOS RESULTADOS	53
3.1. Resultados generales	53
3.1.1. <i>Número de palabras</i>	53
3.1.2. <i>Número de vocablos</i>	55
3.1.3. <i>Densidad léxica e índice de cohesión</i>	56
3.2. Resultados por variable	57
3.2.1. <i>Variable Sexo</i>	57
3.2.2. <i>Variable Carrera</i>	61
3.2.3. <i>Variable Nivel de español</i>	65
3.2.4. <i>Variable Conocimiento de otras lenguas</i>	69
3.3. Conclusiones parciales	72
ANÁLISIS CUALITATIVO DE LOS RESULTADOS	74
4.1. Resultados generales	74
4.2. Resultados por variable	79
4.2.1. <i>Variable Sexo</i>	79
4.2.2. <i>Variable Carrera</i>	83
4.2.3. <i>Variable Nivel de español</i>	87
4.2.4. <i>Variable Conocimiento de otras lenguas</i>	91
4.3. Conclusiones parciales	95

ANÁLISIS DE LAS RELACIONES LÉXICAS	96
5.1. Las relaciones léxico-semánticas	96
5.2. Análisis general	99
5.3. Análisis según la variable Nivel de español	101
5.3.1. <i>CI 01. Partes del cuerpo</i>	101
5.3.2. <i>CI 05. Alimentos y bebidas</i>	106
5.3.3. <i>CI 10. La ciudad</i>	111
5.4. Conclusiones parciales	114
CONCLUSIONES	116
REFERENCIAS BIBLIOGRÁFICAS	119
ANEXO I – MODELO PRUEBA DE DISPONIBILIDAD Y CUESTIONARIO SOCIOLÓGICO	126
ANEXO II – DICCIONARIO DE DISPONIBILIDAD	131

INTRODUCCIÓN

La disponibilidad léxica nació en la mitad del siglo pasado con un estudio conducido por unos lingüistas franceses que idearon una nueva metodología para seleccionar el léxico básico de una lengua uniendo el léxico frecuente y el léxico disponible, es decir, el conjunto de palabras que por primeras vienen a la mente de los hablantes a la hora de hablar de un determinado tema. No obstante, es cierto que su mayor desarrollo en el mundo hispánico se debe a las investigaciones llevadas a cabo por Humberto López Morales en Puerto Rico. Estos estudios pioneros suscitaron tanto interés que a finales de los años noventa nació el *Proyecto Panhispánico de Disponibilidad Léxica* con el objetivo de publicar el léxico disponible común a todo el mundo hispánico y de cada uno de los países que forman esta comunidad.

En estas décadas de historia, las investigaciones se han concentrado mayoritariamente en los hablantes nativos de español, aunque en los últimos años ha aumentado el interés hacia las aplicaciones de la disponibilidad léxica a estudiantes de español como lengua extranjera (ELE). Sin embargo, las investigaciones llevadas a cabo con una muestra compuesta enteramente por aprendices italianos son todavía limitadas. Por esta razón elegimos acercarnos a este tema y estudiar un grupo de estudiantes universitarios italianos en contexto inmersivo.

Este trabajo ha empleado una muestra de estudiantes procedentes de varias universidades italianas que se encontraban en Oviedo para participar al programa Erasmus. Las encuestas de disponibilidad léxica se llevaron a cabo entre los meses de diciembre y enero/febrero, con informantes que llevaban viviendo en España al menos cuatro meses. Hemos elegido centrarnos en tres de los dieciséis centros de interés tradicionales: Partes del cuerpo, Alimentos y bebidas y La ciudad. Los principales objetivos de esta investigación son presentar un estudio introductorio del léxico disponible de estudiantes italianos en contexto inmersivo e individuar las relaciones léxico-semánticas en el corpus de palabras obtenidas.

El presente trabajo se estructura en cinco capítulos. En el primero exponemos el origen y el desarrollo de la disponibilidad léxica, empezamos introduciendo los diccionarios de frecuencia y siguiendo con los estudios llevados a cabo en Francia bajo la dirección de Georges Gougenheim. Nos centramos sucesivamente en los estudios herederos de esta obra tanto en el mundo hispánico, llevados a cabo con hablantes nativos de español, como en las investigaciones más recientes con estudiantes de ELE.

El segundo capítulo se focaliza en la descripción de la encuesta. Como es habitual en las investigaciones de este tipo, presentamos a nuestros informantes dos encuestas: la primera era una prueba de disponibilidad propiamente dicha con el objetivo de obtener las palabras más disponibles en los tres centros de interés objeto de este estudio y la segunda, el cuestionario sociológico, con el objetivo de obtener las variables que usamos para el análisis de los datos. En esta sección central del capítulo exponemos las preguntas del cuestionario y los resultados que pudimos sacar de ellas por lo que se refiere a los datos personales de nuestros informantes, como sexo, edad, carrera, etc. a su conocimiento y uso del español, y a sus actitudes hacia esta lengua. Sucesivamente exponemos brevemente las cuestiones todavía abiertas sobre la adecuación de los dieciséis centros de interés tradicionales y las propuestas que se han avanzado para mejorarlos. Por último, describimos los criterios de edición que empleamos para “limpiar” los datos tanto de manera general como particular por cada centro de interés.

En el tercer capítulo analizamos los datos obtenidos y editados desde un punto de vista cuantitativo. Inicialmente adoptamos una perspectiva general, centrándonos en el número de palabras y de vocablos totales y en los promedios correspondientes, así como en la densidad léxica y en el índice de cohesión. Sucesivamente analizamos estos criterios desglosando los resultados por cada variable que elegimos investigar: sexo, carrera, nivel de español y conocimiento de otras lenguas.

En el cuarto capítulo presentamos el análisis cualitativo de los datos. Empezando por los resultados generales, examinamos las palabras que alcanzan un índice de disponibilidad superior o igual a 0,1 en cada centro de interés. Sucesivamente analizamos los datos de cada centro de interés más detalladamente, desglosándolos según las cuatro variables.

Finalmente, en el quinto capítulo nos centramos en las relaciones léxico-semánticas. Después de una breve descripción del tipo de relaciones que se pueden establecer entre las palabras, analizamos los pares léxicos, es decir los grupos de dos palabras consecutivas que pudimos individuar repetidos en los tres centros de interés. En la segunda parte del capítulo presentamos un análisis de todos los pares léxicos que pudimos encontrar en los tres centros dividiendo los resultados por nivel de español con el objetivo de individuar la influencia del conocimiento lingüístico en la creación de pares, y por consecuencia en la organización del léxico mental.

Por último, adjuntamos en los anexos los materiales empleados y las tablas con los resultados de nuestra investigación. En el anexo I añadimos el modelo de encuesta con el cual

hemos presentado las pruebas de disponibilidad léxica y el cuestionario sociológico a los informantes. En el anexo II presentamos los listados de palabras, o diccionarios de disponibilidad de los tres centros de interés investigados con las indicaciones del índice de disponibilidad, la frecuencia relativa, el porcentaje de aparición y la frecuencia acumulada de cada palabra.

Esperamos que esta investigación, aunque limitada en el número de informantes, pueda representar un primer acercamiento a los estudios de disponibilidad léxica de estudiantes italianos de ELE en contexto inmersivo y abrir el paso a nuevos estudios en este ámbito.

CAPÍTULO I

ESTADO DE LA CUESTIÓN

1.1. Orígenes de la disponibilidad léxica

A lo largo del tiempo, una preocupación constante en la didáctica de las lenguas maternas y extranjeras ha sido la necesidad de identificar las palabras que primero debía aprender un estudiante. La léxico-estadística es la disciplina que se ha ocupado de encontrar los procedimientos que permitieran reflejar el uso real que hacen los hablantes del vocabulario de una lengua. Para realizar estas listas de vocabulario, se usaba el índice de frecuencia, es decir, el número de apariciones de las palabras en los varios textos examinados.

La mayoría de los estudiosos identifica como obra pionera en este campo, aunque su objetivo no era la aplicación a la enseñanza, el diccionario de frecuencia elaborado por J. W. Käding en 1897 que recogía las palabras, los sufijos y las letras más frecuentes del alemán, a partir de un corpus formado por cartas, textos administrativos, políticos, militares, etc. Para el inglés, la primera obra que encontramos es el *Teacher's Word Book* de E. L. Thorndike, publicado en 1921, que presenta una lista de 10 000 lexías sacadas de 4 500 000 palabras de textos diversos. Por el francés hay dos obras iniciales publicadas por Vivian Allen Charles Henmon en 1924 y el *French Word Book* publicado por George E. Vander Bake en 1935 (cf. Gougenheim *et al.*, 1964: 31-32).

Por lo que se refiere al español los primeros diccionarios de frecuencia son: *A Graded Spanish Word Book*, publicado por Milton A. Buchanan en 1927; el *Vocabulario usual, común y fundamental*, publicado por M. Victor García Hoz en 1952, y en el mismo año el *Recuento de Vocabulario español*, dirigido por Ismael Rodríguez Bou. El primero tenía como objetivo crear una lista de palabras frecuentes para la enseñanza del español a anglohablantes a partir de un corpus 40 textos diferentes tanto literarios como técnicos y periodísticos. La obra de García Hoz recogía 400 000 palabras sacadas de cartas familiares, periódicos, documentos oficiales y libros. El *Recuento* se basaba, además de materiales escritos procedentes de periódicos, libros de texto, etc., de textos orales obtenidos de la radio con la finalidad de mejorar la enseñanza del español como lengua materna en Puerto Rico (cf. Carcedo, 1998: 9-10).

Pese a los esfuerzos de los lingüistas para encontrar textos que fueran lo más representativo posibles del uso real de la lengua, en la mitad de los años 50 del siglo XX la comunidad científica empezó a darse cuenta de que el criterio de frecuencia, en el que se basaban estas listas de palabras, por sí solo no era suficiente. Las mayores críticas que se hicieron a los diccionarios de frecuencia fueron que se basaban en un corpus formado, en la mayoría de los casos, únicamente por textos escritos que utilizaban un lenguaje elaborado y alejado del uso cotidiano de la lengua, que su resultado era una simple lista de palabras descontextualizadas, y además, que la gran mayoría de ellas eran unidades funcionales de significado gramatical, como artículos, preposiciones, conjunciones, mientras que las palabras de contenido nocional, comúnmente usadas por los hablantes, como los sustantivos, no estaban incluidas en las listas porque tenían un índice de frecuencia bajo o muy bajo (cf. Nalesso, 2019: 88).

Los primeros que intentaron solucionar este problema fueron los lingüistas franceses Georges Gougenheim, Paul Rivenc, René Michéa y André Sauvageot, que en 1954 publicaron el *Français Elementaire* y diez años más tarde una edición corregida y aumentada conocida como el *Français Fondamental*. El Ministère de l'Éducation Nationale encargó a este grupo la creación de un francés elemental, básico y simplificado para la enseñanza del francés a los extranjeros. En la obra se especifica que esta reducción del vocabulario y de la gramática es destinada a (Gougenheim *et al.*, 1964: 21-22):

1. Enseignement scolaire: a) pratique; b) culture!.
2. Diffusion, dans de jeunes états indépendants, d'une langue donnée choisie comme langue nationale ou langue étrangère privilégiée.
3. Besoins du voyageur, du touriste, du stagiaire étranger, de l'étranger venu travailler dans un autre pays.¹

Para cumplir estos propósitos tuvieron que evaluar y elegir cuáles son los elementos más básicos que es necesario enseñar –y aprender– en los primeros niveles. Así que tuvieron que enfrentarse con el problema de la limitación: los hablantes nativos no utilizan todo el vocabulario de su lengua, y no se espera que un hablante no nativo lo haga; por eso tuvieron que limitar y seleccionar las unidades esenciales para la enseñanza (cf. Gougenheim *et al.*,

¹ Nuestra traducción:

1. La enseñanza escolar: a) práctica; b) cultura.
2. La difusión, en los nuevos estados independientes, de la lengua que han elegido como lengua nacional.
3. Los viajeros, turistas, pasantes extranjeros y los inmigrantes que se mudan a Francia para trabajar.

1964: 11). Es más, la metodología tradicional de enseñanza-aprendizaje de las lenguas implicaba varios años de estudio y sobre todo contacto con textos escritos. Sin embargo, aunque esta metodología podía funcionar con las lenguas clásicas, el latín y el griego, en el siglo XX ya no era compatible con las necesidades de los aprendices de las lenguas modernas que las necesitaban para fines concretos: para viajar, hacer negocio o estudiar al extranjero. La selección del vocabulario tenía que permitir aprender rápidamente los idiomas para hacer un uso práctico y comunicativamente eficaz de ellos; esto fue otro elemento para descartar el criterio de frecuencia por sí solo (cf. Gougenheim *et al.*, 1964: 11-12).

Al analizar el vocabulario de los diccionarios de frecuencias, Michéa distinguió en 1953 entre las palabras atemáticas y temáticas. Las primeras son las palabras más frecuentes, que aparecen con un índice de frecuencia muy alto independientemente del tema tratado en los textos; en cambio, las palabras temáticas son menos frecuentes porque están vinculadas a temas o situaciones específicas, y por eso su frecuencia depende de los textos que se han elegido para formar el corpus. Evidencia también que para el aprendizaje de las palabras atemáticas y funcionales sería suficiente su repetición; sin embargo, las palabras reales, temáticas y conectadas con unas situaciones concretas, resultan más fáciles de aprender mediante asociación de ideas (cf. Gougenheim *et al.*, 1964: 144).

Esta primera distinción se relaciona con lo que el mismo autor llamará vocabulario general (frecuente) y vocabulario de interés (disponible). Los dos son complementarios, puesto que el primero está formado por las palabras necesarias para expresarse, estadísticamente frecuentes de manera estable; en cambio el vocabulario de interés o disponible está compuesto por nombres concretos que se ajustan a las necesidades del hablante según la situación en la que se encuentra y tienen una frecuencia irregular (cf. Michéa, 1953: 341). Desde esta distinción, Michéa definió las unidades léxicas disponibles como:

“En présence d’une situation donnée, les mots que viennent les premiers à l’esprit sont ceux qui sont liés tout spécialement à cette situation et la caractérisent [...]. Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé, et se présente immédiatement à l’esprit ou moment où l’on en a besoin.” (Michéa, 1953: 340)²

² Nuestra traducción: “En una situación determinada, las palabras que salen a la mente por primeras son aquellas que están más conectadas con tal situación y que la caracterizan [...]. Una palabra disponible es una palabra que, a pesar de no ser muy frecuente, es siempre disponible para ser utilizada y que ocurre inmediatamente cada vez que sea necesario.”

Basándose en estas definiciones, el grupo de lingüistas franceses empezó a trabajar seleccionando no solamente las palabras más frecuentes, sino también las más disponibles. Como escriben los mismos autores del *Français Fondamental* (Gougenheim *et al.*, 1964: 13), esta obra “a été établi selon une méthode qui innove, croyons-nous, dans la technique des langues de base. Elle combine en effet:

1. La fréquence des mots dans la langue parlée;
2. La recherche des mots disponibles les plus utiles;
3. L’empirisme rationnel.”³

La innovación en la metodología de esta obra es representada por el empleo de la lengua hablada y de la noción de disponibilidad. De hecho, como ya se ha mencionado, la didáctica de las lenguas extranjeras hasta el momento se basaba en los textos escritos, en cambio los lingüistas franceses decidieron sacar los datos sobre la frecuencia a partir de conversaciones orales. Además, emplearon el concepto de disponibilidad recurriendo a ejercicios asociativos.

En el primer capítulo de la segunda parte de la obra explican muy detenidamente la metodología usada para recoger los datos: dado que una de las finalidades del proyecto era determinar una lengua para un fin práctico y comunicativo, decidieron partir de conversaciones reales que grabaron y transcribieron. Antes de empezar recogieron las informaciones sobre el origen de los informantes, su nivel educativo y su profesión, luego empezaron las conversaciones evitando temas delicados, como la política, la religión o asuntos personales, para evitar que se crearan reticencias en los interlocutores. Los temas abarcados fueron muy variados: profesiones, familia y amigos, viajes y turismo, salud y enfermedad, medios de transporte, escenas de vida cotidiana, literatura, arte y espectáculo, deportes y juegos, alojamiento, ocupaciones femeninas, vida estudiantil, comida, etc. No solamente hablaron con personas desconocidas, sino también les ocurrió registrar las conversaciones durante unas reuniones familiares o con amigos. En un segundo momento se transcribieron estas grabaciones por un total de 312 135 palabras y 7 995 palabras diferentes (cf. Gougenheim *et al.*, 1964: 61-66).

Estas palabras se organizaron en dos listas de frecuencia: la primera compuesta por palabras que tenían una frecuencia igual o superior a 20, es decir, con un mínimo de 20

³ Nuestra traducción: la obra “se basa en una metodología innovadora en la técnica de las lenguas básicas. La metodología combina: 1. La frecuencia de las palabras de la lengua hablada, 2. La búsqueda de las palabras disponibles más útiles, 3. El empirismo racionalista.”

apariciones en el total de 163 textos, en orden decreciente, y la segunda formada por las mismas palabras, ordenadas alfabéticamente con la indicación de la frecuencia (cf. Gougenheim *et al.*, 1964: 68). A partir de estas dos listas empezó el análisis, con el que relevaron que las palabras con mayor frecuencia son palabras gramaticales y que su porcentaje va reduciéndose conforme a la reducción de la frecuencia, decreciendo regularmente de 90,9% a 5,7%. En cambio, los sustantivos tienen la variación opuesta: entre las frecuencias 1 000 y 20 pasan de 8,7% a 53,9%. La variación de los verbos es más irregular: crecen rápidamente de 9,09% a 23% a la frecuencia 500, luego descienden ligeramente a la frecuencia 300 y se mantienen estables alrededor del 20%. Los adjetivos tienen un comportamiento similar al de los sustantivos: aparecen a partir de la frecuencia 1 000 y hasta la frecuencia 60 crecen rápidamente hasta un 17,6%, luego decrecen hasta llegar a 5,7% y sucesivamente se establecen entre el 13,3 y el 16,4% (cf. Gougenheim *et al.*, 1964: 114-116).

A partir de estos resultados, los lingüistas se preguntaron qué importancia tenían que dar a la frecuencia, e identificaron tres posibilidades (Gougenheim *et al.*, 1964: 137):

1. Tenir compte uniquement de la fréquence.
2. N'en tenir aucun compte.
3. Ne demander à la fréquence que ce qu'elle peut donner et la compléter par d'autres moyens.⁴

Era evidente que para diseñar un vocabulario básico no podían usar solo las palabras más frecuentes con significado gramatical, sino que tenían que buscar un método para aplicar la frecuencia de manera distinta para tener acceso también a las palabras de menor frecuencia, pero igualmente básicas y esenciales en la cotidianidad. Uno de los ejemplos que ponen es la palabra *tenedor*:

“prenons le mot fourchette. Voilà bien, dita-t-on, un mot qui doit être fréquent: nous manions cet instrument deux fois par jour, un enfant de trois ans, doué d'une intelligence normale, sait ce que désigne ce mot. Mais quand le prononçons-nous? Quand nous disons à un enfant: “Ne laisse pas tomber ta fourchette” et dans telle autre circonstance. Mais nous pouvons rester des jours et des semaines sans le prononcer.” (Gougenheim *et al.*, 1964: 138-139)⁵

⁴ Nuestra traducción: 1. Considerar únicamente la frecuencia, 2. No considerarla, 3. Considerar la frecuencia solo por lo que puede ofrecer e integrarla con otros métodos.

⁵ Nuestra traducción: “Tomamos la palabra tenedor. Se puede pensar que tiene que ser una palabra frecuente: usamos este instrumento dos veces al día, un niño de tres años dotado de una inteligencia normal sabe a qué se refiere. Pero ¿cuándo la pronunciamos? Cuando decimos a un niño “No hagas caer el tenedor” y en otras situaciones parecidas. Pero podemos llevar días o semanas sin pronunciarla.”

Añaden ejemplos parecidos con las palabras *codo*, *dientes*, *metro* y *rey* y ponen en evidencia que la frecuencia de las palabras concretas es muy inestable. Identifican tres causas de esta inestabilidad: la naturaleza del texto o de la conversación, el país y las circunstancias (cf. Gougenheim *et al.*, 1964: 139-140).

La solución que encuentran para corregir esta inestabilidad es la de complementar la frecuencia con la disponibilidad. A partir de los experimentos de Michéa que le permitieron definir el concepto de disponibilidad, diseñaron una metodología asociativa a partir de unos estímulos para que los participantes produjeran palabras concretas. Se eligieron dieciséis estímulos, áreas temáticas o centros de interés:

1. les parties du corps;
2. les vêtements (peu importe que ce soient des vêtements d'homme ou de femme);
3. la maison (mais pas les meubles);
4. les meubles de la maison;
5. les aliments et boissons des repas (à tous les repas de la journée);
6. les objets placés sur la table et dont on se sert à tous les repas de la journée;
7. la cuisine, ses meubles et les ustensiles qui s'y trouvent;
8. l'école, ses meubles et son matériel scolaire;
9. le chauffage et l'éclairage;
10. la ville;
11. le village ou le bourg;
12. les moyens du transport;
13. les travaux des champs et du jardinage;
14. les animaux;
15. les jeux et distractions;
16. les métiers (les différents métiers et non pas les noms qui se rapportent à un seul métier). (Gougenheim *et al.*, 1964: 152-153)⁶

Después de haber elegido las áreas temáticas que querían investigar, les aplicaron las encuestas a escolares, puesto que eran un colectivo bastante numeroso y podían contar con el

⁶ Nuestra traducción: 1. Las partes del cuerpo, 2. La ropa (no importa si es ropa de hombre o de mujer), 3. La casa (sin los muebles), 4. Los muebles de la casa, 5. Los alimentos y bebidas (en todas las comidas del día), 6. Los objetos colocados sobre la mesa y utilizados en todas las comidas del día, 7. La cocina, sus muebles y los utensilios que hay en ella, 8. La escuela, sus muebles y los materiales escolares, 9. La calefacción y la iluminación, 10. La ciudad, 11. El campo o el pueblo, 12. Los medios de transporte, 13. Los trabajos del campo y jardinería, 14. Los animales, 15. Los juegos y distracciones, 16. Las profesiones (las diferentes profesiones y no los nombres que se relacionan con ella)

apoyo de los profesores (cf. Gougenheim *et al.*, 1964: 153). Además, se pensaba que los niños tenían un conocimiento bastante amplio del vocabulario común, todavía libre de posibles especializaciones. La metodología con la que se presentó la encuesta sigue manteniéndose hoy en día, aunque con algunas variaciones. Los datos se recogieron por escrito, en unos cuadernillos que tenían en la prima página preguntas para obtener informaciones personales como el nombre, apellido, región, sexo, fecha de nacimiento, nivel escolar, fecha del ejercicio, profesión de los padres y rendimiento del alumno⁷. Después se presentaba la tarea pidiendo a los escolares que, sin límite de tiempo, pensasen a los 20 nombres que les parecían más útiles sobre uno de los dieciséis centros de interés que se estaba examinando en aquel momento (cf. Gougenheim *et al.*, 1964: 154). A partir de los datos, llevaron a cabo un análisis de tipo cualitativo, puesto que, al usar listas cerradas de veinte palabras, no se podía hacer también un análisis cuantitativo.

Después de haber redactado las listas de frecuencia y analizado los materiales obtenidos mediante las pruebas de disponibilidad, los lingüistas franceses procedieron con la creación de la lista de palabras del francés fundamental. Hicieron varias exclusiones e inclusiones en el conjunto de palabras hasta llegar a obtener 1 475 palabras diferentes. De estas, el 46,9% son sustantivos, muchos de los cuales fueron introducidos por los estudiosos porque eran palabras disponibles o porque considerados necesarios para los objetivos de la obra, el 22,9% son verbos, el 17,1% palabras gramaticales, el 6,6% adjetivos y el restante 5,6% adverbios de modo (Gougenheim *et al.*, 1964: 197-207).

El *Français Fondamental* fue el primer trabajo de este tipo que se llevó a cabo. Sucesivamente muchos otros lingüistas aplicaron el método de la disponibilidad léxica para investigar en sus lenguas. Al ser el primer trabajo también recibió varias críticas sobre todo por lo que se refiere a la subjetividad de los centros de interés y de la selección de palabras que hicieron antes de crear las listas. Otro aspecto criticado es la presentación de listas en orden alfabético que no ayuda a identificar las palabras más disponibles⁸. Por último, también se le ha criticado que, al ofrecer listas de palabras aisladas del contexto, es imposible solucionar los problemas de homonimia y polisemia.

⁷ Hoy en día las investigaciones sobre disponibilidad léxica siguen preguntando informaciones personales para utilizarlas como variables, sin embargo, la mayoría se realiza de manera anónima, asignando un código a cada informante, no se suele preguntar por la fecha de nacimiento y tampoco una evaluación del rendimiento de los participantes.

⁸ Para solucionar este problema la mayoría de los trabajos sucesivos, como los que forman parte *del Proyecto Panhispánico* presentan los datos ordenados por tema, y no alfabéticamente.

A pesar de algunas críticas, está claro que este primer trabajo sobre la disponibilidad léxica aportó un nuevo campo del saber a la lingüística, con el intento de crear el vocabulario básico de una lengua que no se basase solamente en la frecuencia; pero también ha sido, y sigue siendo, importante para la psicología, puesto que da acceso a las asociaciones mentales entre las palabras. Además, como escribe López Morales (1999: 10), con la aparición de la disponibilidad “se empezaron a deslindar conceptos que hasta la fecha habían sido manejados como sinónimos: palabras frecuentes, palabras comunes y palabras usuales, porque había quedado claro que algunos términos comunes, y hasta usuales, no eran frecuentes”.

Antes de seguir presentando los principales estudios herederos del *Français Fondamental*, nos parece necesario dar unas definiciones de estos conceptos que se empezaron a delimitar a partir de la publicación de la obra francesa. El conjunto de las palabras frecuentes forma el léxico frecuente, es decir, aquellas palabras que tienen el mayor número de ocurrencias en el corpus lingüístico examinado. El léxico básico o usual está formado por las palabras que se utilizan habitualmente en la cotidianidad, está estrechamente vinculado con el uso real de la lengua independientemente del tema tratado. La mayoría de las palabras son gramaticales como verbos, adjetivos y sustantivos de significado general. El léxico disponible, como ya se ha descrito, es el conjunto de las palabras que los hablantes tienen en el lexicón mental, dichas palabras son aquellas que se activan por primeras en una situación comunicativa concreta, y en la mayoría de los casos son unidades de contenido nocional. Es un léxico potencial y no actualizado en el que abundan los sustantivos. El léxico fundamental es el conjunto del léxico básico y del disponible, es decir, la unión entre el vocabulario real de uso, más estable, con el vocabulario concreto vinculado al contexto comunicativo (cf. Nalesso, 2019: 85-87).

1.2. Los herederos del *Français Fondamental*

A partir de la publicación del *Français Fondamental*, muchos estudiosos han empezado a interesarse a este nuevo método de extracción del léxico para la elaboración de vocabularios básicos. La disponibilidad léxica tiene su mayor difusión en el mundo hispánico, sobre todo a partir de la creación del *Proyecto Panhispánico*; sin embargo, hay también numerosos trabajos de investigación sobre el francés y el inglés.

Una de las primeras investigaciones que siguieron al *Français Fondamental* llevada a cabo siempre en ámbito francés fue conducida por William Mackey en Canadá. A partir de los años 1958-1959 se hicieron las primeras encuestas sobre disponibilidad léxica en varias zonas del país, pero dado que había notables diferencias metodológicas, se decidió crear un comité

de investigación que coordinara las actividades y fijara las directrices del proyecto. Los resultados se publicaron en 1971 en *Le vocabulaire disponible du français*. En general se decidió mantener el método utilizado en Francia para poder comparar los resultados obtenidos en los dos países, así como la tipología de encuestados (niños de 9 a 12 años), por otra parte, se añadieron 6 centros de interés más⁹ y algunas preguntas en el cuestionario sociológico. Entre las principales conclusiones de este estudio, cabe destacar que se demostró que el entorno influye en la organización léxica de los hablantes, puesto que hubo bastante divergencia en los grados de disponibilidad de muchas palabras debido a que las realidades más cercanas a los encuestados fueron las más evocadas (cf. Carcedo, 1998: 19-21).

El primer trabajo sobre la disponibilidad léxica de lengua inglesa fue publicado por Naum Dimitrijévic en 1969 con el título *Lexical Availability. A New Aspect of the Lexical Availability of Secondary School Children*, donde expone los resultados de su investigación llevada a cabo con estudiantes de 14 años de varios institutos de Edimburgo y alrededores. Él se distancia de la metodología utilizada por los lingüistas franceses porque propone listas abiertas con el objetivo de identificar cuáles y cuántas eran las palabras más disponibles, sin restricciones; además introduce una limitación de tiempo de 5 minutos por centro de interés para captar las palabras que siempre están disponibles en los estudiantes¹⁰ (cf. Dimitrijévic, 1969: 14 *apud* Sánchez-Saus Laserna 2011: 77); y, finalmente, reduce los centros de interés a 11, eliminando algunos de los tradicionales e incluyendo otros como ciencia y política¹¹. Esta decisión se debe a que consideraba que los centros de interés no debían ser demasiados amplios y generales para generar solapamientos de palabras, pero tampoco demasiados estrechos y técnicos como para limitar el número de palabras (cf. Dimitrijévic, 1969: 16 *apud* Sánchez-Saus Laserna, 2011: 78). Otra novedad fue la de introducir como variable el cociente de inteligencia basándose en un examen que los niños hacían al empezar la secundaria, junto con una evaluación hecha por los profesores.

Entre los primeros estudios de tipo comparativo entre el léxico disponible de dos lenguas encontramos la investigación llevada a cabo en 1971 por Bailey Victory entre español

⁹ 17) “l’église et la vie paroissiale”, 18) “les sports et l’équipement”, 19) “les outils”, 20) “l’automobile et ses parties”, 21) “actions courantes” y 22) “les affaires”. Traducidos al español: 17) la Iglesia y la vida parroquial, 18) los deportes y el quipamiento deportivo, 19) herramientas, 20) el automóvil y sus partes, 21) actos corrientes, 22) negocios.

¹⁰ Para individuar el límite de tiempo realizó un test previo para observar durante cuánto tiempo los encuestados son capaces de seguir añadiendo palabras a la lista.

¹¹ Los centros de interés investigados por Dimitrijévic son: 1. Animals, 2. Countryside, 3. Town, 4. Entertainment, 5. Jobs and professions, 6. Science, 7. Means of transport, 8. Politics, 9. Parts of the house, 10. Food and drink, 11. Clothes.

e inglés, y la de Njock entre francés y basaa publicada en 1979. El primero compara los resultados producidos por alumnos monolingües en inglés y español, y bilingües, con la finalidad de comprobar las diferencias culturales. El vocabulario disponible era muy similar en inglés y en español desde el punto de vista cuantitativo, pero destaca que el grupo bilingüe es el que manifestó más interferencias (cf. Casanova, 2017: 50-51). Njock examina la población bilingüe de Camerún y utiliza solo diez centros de interés, manteniéndose bastante fiel a los del *Français Fondamental*, aunque cambia la formulación de algunos enunciados, y añadiendo “música y danza (géneros)”¹². Entre sus conclusiones destaca que los sujetos bilingües producían más palabras que los individuos monolingües (cf. Casanova, 2017: 23).

Esta breve introducción a los primeros trabajos que siguieron al *Français Fondamental*, demuestra que desde el principio se han llevado a cabo investigaciones en varias lenguas y con grupos de encuestados con características muy distintas. Sin embargo, como se analizará en la siguiente sección, los estudiosos del mundo hispánicos son los que han sabido sacar el mayor provecho de los estudios sobre disponibilidad léxica.

1.2.1. Estudios en el mundo hispánico

Humberto López Morales es el primero que en la década de los setenta empieza a estudiar la disponibilidad léxica en el mundo hispánico. En 1973 lleva a cabo el primer trabajo en la zona metropolitana de San Juan de Puerto Rico con el objetivo de hacer un análisis sociolingüístico de la disponibilidad léxica y determinar la evolución de la adquisición del vocabulario y las variables que la influyen. En su estudio utiliza 10 centros de interés con una muestra de 63 sujetos de dos diferentes niveles socioeconómicos. En los años siguientes publica varios análisis de ampliación de esta investigación, hasta llegar a la publicación del *Léxico disponible de Puerto Rico* en 1999, que tiene, entre otras, la finalidad de mejorar la planificación educativa de enseñanza-aprendizaje del español. A diferencia de los trabajos sobre disponibilidad precedentes, que habían usado como informantes a niños de edad entre los 9 y los 12 años o adolescentes, él se sirve de alumnos universitarios del primer año, de aproximadamente 18 años. De esta manera intentaba acercarse lo más posible a la norma léxica adulta, sin que fuera ya influenciada por tecnicismos propios de las diversas facultades (cf. López Morales, 1999: 27-29). Trabaja con listas abiertas e intenta mantenerse lo más fiel

¹² Los centros de interés empleados fueron: 1) “les parties du corps”, 2) “la nourriture et les boissons”, 3) “les vêtements (pour hommes et pour femmes)”, 4) “la maison (matériaux de construction et meubles)”, 5) “le travail au champ”, 6) “la cuisine (ce qu'on y voit)”, 7) “la salle de classe (ce que tu vois et/ou utilises)”, 8) “les moyens de locomotion (ce qui te permet de te déplacer)”, 9) “la musique et la danse (les genres)” y 10) “les jeux et divertissements”.

posible a los centros de interés usados por los estudiosos franceses para “permitir la comparación entre materiales hispánicos americanos y europeos, en primer lugar, y también entre resultados logrados entre nosotros y el mundo franco hablante, es ya de por sí, de importancia sobresaliente” (López Morales, 1999: 37).

La aparición de estos estudios de disponibilidad léxica, como veremos sucesivamente, despertó el interés de muchos investigadores tanto en América Latina como en España. El número creciente de publicaciones animó a López Morales a coordinar las investigaciones y proponer unos criterios metodológicos y de edición unificados, para permitir la comparación entre los resultados. De esta manera, a finales de los años noventa nació el *Proyecto Panhispánico de Disponibilidad Léxica* (PPHDL) con el objetivo de publicar el léxico disponible común del mundo hispánico y de cada uno de los países que forman la comunidad hispánica. Las características metodológicas, como explica Bartol Hernández en el mismo artículo (2006: 380-381), son:

- a. Se trabaja con 16 centros de interés (campos léxicos)¹³.
- b. Los informantes son alumnos del último año de bachillerato o curso equivalente.
- c. Los informantes responden por escrito a la encuesta y disponen de 2 minutos para cada campo léxico.
- d. Se tienen en cuenta las siguientes variables sociales: sexo, nivel sociocultural, titularidad del centro de estudio y ubicación del centro de estudio.
- e. Se siguen unas mismas normas de edición (homogeneización de las respuestas)¹⁴.
- f. Se aplica el mismo programa informático para calcular el índice de disponibilidad.

A partir de estos criterios es posible dividir las publicaciones en tres bloques: aquellas que siguen todas las características sobre citadas y pertenecen al ámbito del PPHDL, los estudios que no se ajustan enteramente a estas características porque por ejemplo toman informantes de niveles educativos diferentes y, los estudios de disponibilidad léxica aplicados a estudiantes de español como segunda lengua o como lengua extranjera¹⁵.

¹³ Estos corresponden a los centros de interés empleados por los lingüistas franceses que publicaron el *Français Fondamental*. Cf. *Ut supra*.

¹⁴ Las normas de edición se pueden consultar están descritas por José A. Samper (1998). En nuestro estudio también utilizamos estos criterios de edición, como analizaremos sucesivamente.

¹⁵ Para una división de las publicaciones en los tres grupos, véase José A. Bartol Hernández (2006).

Por lo que se refiere a la fórmula para calcular el índice de disponibilidad, tras unas propuestas iniciales Lorán-López Morales¹⁶, se fijó la fórmula de López Chávez y Strassburger que “parte de análisis de una matriz de vectores en la que participan: a) la frecuencia absoluta que obtuvo la unidad en cada posición de la lista, b) la frecuencia absoluta de la unidad, que resulta de sumar las diferentes frecuencias alcanzadas en cada posición, c) el número de sujetos que participan en la encuesta, d) el número de posiciones alcanzadas en el centro de interés analizado y e) las posiciones en que fue escrita cada unidad” (López Morales, 1999: 18). A partir de esta fórmula Moreno Fernández y García de las Heras desarrollaron el programa *Lexidisp*, que además de calcular el índice de disponibilidad léxica¹⁷, calculaba también la frecuencia relativa (número de veces que se aparece una palabra con respecto al total), la frecuencia acumulada (suma de las frecuencias relativas) y el índice de aparición (porcentaje de encuestados que la usan) (cf. Serfati, 2017: 114). Sucesivamente se ha empleado el programa *Dispolex* que permitía también combinar la frecuencia con la posición que obtiene una palabra en los listados de cada centro de interés. Además, permitía calcular el *índice de cohesión*: en una escala entre 0 y 1, si las respuestas de los informantes coinciden, el valor se aproxima a 1, y eso significa que el centro de interés analizado es cerrado y compacto, en cambio, cuando las respuestas son muy diferentes, el valor se acerca 0, y el centro de interés es más abierto o difuso (cf. Nalesso, 2019: 135). El último programa desarrollado por la Universidad de Salamanca en colaboración con la Universidad Miguel Hernández de Elche es *LexPro*. Esta herramienta, además de todos los índices de disponibilidad permite también generar grafos y analizar las redes léxicas.

Otro programa muy útil y frecuentemente usado en los estudios de disponibilidad léxica es *DispoGrafo*¹⁸ que permite individuar las redes semánticas que organizan el lexicón mental. El programa, “basándose en las relaciones de secuencia de las palabras disponibles, permite observar de manera gráfica las relaciones que se establecen entre las unidades, pudiendo analizar, además, la robustez o debilidad de estos vínculos” (Sánchez-Saus Laserna, 2011: 84).

El objetivo del *Proyecto Panhispánico* y de la mayoría de los estudios que mencionaremos después, es el de crear un diccionario del léxico disponible del español tanto

¹⁶ Para una descripción más detallada de las primeras fórmulas matemáticas aplicadas, así como de la propuesta de Lorán-López Morales y de López Chávez-Strassburger, véase López Morales (1999) pp. 13-19.

¹⁷ El índice de disponibilidad une los criterios de frecuencia y orden, y se calcula en función de dos criterios: el número de informantes que incorporan la palabra y la posición en la que fue producida. El resultado es un valor entre 0 y 1, cuanto más se acerque a 1 el índice, más disponible es la palabra.

¹⁸ Para una descripción del funcionamiento del programa cf. Echeverría *et al.* (2008)

de Hispanoamérica como de España, gracias al empleo de estos programas. A pesar de la existencia de las pautas para la uniformidad de la metodología, es cierto que los investigadores han ajustado estos criterios a sus necesidades y a la especificidad de la comunidad analizada. A continuación, vamos a exponer algunos de los principales trabajos, a partir de las investigaciones llevadas a cabo en países hispanoamericanos hasta llegar a España¹⁹.

Después de Puerto Rico, otros estudios pioneros sobre disponibilidad léxica se llevaron a cabo en México. Entre los primeros está el que publica Hortensia Justo Hernández en 1986, en el que utilizó listas cerradas de diez palabras por centro de interés y Alva V. Cañizal Arévalo que en 1987 trabaja con seis centros de interés entre los cuales incluye “defectos físicos y morales” para solicitar los adjetivos evaluando la influencia de las variables sexo, tipo de escuela y tiempo dedicado a ver la televisión. Algunos años más tarde en 1991 conduce una investigación sobre el centro “las partes del cuerpo”, a partir de la cual observa que hay algunas agrupaciones relacionadas con los sentidos, que forman el núcleo de la red semántica y que todos los vocablos alrededor tiene una distribución similar. Por eso señala la importancia de identificar los núcleos y las redes semánticas como índices de los niveles de adquisición del léxico para la planificación de la enseñanza, pero sobre todo para describir la sistematización del vocabulario (cf. Carcedo, 1998: 27-28).

Entre las mayores aportaciones mexicanas cabe mencionar los trabajos de Juan López Chávez y Carlos Strassburger Frías, que entre otras cosas crearon la fórmula matemática todavía en uso para el análisis del índice de disponibilidad. López Chávez en 1992 compara los vocabularios disponibles de Madrid, Las Palmas de Gran Canaria, la República Dominicana y Puerto Rico y llega a la conclusión de que el vocabulario común a las varias zonas es bastante limitado, por eso reafirma la importancia de los estudios de disponibilidad léxica para comprobar las similitudes y las diferencia que hay en el español hablado en las varias regiones hispánicas. Otros dos estudios interesantes son los que lleva a cabo en 1993 con escolares de la ciudad de México de nivel preescolar hasta el sexto grado de la escuela primaria, del cual extrae conclusiones sobre la formación y el desarrollo del lexicón mental, y el estudio publicado en 1994 realizado con estudiantes del quinto curso de bachillerato que sometió a una prueba de disponibilidad léxica para solicitar verbos y observar las diferencias internas de estabilidad de esta categoría gramatical (cf. Carcedo, 1998: 29-30).

¹⁹ Para la elaboración de esta lista nos basado mayoritariamente en las listas presentadas en: Carcedo (1998), Sánchez-Saus Laserna (2011), Casanova (2017).

En los mismos años, Mónica Mena Osorio empieza los primeros estudios de disponibilidad léxica en Chile, aunque es Max Echeverría el que impulsó este sector con la publicación en 1987 de un trabajo de descripción cuantitativa y cualitativa del léxico disponible de estudiantes chilenos. En esta investigación introduce una novedad dividiendo los diez centros de interés estudiados en dos grupos: los que tienen un “alto grado de cohesión semántica” (cuerpo humano, medios de transporte, ropa, muebles, herramientas y materiales de construcción, alimentos y animales) y los con un “bajo grado de cohesión semántica” (naturaleza, la casa, la cocina)²⁰. Además, junto con Alba Valencia, en 1999, publicó *El léxico disponible en estudiantes chilenos* encuestando 2 100 estudiantes.

En la República Dominicana Orlando Alba publica en 1995 el *Léxico disponible de la República Dominicana* conteniente los primeros datos de disponibilidad léxica del país elaborados a partir de los dieciséis centros de interés clásicos y con listas abiertas. En 2000 publica un análisis de los anglicismos y una comparación entre los vocablos del corpus dominicano con los de Puerto Rico, Madrid, México y la ciudad de Concepción en tres centros de interés, el cuerpo humano, medios de transporte y alimentos, demostrando un alto grado de coincidencia del léxico disponible de cada zona (cf. Sánchez-Saus Laserna, 2011: 85).

Marielos Murillo Rojas publica los primeros trabajos de Costa Rica en 1993 y 1994. Él trabajó con niños preescolares a los que hizo dos pruebas a partir de estímulos pictográficos, una al principio y la otra al final del curso para hacer un análisis longitudinal-comparativo y monitorear el desarrollo producido en ese tiempo (cf. Carcedo, 1998: 32-33). En 2006 publica, junto a Sánchez Corrales, un diccionario de disponibilidad de niños entre los cinco años y medio y los seis y medio (cf. Sánchez-Saus Laserna, 2011: 85).

Por lo que se refiere a los otros países hispanoamericanos, mencionamos algunas investigaciones que se han llevado a cabo en los últimos años basándose en los criterios del *Proyecto Panhispánico*, como en Argentina conducidas por Ana María Saine y Julia Dalurzo, en Uruguay por Carlos Jones y Ricardo da Silva Rodrigues, en Colombia por María Clara Henríquez Guarín, en El Caribe por Yolanda Rodríguez, en Panamá por Margarita Vázquez y en Cuba por Nuria Gregori (cf. Bartol Hernández, 2006: 382-284).

En España los estudios de disponibilidad léxica empiezan algunos años más tarde en comparación con los de Hispanoamérica conducidos por López Morales en los setenta. El

²⁰ Esta distinción está basada en el índice de cohesión que presentamos en las páginas anteriores.

primer trabajo peninsular es la tesis doctoral de Azurmendi Ayerbe de 1983, que pretende elaborar un modelo de estudio para la descripción del bilingüismo en general, y en particular del bilingüismo euskera-español. Ella analiza cuatro centros de interés, ofreciendo listas de veinte palabras y diez minutos para cada centro y, considerando las variables sociolingüísticas lleva a cabo un estudio cuantitativo y cualitativo que le permite elaborar varios diccionarios (cf. Carcedo, 1998: 34).

Solamente en 1992 Benítez Pérez publica su investigación incorporándose al Proyecto empezado y dirigido por López Morales. Él trabaja con estudiantes del área metropolitana de Madrid y utiliza los dieciséis centros de interés tradicionales de manera que se pueda conducir un análisis de comparación con otros resultados y para efectuar una evaluación del grado de disponibilidad de una misma forma léxica en distintas comunidades de habla. En los años siguientes el mismo autor se ocupará del análisis de los manuales para la enseñanza del español a extranjeros evidenciando varias insuficiencias²¹ (cf. Carcedo, 1998: 34-35). Estudios más recientes en la comunidad de Madrid han sido publicados por Paredes García, y Gómez Sánchez y Guerra (cf. Sánchez-Saus Laserna, 2011: 87-88).

A partir de estos dos primeros estudios españoles de 1983 y 1992, las investigaciones sobre disponibilidad léxica empezaron a ampliarse a toda la península. Aunque nuestro elenco no tiene la pretensión de considerarse exhaustivo, vamos a citar algunas obras divididas por comunidad autónoma.

En el País Vasco, además de la tesis doctoral de Azurmendi, el siguiente trabajo que se publicó es lo de Maitena Etxebarria en 1996 que investigó la disponibilidad léxica de escolares bilingües del área metropolitana de Bilbao. Ella evalúa la incidencia del modelo de enseñanza (principalmente en español o principalmente en euskera) e identifica pequeñas diferencias entre los resultados de los dos grupos.

En Andalucía los primeros que empezaron a investigar fueron Joaquín García Marcos y M^a Victoria Mateo García al principio de los años 90. García Marcos, en colaboración con López Morales, publicó en 1995 sus resultados procedentes de los dieciséis centros de interés tradicionales a los que añade “los colores” y “la mar”. Mateo García, confrontó los resultados que obtiene en Motril (Granada) con los obtenidos por Benítez Pérez en 1992 en Madrid y Ruiz Basto en 1987 en México sobre el centro “alimentación”. De su análisis, además de encontrar

²¹ El tema de los manuales de ELE se analizará sucesivamente en 1.2.2.

diferencias cuantitativas, emergen diferencias cualitativas notables: a pesar de que la mayoría de los vocablos pertenece al español estándar y hay pocas palabras conectadas con la variedad geográfica, sólo tres vocablos están presentes en las tres listas (cf. Carcedo, 1998: 37-39).

En 2015 Prado Aragonés y Galloso publicaron *El léxico disponible de Extremadura y comparación con el de Andalucía*. En esta obra, además de presentar el léxico disponible de la comunidad autónoma, comparan sus resultados con los obtenidos en las ocho comunidades de Andalucía con el objetivo de comprobar la convergencia entre las dos zonas dialectales cercanas. Otros estudios en las zonas más meridionales de la península fueron llevados a cabo en Ceuta y Melilla y dieron origen a los relativos diccionarios de disponibilidad léxica. La *Disponibilidad léxica en Ceuta* fue publicada en 2006 por María Aroya Esteban y el *Léxico disponible de Melilla* en 2008 por Gérard Fernández Smith. En varios artículos describen la situación de las dos ciudades en las que una parte de la población es monolingüe española, y la otra es bilingüe o trilingüe cuya lengua materna es el árabe dialectal marroquí o el dialecto bereber tamazight. Confrontando los resultados obtenidos en las dos ciudades se vio que los hablantes que tienen como lengua materna el español, producen un mayor número de palabras (cf. Sánchez-Saus Laserna, 2011: 91). Otro caso de bilingüismo analizado es lo de la comunidad de Gibraltar, donde el inglés es la única lengua oficial pero el español se usa en muchos ámbitos cotidianos y sobre todo en el familiar. Luis Escoriza Morera se ha ocupado de analizar el contacto español-inglés y el multilingüismo de esta zona.

En Castilla-La Mancha las investigaciones fueron dirigidas por Natividad Hernández Muñoz y llevaron a la publicación del diccionario de disponibilidad léxica de la comunidad autónoma en 2005. En Castilla y León los estudios fueron realizados por Galloso que empezó por Salamanca, Ávila y Zamora al principio del siglo, mientras que José Antonio Bartol Hernández se ocupó de la provincia de Soria (2004), Fernández Juncal de Burgos (2008), Santiago Guervós de Segovia. Raquel Cruz Alonso, en su tesis doctoral de 2012, presenta el trabajo relativo a toda la región, *El léxico disponible de Castilla y León*.

Por lo que se refiere a la comunidad de Aragón, las investigaciones fueron llevadas a cabo por un equipo de lingüistas dirigidos por María Luisa Arnal que en 2004 publicó *El léxico disponible de Aragón*. Siguiendo las pautas del *Proyecto Panhispánico*, Belén López Meirama publica el *Léxico disponible en español de Galicia* en 2008, aunque modifica la variable urbano-rural dividiéndola en cuatro y añade una quinta variable: “lengua materna”. En las otras comunidades del norte, Carmen Fernández Juncal en 2013 publica *El léxico disponible de*

Cantabria y Alberto Carcedo González en 2001 el de Asturias. En la comunidad de Navarra los estudios han sido conducidos por Maria Areta Lara que en 2009 presenta su tesis doctoral *El léxico disponible de los estudiantes preuniversitarios navarros*. Otros estudios que relacionan la disponibilidad léxica con el contacto de lenguas o los regionalismos han sido publicados en 2008 por Cristina Tabernero. En La Rioja Emma Pérez Jiménez presenta su tesis *El léxico disponible de los estudiantes preuniversitarios en La Rioja* en 2016 (cf. Sánchez-Saus Laserna, 2011: 86-87).

En Cataluña las investigaciones empiezan a aparecer desde 2003 con algunas publicaciones de Maribel Serrano Zapata, que culmina su trabajo con la presentación de su tesis doctoral: *Disponibilidad léxica en la provincia de Lleida*. Ella trabaja con una muestra de 240 informantes y comparado su productividad en las dos lenguas oficiales de la provincia: el castellano y el catalán. En los años siguientes, un equipo de investigadores ha recogido y publicado los materiales de Barcelona, Gerona y Tarragona (cf. Sánchez-Saus Laserna, 2011: 87).

En la isla de Gran Canaria los estudios comenzaron en 1997 con un artículo introductorio de José Antonio Samper Padilla y Hernández Cabrera, en el que explicaban las características metodológicas de su trabajo que tiene las mismas pautas del *Proyecto Panhispánico*, excepto por el utilizo de un centro de interés más, el de los colores, y la consideración de una variable acerca de la ubicación del centro. En los años siguientes se publican varios trabajos sobre todo de carácter comparativo: en 1999 Samper Padi lla compara los resultados grancanarios con los de Puerto Rico obtenidos por López Morales; en 2001 Hernández Cabrera y Samper Hernández los comparan con la norma culta de Las Palmas; en 2003 Samper, Hernández y Bellón con los resultados de Córdoba y en 2008 con los de Aragón por Samper Hernández (cf. Sánchez-Saus Laserna, 2011: 88).

Finalmente, de la Comunidad Valenciana se han ocupado María Begoña Gómez Devís, que, bajo la dirección de José Ramón Gómez Molina, en 2004 publica su tesis doctoral: *La disponibilidad léxica de los estudiantes preuniversitarios valencianos*. Esther Martínez Olmos en 2008 publica lo de Alicante y Manuela Casanova Avalos el léxico disponible de la provincia de Castellón. También se han llevado a cabo estudios de disponibilidad léxica sobre el catalán de Valencia, cuyos resultados han sido publicados por Francesc Llopis Rodrigo en 2008 (cf. Sánchez-Saus Laserna, 2011: 86-87). Además, junto con Gómez Devís lleva a cabo una investigación con los estudiantes de segundo de bachillerato para mostrar el grado de

convergencia semántica del léxico disponible del español y del catalán (cf. Gómez Devís y Llopis Rodrigo, 2010: 53).

Estos son solamente algunos de los trabajos sobre disponibilidad léxica con hablantes nativos de español. Otros recuentos parecidos al nuestro se pueden encontrar en varios artículos y tesis doctorales²². Además, se pueden encontrar otras publicaciones que rastrean la historia de la disciplina en la página *Dispolex*.

En las últimas décadas ha vuelto, en los estudios de disponibilidad léxica, la preocupación que movió los lingüistas franceses de los años cincuenta: ¿cómo mejorar la enseñanza-aprendizaje de las lenguas? ¿Cuáles son las unidades que deben formar parte del primer vocabulario que un alumno tiene que aprender? Es decir, ha renacido el interés para encontrar maneras nuevas y más eficaces para la enseñanza de las lenguas en aprendices extranjeros. En este caso también, como por los hablantes nativos, la mayoría de las investigaciones se llevaron a cabo en ámbito hispánico.

1.2.2. Estudios en ELE

Los primeros estudios aplicados a la enseñanza de Español como Lengua Extranjera (ELE) han empezado a aparecer desde finales de los años noventa, aunque han tenido, y siguen teniendo, su mayor desarrollo en los últimos años. Ya en 1998, Carcedo hizo una lista de los principales objetivos que podría tener la aplicación de los estudios de disponibilidad léxica en la didáctica de ELE (pp. 54-55), más recientemente, en 2015, Paredes García retoma estos objetivos y los reformula de esta manera:

1. analizar el léxico disponible de los aprendices en un determinado momento;
2. definir las etapas de aprendizaje;
3. estudiar las interferencias de la L1 y de otras LE conocidas;
4. cotejar el vocabulario de los participantes con otros grupos de estudiantes y con los nativos;
5. establecer cuáles son las unidades poco rentables o los déficits léxicos;
6. proporcionar comparaciones culturales;
7. examinar la adecuación de los manuales y los otros materiales pedagógicos;
8. eliminar el léxico superfluo de los contenidos didácticos y llevar a cabo una selección de los contenidos apropiada;

²² Para recuentos de trabajos sobre disponibilidad léxica se aconseja: Carcedo (1998), Bartol Hernández (2006), y en las tesis de Sánchez-Saus Laserna (2011) y Casanova (2017).

9. ampliar la cobertura de los análisis hacia la terminología (español para fines específicos).

Antes de exponer las investigaciones llevadas a cabo con aprendices de ELE como muestra (objetivo 1 de la lista de Paredes García), mencionamos algunos estudios que se han ocupado de evaluar la selección del vocabulario en los manuales de español para estudiantes extranjeros (objetivos 7 y 8).

Uno de los primeros lingüistas españoles que evidencia la inadecuación del léxico usado en los manuales es Benítez Pérez que, en 1994, analiza siete manuales publicados en España y tres al extranjero y descubre que en el conjunto de la selección de palabras hay muchas con un alto índice de disponibilidad, pero también hay varias con valores muy bajo o equivalente a 0. El año anterior lleva a cabo, junto con Zebrowski, un análisis de cuatro manuales polacos frecuentemente usados para la enseñanza del español y concluyen afirmando que los autores no han seleccionado de manera adecuada el vocabulario y que no han aprovechado de los estudios sobre disponibilidad léxica y reafirman la idea de que estos estudios podrían ser muy útiles a la hora de crear los materiales didácticos (cf. Benítez y Zebrowski 1993: 229 *apud* Carcedo 1998: 40).

Algunos años más tarde, María Victoria Mateo García analiza el léxico sobre la alimentación de cinco manuales y relevan unas discrepancias notables sobre todo por lo que se refiere a la ausencia de términos considerablemente disponibles, la presentación como sinónimos de palabras que en realidad hacen referencia a dos elementos y la diversificación en los inputs en los materiales. A partir de este estudio concluyen que es necesario seguir evaluando los manuales para poder resolver las faltas en la selección del vocabulario, y que un instrumento muy útil podrían ser los estudios sobre disponibilidad léxica (cf. Sánchez-Saus Laserna 2011: 96).

El primero en aplicar los estudios de disponibilidad léxica a hablantes no nativos es Alberto Carcedo que analiza el vocabulario de hablantes de lengua finesa. Él publica en 1998 y 1999 los resultados de sus estudios conducidos con una muestra de 78 estudiantes de español del último año de bachillerato que examina, en el primer caso con el objetivo de conocer las tendencias generales de estos hablantes y en el segundo para analizar los errores léxicos. En el año 2000 publica una monografía *Disponibilidad léxica en español como lengua extranjera: el caso finlandés* en la que analiza cuantitativa y cualitativamente el léxico disponible de los estudiantes y la influencia de las variables: nivel de estudios, sexo, lengua materna (fines o

sueco), conocimiento de otra(s) lengua(s) románica(s) y libro de texto utilizado. Sucesivamente compara los resultados con los de República Dominicana, Puerto Rico, Cádiz y Zamora y elabora un diccionario de disponibilidad de su muestra. A partir de estos análisis pone en evidencia que el léxico disponible de sus informantes no era conforme a las reales necesidades comunicativas, e invita a tener en cuenta los estudios de disponibilidad léxica a la hora de diseñar los materiales de ELE (cf. Nalesso, 2019: 100-101).

A partir de la publicación de este trabajo, muchos otros estudiosos se han dedicado a la investigación de disponibilidad léxica tanto en el país de origen de los aprendices, como en contexto de inmersión lingüística. De este segundo grupo, señalamos la publicación de Marta Samper Hernández en 2002 que trabajó con una muestra de 45 estudiantes con distintas lenguas maternas de los cursos de español como lengua extranjera de la Universidad de Salamanca. Ella se atiene a los centros de interés clásicos y evalúa las variables: sexo, nivel (inicial, intermedio, avanzado o superior), lengua materna (inglés, italiano, japonés y otros) y conocimiento de otras lenguas. Entre sus conclusiones, destaca que las variables sexo y conocimiento de otras lenguas no influyen en la cantidad de palabras producidas, aunque el nivel de español sí es relevante. De hecho, ella nota un aumento progresivo desde el nivel inicial al avanzado que luego desciende desde el avanzado al superior; sin embargo, subraya que en algunas ocasiones el mayor dominio de la lengua se demuestra no con un mayor número de unidades, sino con la producción de palabras menos frecuentes y más complejas (cf. Sánchez-Saus Laserna, 2011: 97-98).

Otro trabajo interesante de esta autora, es un artículo publicado en 2001, en el que expone unas innovaciones metodológicas necesarias al trabajar con informantes extranjeros. Ella parte de los criterios que se utilizan en el *Proyecto Panhispánico* (cf. Samper Padilla, 1998) y considera necesario adaptarlos a las exigencias de los estudios aplicados a la enseñanza de ELE, puesto que pueden estar presentes palabras de la lengua materna de los hablantes o los errores que se registran pueden ser peculiares e influenciados por la ortografía de la lengua materna, etc. (cf. Sánchez-Saus Laserna, 2011: 98).

Otro de los primeros estudios realizados en el país de origen de los informantes son los realizados por Hugo en Alemania y publicado en 2003, *Disponibilidad léxica en español como lengua extranjera*, en el que se analiza la disponibilidad léxica en español de 84 estudiantes alemanes de la región del Saare en los últimos dos años de instituto. Él también llega a la

conclusión de que el nivel de conocimiento del español es el que más influye en los resultados cuantitativos.

Teniendo en cuenta los resultados obtenidos de la aplicación de los estudios de disponibilidad léxica en muchos países, como Finlandia, Alemania, China, Estados Unidos, Japón, Eslovenia, Portugal, etc.²³, en Italia el interés hacia esta disciplina ha empezado en 2013 y sigue ampliándose en los últimos años.

En 2013 Silvia Caggiula publica *El español como lengua extranjera: un estudio de disponibilidad léxica y su aplicación a la enseñanza*, conteniente los resultados de su investigación en la que examina 50 estudiantes de ELE de nivel preuniversitario, universitario y adultos que estudian en academias privadas utilizando diez centros de interés entre los cuales incluye *El trabajo* (cf. Tomé Cornejo, 2015: 332).

El siguiente trabajo fue conducido por Rubio Sánchez (2015) que analiza el léxico disponible de treinta y seis estudiantes preuniversitarios italianos que se encontraban en Salamanca para asistir a un curso de inmersión lingüística de una semana. Dado que este trabajo tiene una muestra limitada, el autor en las conclusiones se plantea el objetivo de profundizar en el análisis del léxico disponible de estudiantes italianos de español, reconociendo las numerosas posibilidades que la disponibilidad léxica puede ofrecer a la enseñanza de ELE. Algunos años después publica los resultados de otro trabajo llevado a cabo con 173 estudiantes italianos de español en institutos de secundaria de diferentes ciudades italianas. En este trabajo reduce a 14 los centros de interés tradicionales²⁴ y añade siete nuevos: Acciones y actividades habituales, España y los españoles, Aspecto físico y carácter, La lengua española, Italia y los italianos, Partes del cuerpo (en italiano), Comidas y bebidas (en italiano), con el objetivo de obtener palabras de categorías gramaticales distintas, de examinar la concepción que tienen los informantes de la cultura española e italiana y para una comparación entre el número de vocablos producidos en español y en italiano (cf. Rubio Sánchez, 2017: 145).

Varios estudios introductorios han sido publicados por del Barrio de la Rosa en 2016, 2017 y 2018 que trabaja con muestras de estudiantes preuniversitarios y universitarios. En 2018 publica, junto a Michaela Mae Vann el *Disponibilidad léxica de los estudiantes de español en Italia. Estudio y diccionario*. Además de un estudio de disponibilidad léxica, los autores nos

²³ Para más ejemplos de trabajos conducidos con aprendices de ELE véase Tomé Cornejo (2015).

²⁴ Los centros de interés investigados son: 3. Partes de la casa (sin los muebles), 4. Los muebles de la casa, 6. Objetos colocados en la mesa para la comida, 7. La cocina, 9. Calefacción e iluminación, 10. La ciudad, 11. El campo, 12. Medios de transporte y 13. Trabajos del campo y del jardín.

ofrecen unos diccionarios de léxico disponible divididos por grupo de informantes. De hecho, las encuestas se llevaron a cabo con la participación de 133 estudiantes divididos en: estudiantes preuniversitarios, estudiantes preuniversitarios de origen hispánico, nacidos o residentes en Italia que puedan considerarse bilingües, y estudiantes de la Universidad Ca' Foscari de Venecia. Dada la naturaleza distinta de los tres grupos, se realizaron también test acerca de las actitudes hacia el español para los estudiantes bilingües y acerca de las motivaciones que llevaban a los otros estudiantes a aprender español. Después de un análisis de los resultados, concluyen su trabajo confirmando “la utilidad de los listados de disponibilidad a la hora de proporcionar criterios objetivos para la selección y gradación de unidades léxica con fines didácticos” (Del Barrio y Mae Vann, 2018: 81).

Acabamos esta sección dedicada a los trabajos con aprendices italianos mencionando los trabajos conducidos por Nalesso (2018, 2020), cuya tesis doctoral (2019) consiste en un experimento transversal y longitudinal dirigido al estudio de estudiantes italianos de ELE de la Universidad de Padua. En este trabajo realiza un estudio de disponibilidad léxica y de riqueza léxica con el propósito de examinar el vocabulario activo de los informantes. En las conclusiones afirma que los resultados obtenidos “son buenos, no sobresalientes como se esperaba debido a la afinidad que une el italiano al español” (p. 370). De toda manera espera “haber contribuido a poner los cimientos de un nuevo sistema capaz de mejorar la ejecución de los estudios sobre competencia léxica de los aprendices de ELE, basados en la correlación de los datos procedentes de los análisis de disponibilidad léxica y de riqueza léxica” (p. 371).

CAPÍTULO II

DESCRIPCIÓN DE LA ENCUESTA

Antes de describir en detalle la muestra y la prueba, mencionamos la metodología con la que se presentó la encuesta. Para cada participante se prepararon cinco páginas: tres correspondían a la prueba de disponibilidad léxica, una por cada centro de interés, y dos para el cuestionario sociológico. Primero la investigadora se presentó a los participantes y explicó las finalidades académicas del estudio; luego se les explicó en qué consistía la prueba y se suministraron las hojas; se les recordó que todas las pruebas eran anónimas y que en ningún caso iban a ser sujetos a evaluación. También se les aconsejó que escribiesen todas las palabras que se les ocurrieran, aunque no estuviesen seguros de la ortografía porque luego iba a haber una corrección y edición de los datos. Toda esta primera parte de introducción y explicación oral se dio en italiano, mientras que el enunciado de las hojas de la prueba de disponibilidad, así como las preguntas del cuestionario sociológico estaban escritas en español. Preferimos usar su lengua materna por tratarse de estudiantes italianos de ELE, cuyo nivel de español no conocíamos, para que se sintiesen cómodos.

Después de esta explicación inicial, empezó la primera parte de la prueba con la entrega a cada uno de los participantes de la primera página del test al revés; cuando todos la tenían podían darle la vuelta, leer las instrucciones y empezar a escribir. Cada página estaba encabezada por la misma frase: “Escribe todas las palabras que se te ocurran durante 2 minutos sobre el siguiente tema”, seguida por el título del centro de interés y una tabla con dos columnas de 50 líneas numeradas cada una, por un total de 100 líneas. A diferencia de otros trabajos que dejaron también unas líneas sin numeración para dar la idea de que se trataba de listas abiertas (cf. Del Barrio & Mae Vann, 2018: 22), en nuestro caso hemos elegido numerarlas todas porque creemos que un número tan alto puede dar la misma impresión y ofrecer espacio suficiente para rellenarlas en el tiempo dado.

Por lo que se refiere a la limitación temporal, nos hemos conformado con las pautas del *Proyecto Panhispánico*, y como afirma Samper Hernández, la limitación era necesaria para “que no quepan entre las respuestas aquellas palabras que puedan ser producto de un largo –y artificial– tiempo de meditación, [o sea] palabras que el hablante no utilizaría quizás nunca”. Además, como especifica ella misma, aunque se trabaje con estudiantes de ELE y no con

hablantes nativos, “dos minutos era un tiempo suficiente para escribir bastantes unidades e incluso para bucear en su memoria en busca de alguna más” (2002: 21).

Al acabar los dos minutos previstos se recogía la hoja y se entregaba la siguiente, hasta acabar la primera parte de la encuesta. Las hojas con los centros de interés se presentaron una a la vez para evitar que el informante pensase antes de tiempo en las palabras conectadas con el tema después. Sucesivamente se entregaron las dos páginas de cuestionario sociológico y se dejó que los participantes contestasen libremente sin límites de tiempo. Como se analizará posteriormente en la segunda parte de este capítulo, el cuestionario fue necesario para recoger los datos personales de los participantes, su conocimiento y uso del español, así como sus actitudes hacia esta lengua.

Una vez finalizadas las encuestas y editado los datos, siguiendo las pautas que se expondrán en la sección 2.3.2., se procedió con el análisis de los mismos a través de la herramienta *LexPro*, accesible en la página web dispogram.usal.es/lexpro/.

2.1. La muestra

2.1.1. Los informantes

Las encuestas de disponibilidad léxica de las que parte esta investigación fueron llevadas a cabo en los meses de diciembre 2023, enero y febrero 2024 sobre 14 estudiantes universitarios italianos. Todos se encontraban en España para participar a un programa de intercambio entre su universidad italiana y la Universidad de Oviedo. Dado que uno de los objetivos secundarios de este trabajo es investigar sobre la influencia que tiene el contexto inmersivo en el léxico disponible de los estudiantes italianos, las encuestas se llevaron a cabo a finales del primer semestre académico, o al principio del segundo, así que todos los encuestados se encontraban en España desde hace cuatro, cinco o más meses.

En la encuesta participaron 7 mujeres y 7 hombres de edad comprendida entre los 20 y los 28 años. A pesar de ser una muestra reducida, hemos elegido no seguir buscando otros informantes para mantener el equilibrio que habíamos alcanzado entre los dos sexos. Todas las mujeres estudian carreras humanísticas, en concreto, cinco de las participantes estudian lenguas extranjeras, una, ciencias de la educación (Grado en Maestro de Educación Primaria) y otra DAMS (Historia y Ciencias de la Música en la Universidad de Oviedo). En cambio, solo dos de los hombres estudian una carrera humanística (lenguas extranjeras y filosofía), mientras que los demás pertenecen a los departamentos científicos: un participante estudia medicina y los otros cuatro, física. De este último subgrupo aclaramos que dos de los informantes no son

estudiantes de grado o máster, sino de doctorado. Los hemos aceptados dada la dificultad en encontrar otros participantes para alcanzar el equilibrio entre los dos géneros y porque a pesar de ser doctorandos, llevaban en España el mismo tiempo que los de los otros participantes.

Como es usual en otros trabajos de disponibilidad léxica, las variables sociolingüísticas se seleccionaron después de la realización de las pruebas, así que la selección de los informantes ha sido totalmente aleatoria, los únicos requisitos eran ser estudiantes italianos y residir temporáneamente en España desde el principio del año académico.

2.1.2. Las variables seleccionadas

Las variables seleccionadas para este estudio difieren de las que se emplean en los trabajos que forman parte del *Proyecto Panhispánico*. De hecho, en los estudios que se han llevado a cabo en este ámbito con estudiantes de español como lengua materna, las variables consideradas son las siguientes: sexo, tipo de centro escolar (público o privado), localización del centro (rural o urbano) y estrato sociocultural (Bartol Hernández, 2006: 381). En cambio, cuando los estudios sobre disponibilidad léxica se han llevado a cabo con informantes bilingües o aprendices extranjeros de ELE, las variables consideradas se han adaptado a la población objeto del estudio. Por ejemplo, en el primer estudio con hablantes no nativos publicado por Carcedo en 2000, que trabaja con estudiantes finlandeses, se consideran las variables: sexo, lengua materna, conocimiento de otras lenguas románicas y nivel de estudio. Otros trabajos sucesivos consideran: sexo, conocimiento de otras lenguas, nivel de español y lengua materna (cf. Samper Hernández, 2002: 19-20), mientras que Marta Sánchez-Saus Laserna en su estudio de 2011 considera además una quinta variable, la frecuencia de uso del español (pp. 35-36).

Uno de los últimos trabajos italianos, publicado por Del Barrio y Mae Vann en 2018 incluye tres grupos distintos de informantes y adapta las variables a la población, por eso nos encontramos con variables comunes a los tres: sexo, edad y país de nacimiento, a las que se añaden unas variables específicas que, en el caso de los estudiantes preuniversitarios son el curso, las actitudes instrumentales, las actitudes integradoras y la facilidad, por el grupo de estudiantes universitarios a estas siete se añade también la variable sobre los conocimientos previos de español, mientras que por el grupo de informantes bilingües se consideran también cuantos años llevan en Italia (pp. 39-53).

En nuestro caso, dado que queremos comparar nuestros resultados con los de estudiantes italianos en contexto no inmersivo, hemos adaptado nuestras variables a las que se han empleado en estos otros trabajos. Por eso, las variables que seleccionamos son: sexo, carrera,

nivel de español y conocimiento de otras lenguas. En la tabla 2.1. se puede ver la distribución de los participantes por cada una de ellas.

Variable	Número de informante
SEXO	
Mujer	7
Hombre	7
CARRERA	
Lenguas extranjeras	6
Humanidades	3
Medicina	1
Física	4
NIVEL DE ESPAÑOL	
Básico (A)	5
Intermedio (B)	4
Avanzado (C)	5
CONOCIMIENTO DE OTRAS LENGUAS	
Lenguas romances	3
Lenguas no romances	11

Tabla 2.1. Distribución de los encuestados por cada variable

Antes de presentar la distribución de los participantes en función de los factores que hemos tenido en cuenta, queremos especificar las motivaciones que nos llevaron a su selección. La primera variable que vamos a considerar es el “sexo”. Esta es uno de los hilos que conectan todos los trabajos de disponibilidad léxica; en nuestro caso la hemos elegido para comprobar si, como ocurre en varios otros estudios (cf. Del Barrio & Mae Vann, 2018), se puede detectar una mayor producción léxica en las mujeres que en los hombres. La segunda variable que vamos a considerar es la “carrera”. Hemos dividido nuestros informantes en cuatro grupos según la carrera que estudian, así tenemos los alumnos de lenguas, de cursos humanísticos, de física y de medicina. Aunque tuvimos solo un participante que frecuenta la facultad de medicina, decidimos no incluirlo en otro grupo para demostrar como los conocimientos técnicos influyen en la producción de palabras en el centro de interés 01. Partes del cuerpo. Nuestra tercera variable es el “nivel de español”. Como también han hecho otros investigadores (cf. Sánchez-Saus Laserna, 2016), hemos dividido nuestros participantes en tres grupos: A, B, C según las indicaciones que nos dieron a la hora de completar el cuestionario sociológico. En la misma publicación, Marta Sánchez-Saus Laserna explica que no es necesaria una mayor

segmentación dentro de esta variable ya que la mayor diferencia en la producción de léxico se da entre el nivel inicial y el intermedio, entre A y B, mientras que, entre el nivel intermedio y avanzado es más contenida (p. 36). La última variable que consideramos es el “conocimiento de otras lenguas”, en concreto, nos preguntamos si el conocimiento de otra lengua romance como lengua extranjera, además del español, pueda influir en el léxico de esta última. Por eso dividimos nuestros informantes entre los que conocen otras lenguas, pero ninguna romance (LENoRom) y los que conocen otras lenguas romances (LERom). Sánchez-Saus Laserna también considera esta variable, pero la evalúa solo para los estudiantes de español que no tengan una lengua romance como lengua materna. De hecho, ella excluye de su grupo, para esta variable, los estudiantes italianos y franceses, y considera solamente los informantes de lengua inglesa, finesa, polaca y alemana teniendo en cuenta si, además del español, conocen otra lengua extranjera romance, si ninguna de las lenguas extranjeras que conocen es romance y si no conocen ninguna lengua extranjera. Ella parte de la hipótesis de que, si un estudiante conoce otra lengua romance, además del español, esta pueda influir en la amplitud del léxico en español, respecto a los alumnos cuya única lengua extranjera romance sea el español (pp. 36-37).

		SEXO	
		Hombres	Mujeres
CARRERA	Lenguas extranjeras	1	5
	Humanidades	1	2
	Medicina	1	0
	Física	4	0
NIVEL	Básico (A)	5	0
	Intermedio (B)	1	3
	Avanzado (C)	1	4
OTRAS LENGUAS	LENoRom	6	5
	LERom	1	2

Tabla 2.2 Distribución de los participantes en función de la variable Sexo

Como se puede ver a partir de la tabla 2.2., que ilustra la distribución de los participantes en función de la variable “sexo”, las mujeres son todas estudiantes de lenguas extranjeras o de humanidades, la mayoría de ellas tiene un nivel avanzado de español, algunas certificadas mediante DELE, otras a través de otras calificaciones. Los hombres, en cambio, tienen una distribución más variada según la variable “carrera”, de hecho, la mayoría son estudiantes de física (4/7), los otros tres están distribuidos uno por cada una de las otras carreras. Probablemente esto influye en su nivel de español: cinco de ellos tienen un nivel inicial o nunca

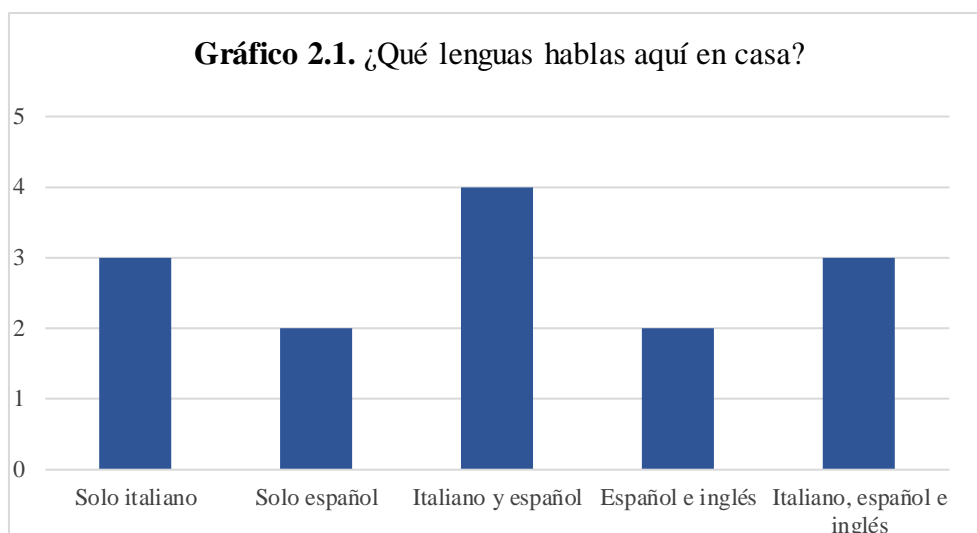
lo han estudiado como lengua extranjera, uno tiene un nivel intermedio y el único estudiante que tiene un nivel avanzado es el estudiante de lenguas extranjeras.

Por lo que se refiere al conocimiento de otras lenguas, todos los informantes declaran de conocer y haber estudiado inglés. Destaca que solamente tres informantes, además del español conozcan otra lengua romance (el francés o el portugués), y los tres son estudiantes de lenguas. En cambio, la mayoría de los demás ha estudiado solo una lengua extranjera (el inglés) y algunos, otra lengua no romance (el alemán).

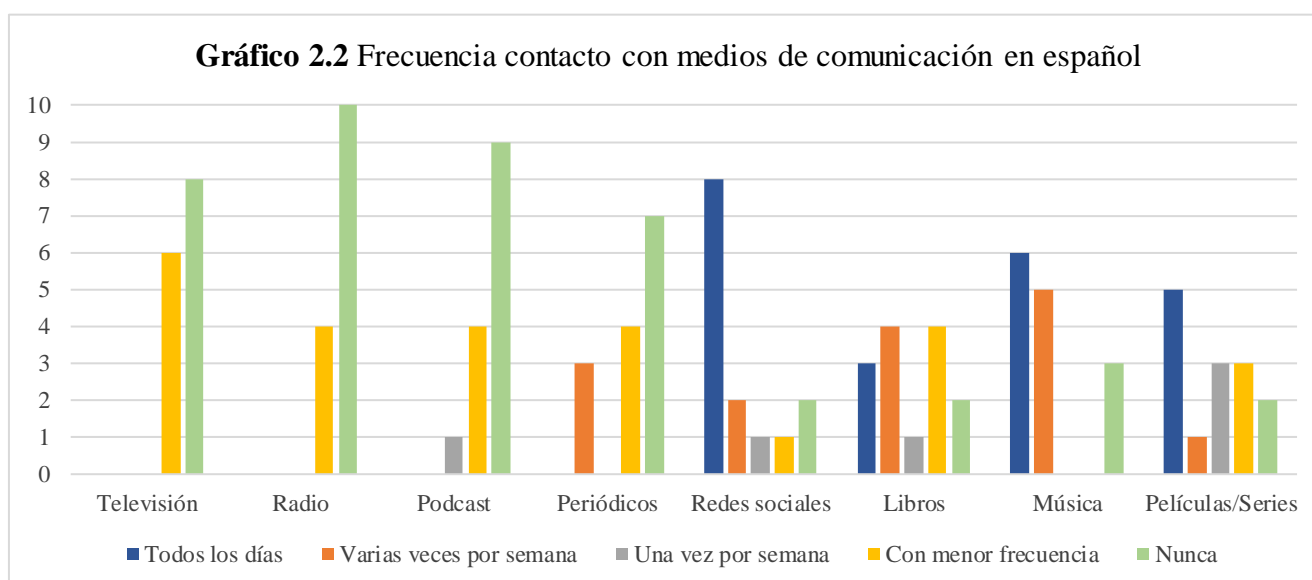
2.2. El cuestionario sociológico

Después de haber hecho la primera parte de prueba de disponibilidad léxica propiamente dicha, se les ha pedido a los estudiantes que respondieran a unas preguntas sociológicas. Este cuestionario está compuesto por dos partes: una primera para recoger sus datos personales y su uso del español, y una segunda para ver su opinión y actitud hacia esta lengua. No se ha establecido un límite de tiempo para rellenar esta parte, sino que, al ser presentada como última fase de la encuesta, cada estudiante tenía todo el tiempo que necesitaba para completarla.

Además de las informaciones que hemos usado como variables (sexo, carrera, nivel de español y conocimiento de otras lenguas), en el cuestionario sociológico también hicimos otras preguntas acerca del uso de las lenguas que conocen. De las respuestas emerge que todos los informantes hablan italiano en su casa en Italia, en cambio, en casa en Oviedo la situación es más variada. Como representa el gráfico 2.1., la mayoría sigue hablando total o parcialmente en italiano, en cambio solo cuatro personas no lo usan cotidianamente con sus compañeros de piso. Por lo que se refiere al uso del español 11 informantes lo usan como única lengua o junta con otras, eso nos permite evaluar con un grado mayor la influencia del contexto inmersivo.

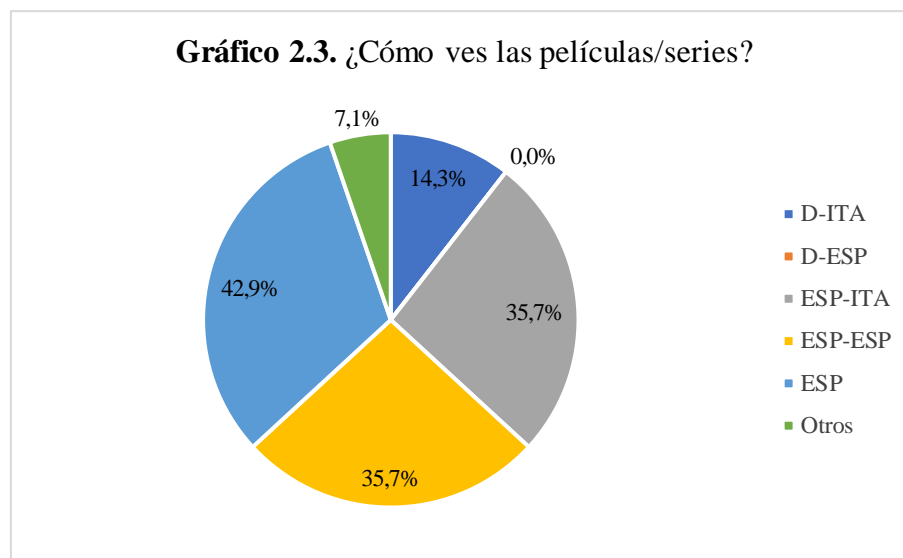


Además, acerca del contacto y uso del español, ante la pregunta de si consuman materiales de los medios de comunicación en lengua española, la mayoría ha respondido que tiene contactos escasos o nulos con la televisión, la radio y los podcasts. La frecuencia de uso aumenta ligeramente con los periódicos, mientras que los medios de comunicación más usados son las redes sociales (8 informantes afirman usarlas diariamente), la música (6 escuchan música en español cada día) y las películas o las series TV (5 informantes). Los libros son, junto con las redes sociales, los medios de comunicación con el que casi todos tienen contacto, de hecho, solamente dos personas afirman que no leen y que no usan las redes sociales.



Relacionado con la pregunta anterior, a través de una pregunta con respuesta múltiple, preguntamos en qué lengua suelen ver las películas y las series TV. Como nos enseña el gráfico 2.3., dos personas (el 14,3%) ven las películas dobladas en italiano, mientras que nadie dobladas en español. Cinco personas ven las películas o las series en lengua original (español)

y con subtítulos en italiano y otras cinco en lengua original (español) con subtítulos en español. Seis personas las ven en lengua original (español) sin subtítulos y solo una persona suele ver películas o series de otra manera (en concreto en lengua original (inglés) con subtítulos en inglés).



Otra pregunta de esta primera parte del cuestionario sociológico era acerca de sus precedentes estancias en España o en otros países hispanohablantes. Doce participantes ya habían estado en España de vacaciones para unos días, y solamente dos de estos ya habían hecho también una experiencia de intercambio de seis meses y de un año. Por lo que se refiere a los otros países hispanohablantes solo tres personas visitaron uno o más países por un período de un mes.

Por lo que se refiere a la segunda parte del cuestionario, pedimos a los estudiantes de evaluar, a través de una escala Likert, cuanto estuviesen de acuerdo con cinco afirmaciones acerca de su opinión y actitud hacia el español. La escala está compuesta por 4 niveles: 1. Muy de acuerdo, 2. De acuerdo, 3. Poco de acuerdo, 4. En desacuerdo.

Gráfico 2.4. El español me parece fácil de aprender

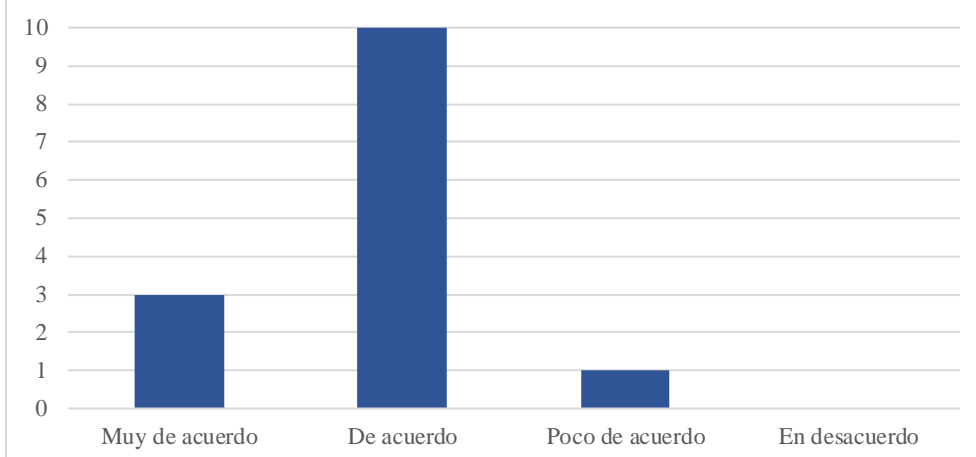


Gráfico 2.5. Aprender español en España me parece más fácil

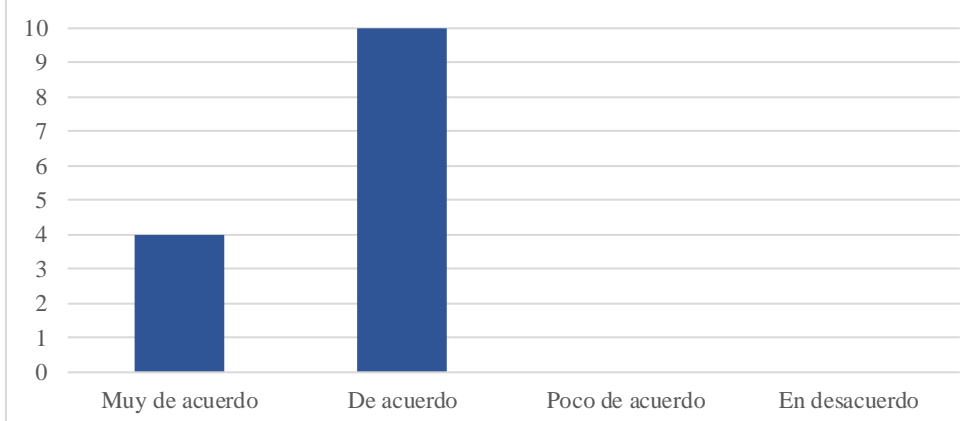
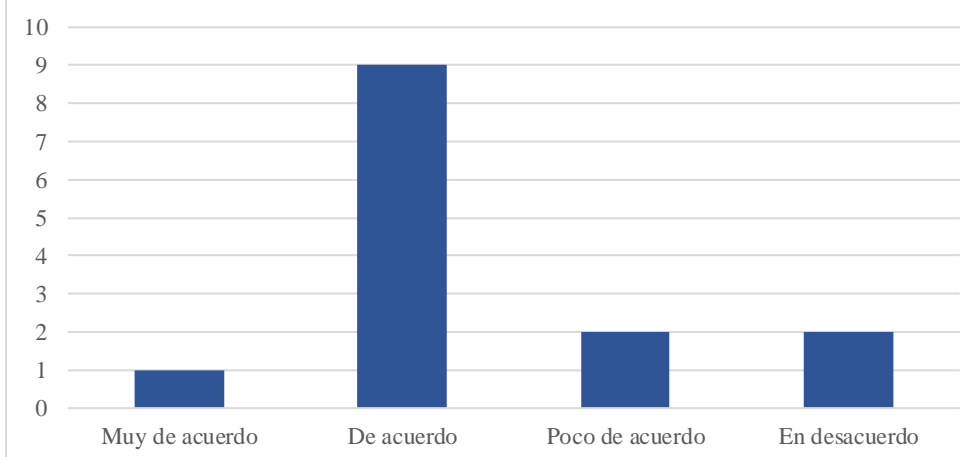
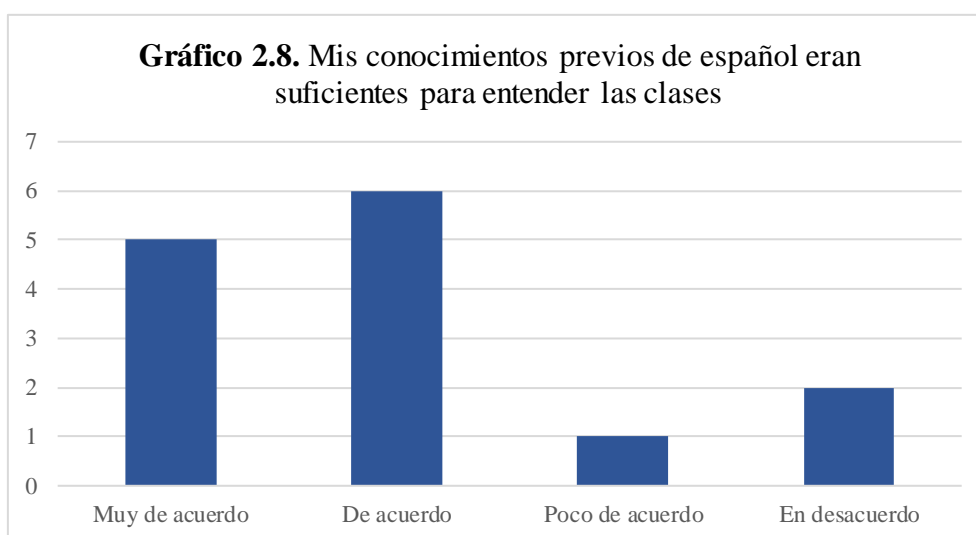
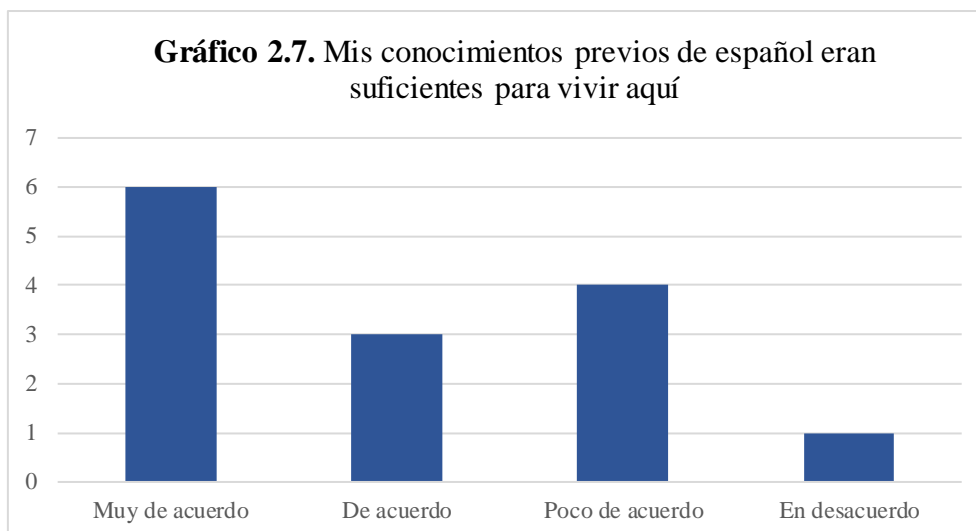


Gráfico 2.6. El mejor español se habla en España





Por lo general, la mayoría está de acuerdo o muy de acuerdo con las afirmaciones de que el español es fácil de aprender (Gráfico 2.4.) y que aprenderlo en España es más fácil (Gráfico 2.5.). Nueve estudiantes consideran que el mejor español se habla en España, mientras que cuatro están poco de acuerdo o en desacuerdo con esa afirmación (Gráfico 2.6.). Por último, les preguntamos también si sus conocimientos previos eran suficientes para vivir en España (Gráfico 2.7.) y para entender las clases (Gráfico 2.8.). Generalmente están muy de acuerdo o de acuerdo con estas dos afirmaciones, aunque cinco estudiantes piensan que sus conocimientos previos no eran suficientes para vivir en Oviedo y tres no los consideraron suficientes para entender las clases.

2.3. La prueba de disponibilidad léxica

2.3.1. Los centros de interés

Para nuestra investigación, hemos elegido examinar tres centros de interés: Partes del cuerpo, Alimentos y bebidas y La ciudad. Estos forman parte de la lista de los centros de interés tradicionales usados en la mayoría de los estudios de disponibilidad léxica, propuestos por primera vez por los estudiosos franceses que crearon el *Français Fondamental* y que se han adoptado también en las investigaciones del *Proyecto Panhispánico* (cf. Bartol Hernández 2006: 380). Los dieciséis centros de interés empleados en estos estudios, así como enunciados en varios trabajos, son:

01. Partes del cuerpo
02. La ropa
03. Partes de la casa (sin muebles)
04. Los muebles de la casa
05. Alimentos y bebidas
06. Objetos colocados en la mesa para la comida
07. La cocina y sus utensilios
08. La escuela: muebles y materiales
09. Calefacción e iluminación
10. La ciudad
11. El campo
12. Medios de transporte
13. Trabajos del campo y del jardín
14. Los animales
15. Juegos y distracciones
16. Profesiones y oficios

A pesar de que los autores del *Français Fondamental* intentaron abarcar los temas más importantes del universo léxico de los hablantes con la finalidad de elaborar un diccionario que pudiera ser empleado en la enseñanza del francés a los extranjeros (cf. Gougenheim *et al.*, 1964: 21-22), en los últimos años muchos estudiosos se están planteando varias preguntas acerca de la funcionalidad de estos centros de interés clásicos. A continuación vamos a analizar algunas de ellas y las soluciones que se han propuesto.

Una de las críticas es que algunos centros de interés, como por ejemplo 09. *Calefacción e iluminación* y 13. *Trabajos del campo y del jardín*, son poco productivos porque suelen ser más alejados de la vida de los informantes y de la realidad en la que viven (Tomé Cornejo, 2015: 196). De hecho, cada vez más los intereses humanos aumentan y se diferencian, eso nos obligaría a modificar los centros de interés que investigamos. López Morales ya hace muchos años se preguntaba “¿con cuántos centros nos obligaría a trabajar una investigación que pretendiera ser exhaustiva?” (1999: 33). Algunos estudiosos han propuesto aumentar los centros de interés añadiendo esos campos del saber o temas más actualizados con la sensibilidad de los informantes, como por ejemplo la *Naturaleza o Medio ambiente*, los *Medios de comunicación social*, las *Nuevas tecnologías*, la *Vida social*, los *Sentimientos y emociones*, etc. (Gómez Devís, 2004: 56), o las *relaciones familiares y no familiares*, la *música*, el *cine*, la *enfermedad*, las *acciones cotidianas*, el *carácter*, etc. (Izquierdo Gil, 2003: 424).

Este problema no se ha discutido solo por las pruebas de disponibilidad léxica con hablantes nativos, sino también con los estudiantes de ELE. Bartol Hernández en 2010 comparó los dieciséis centros de interés clásicos con los temas y subtemas del *Marco Común Europeo de Referencia para las lenguas* (MCER) y reconoció que hay bastantes coincidencias. Pone de relieve que la mayoría de los centros de interés está bien representada en los temas del MCER, sin embargo, otros, como 06. *Objetos colocados en la mesa para la comida*, 09. *Iluminación, calefacción y medios de airear un recinto* y 13. *Trabajos del campo y del jardín* no están incluidos, mientras que 08. *La escuela* no está muy bien representada o, mejor dicho, podría encajar en la Educación que implica un ámbito más amplio. Pese a las similitudes, evidencia la falta de centros de interés “relativos a vida cotidiana, relaciones con otras personas, la salud, el tiempo, las creencias, y otras no están bien representadas, como la educación” (2010: 95). Además, al final del mismo artículo, contesta a la pregunta de López Morales diciendo que intentar encasillar todos los intereses humanos en los centros de interés sería “irrealizable e inviable”, puesto que nos llevaría a trabajar con un número demasiado elevado (p. 104), aunque no descarta la posibilidad de ampliar el número y el tipo de centros de interés.

Con respecto al aumentar el número, esto sería posible siempre y cuando cada informante no tuviera que responder a todos ellos (cf. Tomé Cornejo, 2015: 201). Por ejemplo, la misma autora, en su tesis doctoral utiliza 47 centros de interés repartidos en cuatro cuestionarios de manera que cada estudiante tenía que contestar solo a 11 o 12 de ellos (cf. Tomé Cornejo, 2015).

Refiriéndose al tipo, Bartol Hernández contesta también a una de las otras grandes críticas, o sea que con estos centros de interés se obtienen mayoritariamente sustantivos concretos, en algunos casos verbos, pero muy pocas palabras de otras categorías gramaticales. Él, como muchos otros estudiosos, reconoce la necesidad de modificar algunos enunciados para que nos permitan obtener palabras de todas las categorías. Como se observará sucesivamente, en nuestro estudio no hemos encontrado ningún verbo ni adjetivo, sin embargo, en otros estudios (cf. Del Barrio & Mae Vann, 2018) que han tomado en consideración los dieciséis centros de interés, sí han aparecido adjetivos y verbos, sobre todo en 02. *La ropa*, 05. *Alimentos y bebidas*, 09. *Iluminación y medios de refrigeración y calefacción*, 10. *La ciudad*, 11. *El campo*, 12. *Medios de transporte*, 13. *Trabajos del campo y del jardín* y 15. *Juegos y distracciones*. A pesar de algunas apariciones, en este estudio también la mayoría de las palabras que los informantes han producido han sido sustantivos, por eso los mismos autores del libro se plantean el objetivo de añadir otros centros de interés como los colores, el aspecto físico, el carácter, las actividades habituales, etc. para obtener un vocabulario más variado desde el punto de vista categorial (p. 12).

La voluntad de añadir nuevos centros de interés responde a la necesidad de captar el léxico que los dieciséis tradicionales no alcanzan. Como expone Tomé Cornejo (2015: 220), en líneas generales, los centros de interés novedosos que se han empleado en varios estudios quieren obtener:

- el léxico disponible de ámbitos no investigados, como la política, el arte, la ciencia, los colores, la economía, etc.
- el léxico de los intereses humanos más recientes como las nuevas tecnologías o el terrorismo,
- las palabras de categoría gramatical diferente, considerando por ejemplo las acciones, sensaciones y sentimientos, aspecto físico y personalidad,
- el léxico que refleja particularidades geográficas o sociales de la población estudiada (la mar, Portugal, etc.)
- el vocabulario especializado de algunos sectores sociales²⁵.

²⁵ Para consultar una lista de los centros de interés novedosos empleados en varios trabajos véase Tomé Cornejo (2015) pp. 220-228.

Además de añadir centros de interés nuevos, en ocasiones también se han modificado los enunciados. Estas reformulaciones pueden no tener grandes consecuencias en los resultados, como por ejemplo proponer *Juegos y diversiones* en lugar de *Juegos y distracciones*, o añadir nuevas referencias temáticas, omitir una parte del enunciado, invertir el orden de los sintagmas que lo componen o cambiar el nivel de inclusividad del mismo²⁶ (cf. Tomé Cornejo, 2015: 210).

La formulación de los enunciados de los centros de interés es otro aspecto criticado puesto que algunos, como 06. *Objetos colocados en la mesa para la comida*, son muy restrictos y tienen un vínculo fuerte con las palabras que requieren, mientras que otros implican relaciones heterogéneas, de carácter asociativo, como por ejemplo 10. *La ciudad* o 11. *El campo* (cf. Tomé Cornejo, 2015: 197). Esta heterogeneidad ha llevado, en algunos casos, a percepciones muy distinta entre los encuestados: por ejemplo, en su estudio conducido en 2012, Rebeca Delgado Fernández encuentra que sus informantes tienden a interpretar el centro de interés *La ciudad* como *La ciudad de Burgos*, aportando nombres propios de calles, barrios y monumentos de la misma (cf. Delgado Fernández, 2012: 21 *apud* Tomé Cornejo, 2015: 197). Por lo general, la comunidad científica se divide en dos posiciones: los estudiosos que consideran los estímulos demasiado amplios, y que preferirían especificarlos más (cf. Izquierdo Gil, 2003) y los que sostienen que algunos son demasiado concretos, y que proponen trabajar con categoría más abiertas como Victory (1971).

Otro aspecto relacionado con la amplitud, es que algunos centros de interés hacen referencia a una sola categoría semántica, aunque vasta, como por ejemplo 01. *Partes del cuerpo*, mientras que otros se refieren a más de una categoría, como 09. *Calefacción e iluminación*. Esto puede complicar la evocación de palabras puesto que el límite de tiempo de dos minutos que se suele dejar por cada centro no varía según su amplitud. Para evitar este problema, algunos autores, como Izquierdo Gil (2003) han decidido dividir los centros más amplios en campos semánticos más pequeños: por ejemplo 05. *Alimentos y bebidas* se ha dividido en “los alimentos”, “las bebidas” y “la fruta” (p. 425).

Vinculado con la amplitud y la formulación de los enunciados, cabe destacar otros dos aspectos: las interrelaciones entre centros de interés y el efecto *priming*. Ya Mackey en 1971 decía que no le iba a sorprender encontrar palabras duplicadas en más de un centro de interés,

²⁶ Para consultar los principales cambios que se han introducidos en los enunciados de los centros de interés tradicionales, véase Tomé Cornejo (2015) pp. 210-219.

como por ejemplo en lo de la ciudad y del campo, porque el universo en el que vivimos no está organizado en compartimentos estancos. En nuestro caso, al analizar tres centros que se refieren a tres realidades muy distintas entre sí, no hemos encontrado interrelaciones, sin embargo, hemos encontrado palabras en 10. *La ciudad* que podrían más bien encajar en 12. *Medios de transporte*, como *coche*, *autobús* y *taxi*. En otros estudios que han considerado todos los centros de interés se han obtenido palabras compartidas en más de un centro: por ejemplo *árbol*, *casa*, *hierba*, *jardín*, etc. tanto en 10. *La ciudad* como en 11. *El campo* (cf. Del Barrio & Mae Vann, 2018). Como nos acuerda Tomé Cornejo (2003), esta repetición “es consecuencia de la interrelación que existe entre las categorías semánticas: una palabra pertenece a distintos campos semánticos –ya sea con el mismo significado o con otro, en el caso de las palabras polisémicas–, y en cada uno presenta una disponibilidad diferente” (p. 202).

En cuanto al efecto *priming*, este implica que determinados estímulos, denominados *primes*, activen algún tipo de información (bien sea semántica, fonológica, ortográfica, etc.) que promueve o inhibe la activación, selección o producción de otras lexías (cf. Álvarez, Alameda, Domínguez, 1999: 103). Algunos estudiosos han puesto en evidencia que enunciados formados por dos o más sintagmas coordinados, como 05. *Alimentos y bebidas*, suelen favorecer este efecto, así que los informantes responden primero al último campo que escuchan o leen, puesto que es el que permanece en su memoria²⁷. Además, también se han detectado ejemplos del efecto *priming* entre centros de interés o al interior de uno en concreto. Por lo que se refiere al primer caso, está demostrado que el orden en el que se presentan los estímulos puede favorecer la repetición de términos en centros de interés próximos y de temática similar (cf. Tomé Cornejo, 2015: 190). También se dan ejemplos de *priming* adentro de un centro de interés: Hernández Muñoz y Tomé Cornejo, analizando 14. *Los animales*, explican que se pueden producir asociaciones tanto semánticas (*gato-ratón*) como fonológicas (*gato-pato*) (2017: 105).

Por último, aplicando los estudios de disponibilidad léxica a los estudiantes de ELE, se ha evidenciado como algunos centros de interés, poco productivos con los hablantes nativos, resultan igualmente poco productivos también con los extranjeros. De ahí que varios estudiosos decidieron modificarlos para adaptarlos a sus informantes, reformulando el enunciado, juntando varios campos en uno o eliminando los menos productivos. Por ejemplo, Sánchez-Saus Laserna (2016) une 03. *Partes de la casa (sin muebles)* y 04. *Los muebles de la casa* en

²⁷ En nuestro caso en el centro 05. *Alimentos y bebidas*, cinco informantes sobre catorce han sido sujetos al efecto *priming*. De hecho, las primeras dos o tres palabras que produjeron han sido bebidas, seguidas por los alimentos.

el centro *La casa* puesto que se pueden establecer vínculos semánticos cercanos entre los dos. En cambio, Rubio Sánchez (2017) decidió eliminar completamente los centros menos productivos y añadir otros más novedosos, como Acciones y actividades habituales, Aspecto físico y carácter para perseguir palabras de diferentes categorías gramaticales, La lengua española, España y los españoles, Italia y los italianos para examinar las opiniones y concepciones acerca de la cultura española e italiana, y Partes del cuerpo (en italiano), Comidas y bebidas (en italiano) para confrontar el número de vocablos que saben en español y en su lengua materna, el italiano (p. 145). Muchos otros estudios introducen nuevos centros de interés²⁸, algunos de los cuales los sacan de los planes curriculares de enseñanza de ELE, como por ejemplo Jiménez Berrio (2013) que añade La familia o Sánchez-Saus Laserna (2016) que introduce El dinero.

Como se ha brevemente expuesto en esta sección, varios autores han encontrado algunas criticidades en la formulación y adecuación de los dieciséis centros de interés tradicionales, sin embargo, seguir utilizándolos tiene la ventaja de permitir comparar los nuevos trabajos con los anteriores. En todo caso, a la comunidad científica parece evidente la necesidad de investigar en nuevos centros más actuales y productivos.

2.3.2. *La edición de los datos*

Después de haber seleccionado los centros de interés y diseñado y aplicado la encuesta se ha procedido a la edición de los datos obtenidos. Esta labor permite unificar las variantes dadas por los informantes bajo una única forma para el sucesivo análisis. Los criterios de edición que hemos aplicado fueron propuestos por José Antonio Samper Padilla en 1998 y son los mismos que han servido como base en las investigaciones del *Proyecto Panhispánico*. En nuestro caso hemos adaptado algunos de los 8 expuestos.

Durante el trabajo de edición de los datos se nos presentó una cuestión metodológica acerca del alcance de las relaciones que podíamos aceptar y las que teníamos que descartar. Por lo general hemos elegido mantenernos lo más fieles posible a las palabras producidas por los informantes, de ahí que hayamos aceptado tanto las palabras que se relacionan estrictamente con el centro de interés propuesto, como las que tienen una relación de tipo asociativo.

Como fuente para resolver nuestras dudas hemos tomando como referencia el *Diccionario de lengua española* de la Real Academia Española (de aquí en adelante *DLE*) en

²⁸ Para consultar una lista de los centros de interés novedosos empleados en varios trabajos con estudiantes de ELE véase Tome Cornejo (2015) pp. 239-240.

la versión electrónica de la 23ª edición. Antes de analizar los criterios de edición distinguiendo por cada centro de interés objeto de esta investigación, vamos a exponer los criterios comunes a los tres.

1. Eliminación de los términos repetidos. En una sola ocasión hemos encontrado una repetición (*pies - pies*) y hemos descartado la palabra que se encontraba en la posición más baja de la lista.
2. Corrección de la ortografía. Hemos encontrado varios errores ortográficos que hemos corregido y aceptado siempre y cuando se podía entender sin dificultad la palabra a la que el informante hacía referencia. Lo que es evidente es que muchas veces los fallos ortográficos son debidos a un desconocimiento del español o a errores habituales en los estudiantes extranjeros, como la colocación de la tilde o el empleo equivocado de *h, s* y *c/z* o *g* y *j*. Es oportuno señalar que, al explicar las instrucciones del test, se han informado a los participantes que la ortografía no iba a tener importancia en el análisis porque sus respuestas habrían sido uniformadas y corregidas. Con esto intentamos prevenir que los participantes evitaran escribir una palabra sólo para no incurrir en un error²⁹.

La falta más frecuente fue el posicionamiento equivocado de la tilde (*bibliotécas, catedral, taxi*) o su omisión (*platano, azucar, jamon, meloconton, lemon, estomago, calabacin, higado, pulmon, riñon, cartilagos, musculos, torax, gluteo, corazon, semaforo, automovil, peluqueria, arbol, historico, autonomas, tipica, sidreria, habitacion, pescaderia, carniceria, fruteria, cafeteria, frio*).

También hubo varios problemas con la *h*, que se ha repuesto cuando absente: *zanaoria > zanahoria, cipirones > chipirones, alcolica > alcohólica*, quitado cuando insertada erróneamente: *thé > té, cathedral > catedral* o recolocado, cuando estaba puesta en el sitio incorrecto: *zhanaoria > zanahoria, albacha > albahaca*.

Otro error frecuente ha sido la confusión entre los grafemas ⟨*s*⟩, ⟨*c*⟩ y ⟨*z*⟩. En concreto en el uso de ⟨*s*⟩ en lugar de ⟨*z*⟩ (*sanahorias > zanahorias, sumo > zumo*), de ⟨*c*⟩ en lugar de ⟨*s*⟩ (*cidra > sidra*), de ⟨*c*⟩ en lugar de ⟨*z*⟩ (*bracos > brazos, cerveca > cerveza*), de ⟨*z*⟩ en lugar de ⟨*s*⟩ (*piez > pies*) y de ⟨*z*⟩ en lugar de ⟨*c*⟩ (*zebra >*

²⁹ Aunque especificar a los participantes esta intervención de corrección ortográfica es necesaria para que escriban también las palabras sobre las cuales tienen alguna duda, si es verdad que de esta manera se les desalientan a que escriban correctamente. Esto es evidente sobre todo en los participantes que han declarado un conocimiento avanzado de lengua y que, por ejemplo, no ponen la tilde en palabras de nivel básico.

cebra). Se registra también un caso de uso del grafema ⟨ç⟩ ajeno al español (*coraçón* > *corazón*) probablemente debido a una baja competencia ortográfica del informante, y un caso de uso de ⟨f⟩ por ⟨c⟩ (*morfilla* > *morcilla*).

Se han encontrado también errores en la escritura de ⟨g⟩ que se ha producido como ⟨j⟩: *barrija* > *barriga*, *colejio* > *colegio*, *jemelos* > *gemelos*, o de ⟨j⟩ como ⟨y⟩: *oreyas* > *orejas*. El mismo grafema se ha erróneamente escrito también con el dígrafo -hi- (*ohio* > *ojo*) o con -n- (*menilla* > *mejilla*).

Han ocurrido confusiones entre ⟨d⟩ y ⟨t⟩: *detos* > *dedos*, *catetral* > *catedral*. Se registran ejemplos de reduplicación de ⟨l⟩: *collegio* > *colegio*, *tallón* > *talón*, probablemente debidos a la influencia de las palabras italianas *collegio* y *tallone*. Hemos encontrado también un error en la escritura de ⟨r⟩ producida como ⟨l⟩: *celebro* > *cerebro*.

Otros errores esporádicos que hemos encontrado han sido la apocope (supresión de un sonido al final de palabra): *abdome* > *abdomen* o la sincopa (supresión de un sonido al interior de palabra): *ombigo* > *ombligo*, dificultades en la escritura del dígrafo para representar la velar oclusiva sorda /k/: *cheso* > *queso*, o en la escritura de algunas palabras: *cedro* > *cerdo*, *ajuntamiento* > *ayuntamiento*, *districtos* > *distritos*.

Por lo que se refiere a las vocales, se han corregido los errores debido a la confusión de átonas (*espinaces* > *espinacas*, *eglesia* > *iglesia*, *lemon* > *limón*) y tónicas (*ciudadero*, *ciudadino* > *ciudadano*). Se ha encontrado un caso de diptongación equivocada, probablemente por influencia del italiano: *cualidad* > *calidad* (it. *qualità*), un caso de apocope (*vin* > *vino*) y un caso de escritura equivocada de la vocal i (*gymnasio* > *gimnasio*).

3. Edición de las variantes flexivas. Dado que elegimos mantenernos lo más fieles posible a la forma producida por los participantes, decidimos dejar las palabras que aparecen tanto al singular como al plural para examinar la disponibilidad de cada una de las formas. Por eso aparecerán tanto *naranja* como *naranjas*, *ojo* y *ojos*, *calle* y *calles*, etc. Solo se registra un caso de palabra repetida tanto al singular como al plural por el mismo participante (*pelo* – *pelos*), en cambio, no se registran sustantivos con distinción entre femenino y masculino.

4. Edición de derivados regulares. Registramos solo un diminutivo (*quesito*) que hemos mantenido, dado que tiene una voz propia en el *DLE*, y no aparece vinculado como diminutivo a la palabra *queso*.
5. Acortamientos y grupos sintagmáticos. Sólo hemos encontrado dos acortamientos: *bus* por *autobús* y *cine* por *cinematógrafo*. En los dos casos decidimos no uniformarlos a la forma plena, para poder comprobar su índice de disponibilidad frente a esta y porque, en el segundo caso, solo aparece el acortamiento. Por lo que se refiere a los grupos sintagmáticos, al contrario de lo que se ha hecho en otras investigaciones, elegimos mantener todas las especificaciones sin poner entre paréntesis el elemento que se podría elidir: por eso aparecerán *zum* y *zum* de *naranja*, *vino* y *vino* *tinto*, de la misma manera, se mantienen *bebidas* *alcohólica*, *bebidas* *energizante*, *bebidas* *nutrientes*, *bebidas* *sin azúcar*, *calidad* *de vida*, *calidad* *del aire*, *casco* *antiguo*, *casco* *histórico*, *centro* *comercial*, *centro* *de ocio*, *centro* *deportivo*, *centro* *salud*, *ciudad* *rural*, *columna* *vertebral*, *comida* *típica*, *comunidades* *autónomas*, *globo* *ocular*, *manera* *de moverse*, *palma* *de la mano*, *platos* *compuestos* dado que tienen una unicidad referencial, es decir, funcionan como unidades simples.
6. Marcas comerciales. Solo hemos encontrado tres marcas comerciales (*Coca-Cola*, *San Martin* y *Uber*) que se han mantenido editadas en mayúsculas.

Después de haber expuesto los criterios de edición generales, usados por los datos de los tres centros de interés, pasamos ahora al análisis detallado de cada uno de los centros, evidenciando sus particularidades.

01. Partes del cuerpo

Este centro de interés no es muy problemático, hubo varios errores ortográficos que ya comentamos anteriormente. Todas las palabras producidas son sustantivos y destaca que por lo general abundan los plurales (*ojos*, *piernas*, *pies*) frente a las formas singulares. Aparecen algunas interferencias con el italiano, como *bulbo ocular*, *ditas*, *pulso* y *rotulas* que hemos editado en *globo ocular*, *dedos*, *muñeca* y *rodillas*.

05. Alimentos y bebidas

La mayoría de las expresiones de este centro se refiere a alimentos, tipos de bebidas y comidas. No hemos encontrado adjetivos ni verbos, de hecho, todas las entradas son

sustantivos. Hemos eliminado las palabras *vaso* y *tabla* porque lejanas al tema propuesto, así como dos nombres propios de dos sidrerías de Oviedo.

Se han “castellanizado” algunas palabras que aparecían influenciadas por el italiano (*lasagna* > *lasaña*, *pez* > *pescado*, *salsichas* > *salchicas*, *vegetables* > *verduras*), mientras se ha quitado la palabra *burro* (que corresponde a ‘mantequilla’ en español) porque su traducción ya aparecía en una posición más alta de la lista. Hemos también encontrado, y aceptado los extranjerismos *vodka*, *pasta* y *pizza* puesto que están reconocidos en el *DLE*, aunque este último aparece como un préstamo italiano.

Aceptamos nombres de platos y bebidas como: *arroz con leche*, *cachopo*, *calimoch*, *chorizo*, *churros*, *fabada*, *paella*, *pincho*, *pollo asado*, *tortilla*, *sangría*, *sidra*, *tarta de queso*; así como las especificaciones (*vino tinto*, *zum* *de naranja*); hemos aceptado también los dialectismos como *frijoles* y *papas*.

10. La ciudad

Este centro de interés es el que más se presta, entre los que analizamos, a asociaciones no estrictamente cercanas al tema propuesto. Después de haber eliminado los nombres de las ciudades, aunque correctamente escritos (*Gijón*, *Madrid*, *Oles*, *Oviedo*, *Valencia*, *Valladolid*, *Villagraciosa*), y la palabra inglesa *avenue*, hemos procedido al análisis.

Es posible individuar varios subgrupos: la división de la ciudad (*casco antiguo*, *casco histórico*, *distritos*), los edificios (*casas*, *catedral*, *edificio*, *habitación*, *iglesia*, *piso*), los servicios y tiendas (*almacén*, *ayuntamiento*, *carnicería*, *centro comercial*, *centros de salud*, *esteticista*, *farmacias*, *frutería*, *hospitales*, *librerías*, *mercado*, *negocios*, *oficinas*, *palacios*, *peluquería*, *pescadería*, *restaurante*, *sidrería*, *supermercado*, *talleres*, *tienda*, ...), algunos de los cuales también están relacionados con la educación (*bibliotecas*, *campus*, *colegio*, *escuela*, *universidad*) o son espacios de ocio o deporte (*bar*, *cafeterías*, *centro deportivo*, *centros de ocio*, *cine*, *discotecas*, *fiestas*, *gimnasio*, *pub*, *teatro*, ...). Se puede individuar el subcampo de los medios de transporte y sus espacios (*aeropuerto*, *andén*, *aparcamiento*, *autobús*, *automóvil*, *carro*, *coche*, *estación de trenes*, *manera de moverse*, *parking*, *puerto*, *taxi*, *Uber*), los espacios de la ciudad (*acera*, *calle*, *camino*, *carretera*, *esquina*, *paso de cebra*, *plaza*, *rotonda*), varios elementos como *farola*, *luc* *es*, *semáforo*, y las personas que viven en la ciudad (*ciudadanos*, *gente*, *personas*, *población*).

Por otra parte, aparecen también palabras conectadas con el opuesto de la ciudad, la naturaleza (*árbol, flores, jardín, parque, prados*). Además, hemos encontrado palabras que encajan de manera más amplia con el campo de la ciudad como *comida típica, frío, lluvia y ruido*, y también *arte, capital, ciudad rural, compras, comunidades autónomas, calidad de vida, calidad del aire, economía, historia y ventanas*.

Se han mantenido tres nombres que provienen del inglés (*parking, pub y taxi*) porque están reconocidos en el *DLE*, y también el italianismo *negocios* con el significado de “local en que se negocia o comercia” y se ha traducido *fontana* en *fuelle*. En cambio, hemos elegido dejar la palabra *Fontán* así como una participante la escribió porque probablemente se estaba refiriendo a la plaza del *Fontán* (conocida también como *El Fontán*) en Oviedo.

CAPÍTULO III

ANÁLISIS CUANTITATIVO DE LOS RESULTADOS

En este capítulo presentamos el análisis cuantitativo de los resultados obtenidos con las encuestas de disponibilidad. Primero nos centramos en los resultados generales considerando el número total de unidades léxicas producidas por los informantes, marcando la diferencia entre palabras y vocablos, presentamos también los promedios de respuestas por persona, la densidad y el índice de cohesión de cada uno de los centros de interés investigados. Sucesivamente analizamos estos datos según las cuatro variables que consideramos en este estudio: sexo, carrera, nivel de español y conocimiento de otras lenguas. Al final presentamos un resumen de las conclusiones que hemos podido derivar a partir del análisis de los resultados obtenidos.

3.1. Resultados generales

Después del proceso de edición de los datos hemos obtenido un corpus de 686 palabras y 341 vocablos totales repartidos en los tres centros de interés. En los estudios de disponibilidad léxica es común distinguir entre palabras (o *tokens*) y vocablos (o *types*): las palabras indican todas las unidades que han aparecido en el corpus, incluidas las repeticiones, en cambio, los vocablos son las unidades diferentes que se han encontrado dentro de la totalidad de las palabras obtenidas con las encuestas.

El total de palabras y de vocablos están presentados en la segunda y tercera columna de la tabla 3.1. En la cuarta y quinta columna se presentan los promedios generales por informante.

Informantes	Palabras	Vocablos	Promedio palabras	Promedio vocablos
14	686	341	49	24,35

Tabla 3.1. Resultados generales

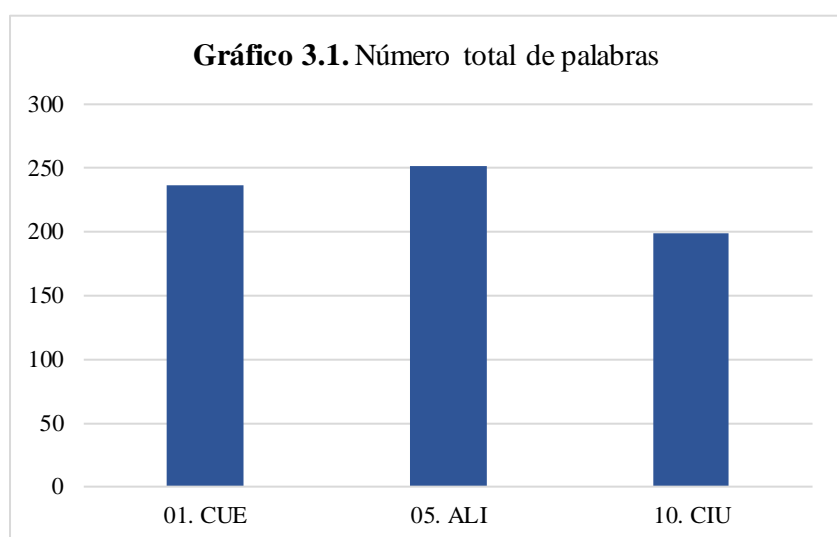
3.1.1. Número de palabras

En la tabla 3.2. mostramos el total de palabras producidas por los estudiantes en cada centro de interés, así como el promedio de palabras por informante en cada uno de ellos. En la primera columna, indicamos el rango, es decir, el orden de los centros en función de su mayor o menor número de palabras.

Rango	Centro de interés	Palabras	Promedio por informante
2	01. Partes del cuerpo	236	16,86
1	05. Alimentos y bebidas	251	17,93
3	10. La ciudad	199	14,21
TOTAL		686	49

Tabla 3.2. Número de palabras y media por informante por centro de interés

En el gráfico 3.1.³⁰ mostramos la productividad de cada centro de interés. Es evidente que la distribución de palabras en los tres centros es bastante uniforme, aunque dos superan la media de palabras producidas por cada centro (228,6) mientras que uno, 10. La ciudad, queda debajo de este valor.



El centro de interés que resulta ser el más productivo es 05. Alimentos y bebidas (251 palabras; 17,93 de media por informante), seguido por 01. Partes del cuerpo (236 palabras; 16,86 de media por informante). 10. La ciudad es el centro menos productivo con una diferencia de más de 50 palabras con respecto al primero (199 palabras; 14,21 media por informante).

Analizando las medias, podemos ver como la diferencia en la productividad por informante es varia: en el centro más productivo, 05. Alimentos y bebidas, los informantes escriben mediamente 17,93 palabras cada uno, en el segundo escriben casi una palabra menos por persona (16,86 palabras) y finalmente el centro menos productivo, 10. La ciudad, tiene una

³⁰ En los siguientes gráficos y tablas se hará referencia a los centros de interés con las abreviaturas: 01. CUE: Partes del cuerpo, 05 ALI: Alimentos y bebidas, 10. CIU: La ciudad.

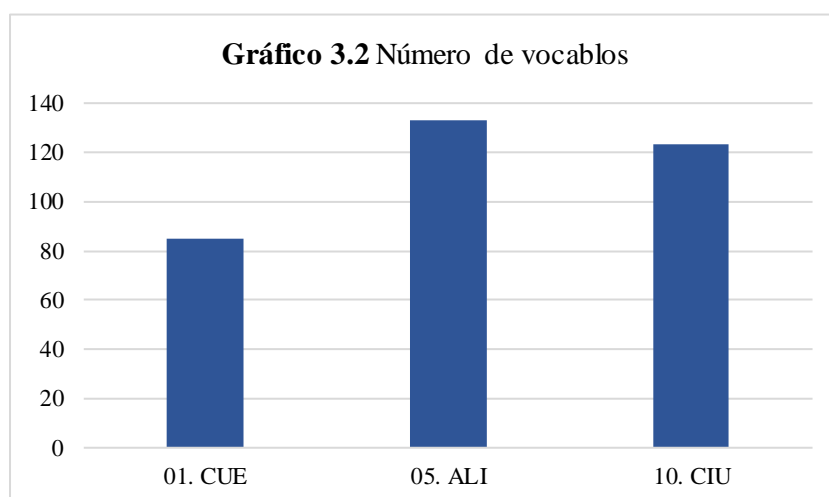
media de 14,21 palabras por persona, es decir, más de dos palabras menos por persona con respecto al centro de rango 2.

3.1.2. Número de vocablos

Pasando al análisis la variedad léxica encontramos 341 vocablos distribuidos en los tres centros de interés. El total de vocablos es casi la mitad en comparación con las palabras, de ahí que también el promedio total por informante sea inferior.

Rango	Centro de interés	Vocablos	Promedio por informante
3	01. Partes del cuerpo	85	6,07
1	05. Alimentos y bebidas	133	9,5
2	10. La ciudad	123	8,78
TOTAL		341	24,35

Tabla 3.3. Número de vocablos y media por informante por centro de interés



Analizando el total de vocablos el centro más rico es, como ocurría por las palabras, 05. Alimentos y bebidas con 133 vocablos, seguido por 10. La ciudad con 123 vocablos, que es el menos productivo. El centro menos rico es 01. Partes del cuerpo con solo 85 vocablos.

Por lo que se refiere a la media de vocablos producidos por cada informante, los dos centros más ricos, 05. Alimentos y bebidas y 10. La ciudad, presentan un número de vocablos superior a la media de 8,11, respectivamente 9,5 y 8,78 unidades por persona. El centro menos rico, 01. Partes del cuerpo, queda debajo de este valor con un promedio de 6,07 vocablos.

Comparando palabras y vocablos producidos por cada centro de interés, podemos afirmar que no hay relación entre productividad y riqueza léxica puesto que el rango, u orden, de los dos difiere. De hecho, se reafirma el centro 05. Alimentos y bebidas como el más

productivo y el más rico, en cambio, en el caso de los otros dos centros, su orden se invierte: 01. Partes del cuerpo resulta ser el segundo centro más productivo, pero es el menos rico, mientras que 10. La ciudad es el menos productivo, pero es el segundo en variedad léxica.

3.1.3. Densidad léxica e índice de cohesión

Nos centramos ahora en el análisis de la densidad léxica de los tres centros de interés y en su índice de cohesión. Estos dos valores tienen que ver con las unidades producidas por los informantes, es decir, las unidades pueden ser actualizadas por muchos informantes o sólo por un número limitado de ellos.

La densidad léxica, o media de repeticiones de las palabras, se calcula dividiendo el número de palabras totales de cada centro de interés por el número de vocablos del mismo. A una mayor densidad corresponde una media mayor, y entonces una mayor homogeneidad. El índice de cohesión, como ya mencionamos al presentar los índices empleados en los estudios de disponibilidad léxica (Cf. apartado 1.2.1.), se calcula dividiendo la media de palabras por cada informante por el número de vocablos que el mismo centro de interés incluye. En una escala de 0 a 1, cuanto más el valor se acerca a 1, más el centro se puede considerar cerrado o compacto puesto que las respuestas de los informantes coinciden en gran medida; por el contrario, si el valor está cerca del 0, el centro se considera abierto puesto que las unidades aportadas difieren más.

Como afirma Sánchez-Saus Laserna, estos dos índices nos dan informaciones sobre la constitución interna de los centros de interés y eso tendrá influencia en el análisis cualitativo (2011: 263). De hecho, un índice de cohesión bajo implica la aparición de un número más variado de asociaciones al interno de un centro de interés.

Centro de interés	Rango	Densidad léxica	Índice de cohesión
01. Partes del cuerpo	1	2,77	0,20
05. Alimentos y bebidas	2	1,88	0,13
10. La ciudad	3	1,61	0,12

Tabla 3.4. Densidad léxica e índice de cohesión por centro de interés

Como se puede ver en la tabla 3.4., el centro más compacto, cuyas unidades son más unificadas, es 01. Partes del cuerpo (0,20), seguido por 05. Alimentos y bebidas (0,13), mientras que el centro más abierto es 10. La ciudad (0,12). El mismo orden, o rango, se da también al evaluar la densidad léxica.

CI	Palabras	Rango	Vocablos	Rango	Densidad léxica	Índice de cohesión	Rango
01. CUE	236	2	85	3	2,77	0,20	1
05. ALI	251	1	133	1	1,88	0,13	2
10. CIU	199	3	123	2	1,61	0,12	3

Tabla 3.5. Número palabras, número vocablos, densidad léxica e índice de cohesión y sus rangos

Es interesante ver cómo, a partir del análisis de las palabras, de los vocablos, o del índice de cohesión, el orden o rango de los centros de interés cambia. El centro 01. Partes del cuerpo es el segundo más productivo y es el más compacto ya que su densidad e índice de cohesión son notablemente superiores a los demás (rango 1), pero es el menos rico (rango 3). El centro 05. Alimentos y bebidas es al mismo tiempo el más productivo y el más rico, pero no es el más compacto (rango 2). Por último, el centro 10. La ciudad es el menos productivo (rango 3) y el menos compacto (rango 3), pero es el segundo centro más rico.

3.2. Resultados por variable

3.2.1. Variable Sexo

A diferencia de lo que ocurre en la mayoría de los otros estudios de disponibilidad léxica, que cuentan con más mujeres que hombres³¹, en nuestro caso tenemos un grupo de informantes homogéneo compuesto por 7 mujeres y 7 hombres.

Informantes	Palabras	Vocablos	Promedio palabras	Promedio vocablos
Mujer (7)	414	260	59,14	37,14
Hombre (7)	272	175	38,85	25

Tabla 3.6. Índices generales según la variable Sexo

Las mujeres son las que actualizan una cantidad mayor tanto de palabras como de vocablos. En concreto las informantes escribieron un 34,29% de palabras más que los hombres, y un 32,69% de vocablos más.

³¹ Muchos estudios tienen esta distribución, como por ejemplo: Samper Hernández (2002), Sánchez-Saus Laserna (2011), Rubio (2017), Del Barrio & Mae Vann (2017), Nalesso (2019).

Mujer						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	158	1	22,57	69	3	9,86
05. ALI	136	2	19,43	96	1	13,71
10. CIU	120	3	17,14	95	2	13,57

Tabla 3.7. Índices según la variante Mujer

El centro de interés más productivo de las mujeres es 01. Partes del cuerpo, en el cual realizan 158 palabras. 05. Alimentos y bebidas está en rango 2, con 136 palabras, es decir 22 palabras menos que el primero. El centro menos productivo es 10. La ciudad con 120 palabras, 16 menos que el de rango 2. Los vocablos tienen una distribución distinta: el centro 01. Partes del cuerpo es el menos rico, con 69 unidades, el centro intermedio es 10. La ciudad con 95 vocablos y, finalmente el centro más rico es 05. Alimentos y bebidas con 96 unidades.

Hombre						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	78	3	11,14	45	3	6,42
05. ALI	115	1	16,43	77	1	11
10. CIU	79	2	11,29	53	2	7,57

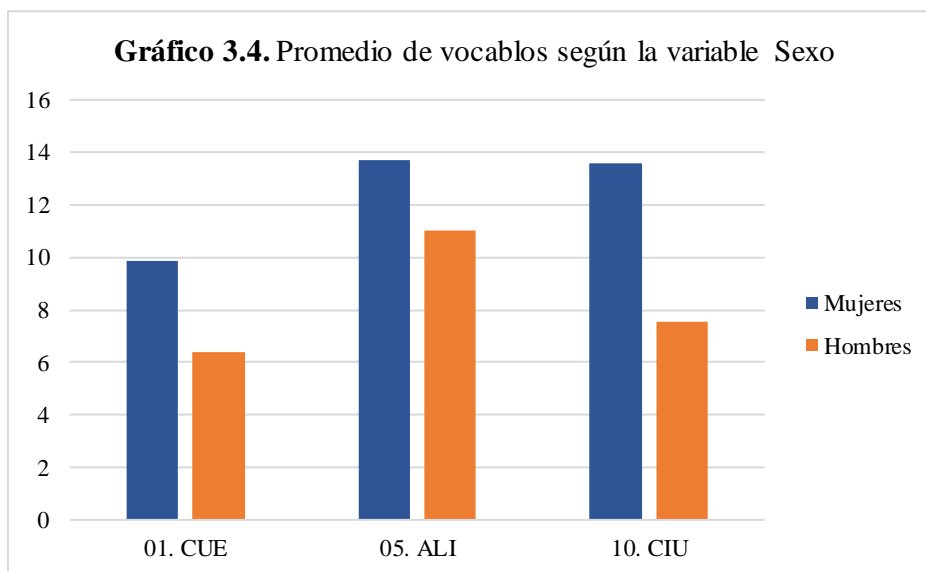
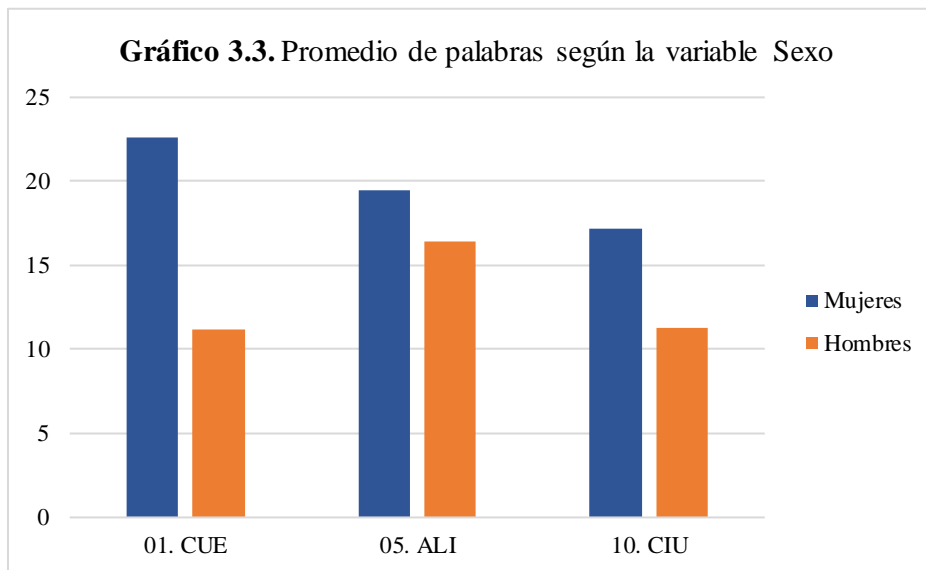
Tabla 3.8. Índices según la variante Hombre

Considerando la variante hombre, el centro menos productivo es 01. Partes del cuerpo, con 78 palabras; poco más productivo es 10. La ciudad con un incremento de una palabra. El centro con el mayor número de unidades es 05. Alimentos y bebidas que supera el de rango 2 de 36 palabras. Analizando la variedad, encontramos la misma ordenación: el centro menos rico es 01. Partes del cuerpo con 45 vocablos, en el centro 10. La ciudad los hombres actualizaron 8 vocablos más y, finalmente el centro más rico es 05. Alimentos y bebidas con 77 unidades, 24 más que el anterior.

A partir de las tablas 3.7. y 3.8. comparamos los índices que acabamos de describir según las dos variantes. Por lo que se refiere a la productividad, el orden de los centros de interés cambia notablemente: el que para las mujeres es el centro más productivo, es decir, 01. Partes del cuerpo, en el caso de los hombres está situado en rango 3. El centro 05. Alimentos y bebidas es el segundo más productivo de las mujeres y el más productivo de los hombres. Mientras que el centro menos productivo de las mujeres, 10. La ciudad, es el segundo más

productivo del grupo opuesto. Ambos grupos se alejan de la ordenación que presentamos anteriormente en la tabla 3.2. que representaba el rango general de las palabras por centro de interés. Además del orden, como analizaremos sucesivamente a través de los promedios, cambia también el número de palabras actualizadas: las mujeres en cada centro de interés son más productivas que los hombres.

Por lo que se refiere a la variedad, los dos grupos coinciden en la ordenación de los centros de interés: el centro más rico, en el cual han producido más vocablos, es 05. Alimentos y bebidas, seguido por 10. La ciudad y finalmente, el menos rico es 01. Partes del cuerpo. En este caso también, como en la productividad, los dos grupos tienen una distinción en los vocablos actualizados que se analizará de manera detallada a través de los gráficos 3.3. y 3.4.



A través de los gráficos 3.3. y 3.4. queremos mostrar visivamente las diferencias de los promedios de las palabras y de los vocablos actualizados por cada uno de los dos grupos en cada centro de interés. Por lo que se refiere al promedio de palabras, es evidente que las mujeres son más productivas, sobre todo en el centro 01. Partes del cuerpo en el cual, mediamente, escriben más del doble de las palabras de los hombres (+ 50,64%). En el caso de los otros dos centros de interés las mujeres siguen siendo más productivas, aunque la diferencia entre los promedios es inferior: en 10. La ciudad (+ 34,13%) y en 05. Alimentos y bebidas (+ 15,44%). Esta diferencia en el promedio de palabras se puede observar también en otros trabajos en los que las mujeres resultan ser más productivas que los hombres en todos los centros de interés investigados, como por ejemplo en Sánchez-Saus Laserna (2016) o Rubio Sánchez (2017). En otros casos se alternan en la mayor productividad de los centros, como en Nalesso (2019).

Analizando el gráfico 3.4., que representa la variedad léxica media, podemos observar que las mujeres actualizan más vocablos que los hombres, de hecho, los tres centros de interés del grupo femenino son más ricos. La diferencia mayor entre los dos grupos se da en el centro 10. La ciudad (+ 44,21%), seguido por 01. Partes del cuerpo (+ 34,88%), y el centro con menor diferencia entre los promedios de vocablos es 05. Alimentos y bebidas (+ 19,76%). Una situación opuesta la obtiene Nalesso (2019), donde el promedio de vocablos por centro de interés de los hombres sobrepasa, y en algunos casos casi triplica, el promedio de las mujeres.

CI	Densidad léxica		Índice de cohesión	
	Mujer	Hombre	Mujer	Hombre
01. CUE	2,29	1,73	0,33	0,25
05. ALI	1,42	1,49	0,20	0,21
10. CIU	1,26	1,49	0,18	0,21

Tabla 3.9. Densidad léxica e índice de cohesión por centro de interés por Sexo

La tabla 3.9. muestra los valores de densidad léxica y de cohesión de los dos grupos por cada centro de interés. Los resultados de las mujeres alcanzan una cohesión más alta en el centro de interés 01. Partes del cuerpo, mientras que los resultados de los hombres son más altos en los otros dos centros. Esto significa que el conjunto de las unidades producidas por el grupo de los hombres es ligeramente más compacto en estos dos centros, mientras que las mujeres tienden a variar más sus respuestas.

3.2.2. Variable Carrera

La segunda variable que vamos a considerar para el análisis de nuestros resultados es la carrera. Como ya explicamos en la descripción de las variables seleccionadas, dividimos nuestros informantes en cuatro grupos: los estudiantes de lenguas extranjeras, de humanidades, de medicina y de física.

Cabe mencionar que todos los estudiantes de lenguas extranjeras han estudiado por dos o más años español como LE, lo mismo pasa con el estudiante de medicina, en cambio, solo dos de los tres estudiantes de humanidades habían estudiado español antes de llegar a Oviedo, y ninguno de los de física tenía conocimientos previos.

Carrera	Palabras	Vocablos	Promedio palabras	Promedio vocablos
Lenguas extranjeras (6)	345	226	57,5	37,6
Humanidades (3)	140	118	46,6	39,3
Medicina (1)	68	68	68	68
Física (4)	134	102	33,5	25,5

Tabla 3.10. Índices generales según la variable Carrera

Dado que hay una considerable diferencia en la constitución de las cuatro agrupaciones, presentamos solamente un análisis del rango de palabras y de vocablos, así como de los promedios, puesto que un análisis de los datos generales nos podría llevar a conclusiones incorrectas. Antes de hacer un análisis comparativo de los promedios de cada grupo, presentamos una breve explicación de los datos para cada uno de ellos.

Lenguas extranjeras						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	133	1	22,17	59	3	9,83
05. ALI	119	2	19,83	89	1	14,83
10. CIU	93	3	15,50	78	2	13

Tabla 3.11. Índices según la variante Lenguas extranjeras

Para el grupo de estudiantes de lenguas extranjeras el centro de interés más productivo es 01. Partes del cuerpo con 133 palabras, seguido por 05. Alimentos y bebidas con 119, mientras que el menos productivo es 10. La ciudad con 93, es decir, 40 menos que el primero. Analizando la variedad léxica el orden cambia, de hecho, el centro más rico es 05. Alimentos

y bebidas con 89 vocablos, el segundo es 10. La ciudad con 78 y el menos rico es 01. Partes del cuerpo con 59.

Humanidades						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	48	1	16	38	2	12,66
05. ALI	48	1	16	40	1	13,33
10. CIU	44	2	14,67	40	1	13,33

Tabla 3.12. Índices según la variante Humanidades

En el caso del grupo de estudiantes de humanidades, nos encontramos con dos centros de interés igualmente productivos: 01. Partes del cuerpo y 05. Alimentos y bebidas con 48 palabras cada uno, mientras que ligeramente menos productivo es 10. La ciudad con 44 unidades. Por lo que se refiere a los vocablos, estos informantes han producido el mismo número de unidades en el centro 05. Alimentos y bebidas y en 10. La ciudad (40), mientras que 01. Partes del cuerpo es menos rico (38).

Medicina						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	29	1	29	29	1	29
05. ALI	24	2	24	24	2	24
10. CIU	14	3	14	14	3	14

Tabla 3.13 Índices según la variante Medicina

Como nos esperábamos, el centro más productivo y rico en el caso del estudiante de medicina es 01. Partes del cuerpo en el que actualiza 29 unidades, seguido por 05. Alimentos y bebidas con 24, mientras que el menos productivo y rico es 10. La ciudad con 14 unidades.

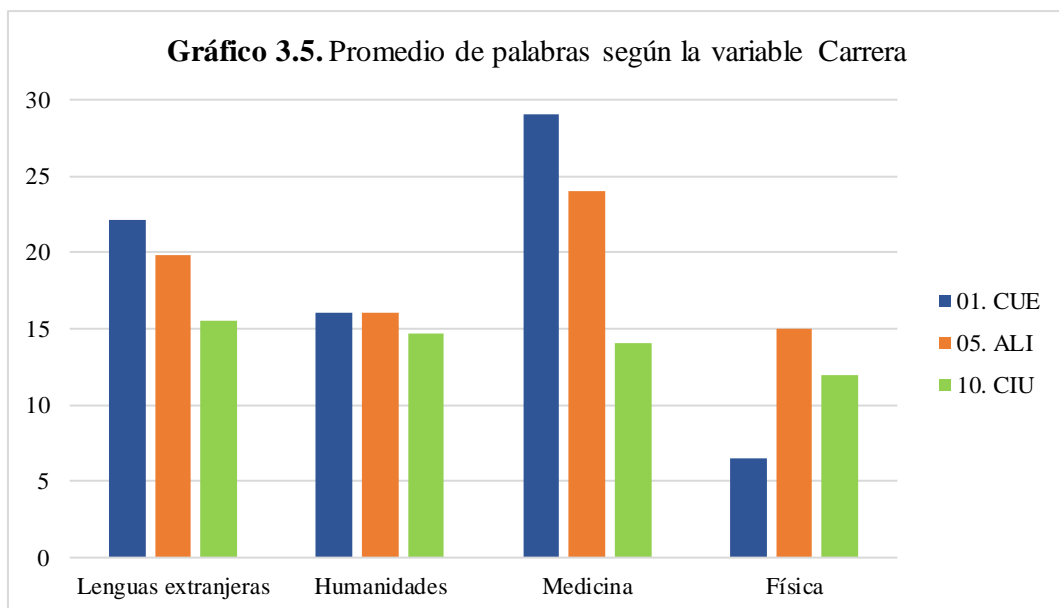
Física						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	26	3	6,5	18	3	4,5
05. ALI	60	1	15	49	1	12,25
10. CIU	48	2	12	35	2	8,75

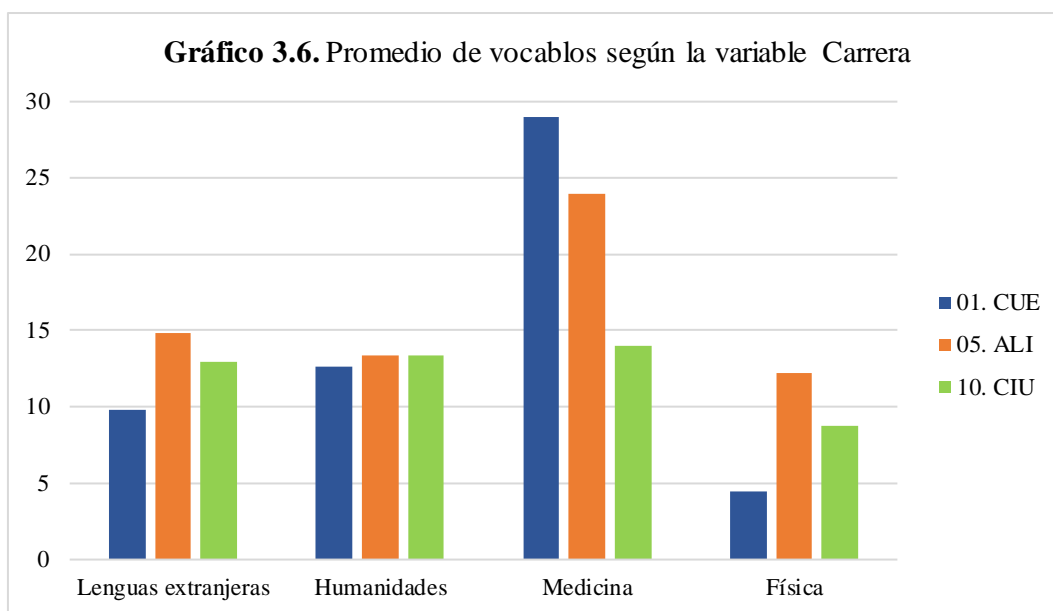
Tabla 3.14. Índices según la variante Física

El grupo de estudiantes de física es más productivo en el centro de interés 05. Alimentos y bebidas en el cual produce 60 palabras, mientras que es menos productivo en 01. Partes del

cuerpo con 26. La variedad léxica sigue el mismo orden: el centro 05. Alimentos y bebidas es el más rico con 49 vocablos, seguido por 10. La ciudad con 35, y por último 01. Partes del cuerpo, en el que realizan menos de la mitad de los vocablos del centro más rico. Probablemente estas diferencias, sobre todo en la variedad, dependen del hecho de que estos informantes nunca estudiaron español como LE, y por eso no entraron en contacto con áreas temáticas que suelen presentarse en los libros y en las clases de ELE como por ejemplo el cuerpo humano. En cambio, los otros dos centros, sobre todo 05. Alimentos y bebidas, resultan ser los más ricos porque tienen que ver con las necesidades cotidianas de cualquier persona.

Por lo que se refiere a los rangos, los estudiantes de lenguas extranjeras, de humanidades y de medicina tienen el mismo orden de la productividad: en los tres casos el centro más productivo es 01. Partes del cuerpo, seguido por 05. Alimentos y bebidas, y finalmente 10. La ciudad. En cambio, el grupo de estudiantes de física ha sido más productivo en 05. Alimentos y bebidas, seguido por 10. La ciudad y han sido menos productivos en 01. Partes del cuerpo. Analizando los rangos de la variedad léxica, la jerarquía cambia: los estudiantes de lenguas extranjeras y los de física producen el mayor número de vocablos en 05. Alimentos y bebidas y el menor en 01. Partes del cuerpo. Los estudiantes de humanidades coinciden en el centro más rico con estos dos grupos, pero su centro menos rico es 10. La ciudad. Como era esperable, el centro en el cual el estudiante de medicina ha escrito más vocablos es 01. Partes del cuerpo, seguido por 05. Alimentos y bebidas y finalmente por 10. La ciudad.





En los gráficos 3.5. y 3.6. se muestran los promedios de palabras y vocablos por informante por centro de interés. El estudiante de medicina resulta ser el más productivo en los centros 01. Partes del cuerpo y 05. Alimentos y bebidas. En general los estudiantes de lenguas extranjeras son los más productivos después de él con respecto a los demás estudiantes; incluso son los más productivos entre todos en el centro 10. La ciudad. Los estudiantes de humanidades son los que mediamente producen casi el mismo número de palabras en los tres centros de interés. Finalmente, los de física son los que tienen los promedios más bajos, probablemente debido a que no tenían conocimientos previos de ELE antes de llegar a España.

Por lo que se refiere a los promedios de variedad léxica, el estudiante de medicina sigue siendo el que produjo más vocablos en el centro 01. Partes del cuerpo (+ 42,2% con respecto a los de humanidades que lo siguen). Los que escribieron menos vocablos en este centro son los estudiantes de física, con una media de 4,5 vocablos por persona. El centro 05. Alimentos y bebidas es por casi todos los grupos el más rico, excepto por el grupo de medicina. Finalmente, el centro 10. La ciudad es el que tiene menos variación, es decir, la diferencia entre el grupo que produjo más vocablos y el que produjo menos es mínima: el estudiante de medicina es el que produjo más unidades (14) y el grupo de física el que produjo menos (8,75 por informante) con una variación de 37,5%.

3.2.3. Variable Nivel de español

La tercera variable que consideramos en nuestro estudio es el nivel de español. A través de las repuestas al cuestionario sociológico hemos dividido nuestros participantes en tres grupos según el nivel con referencia al MCER: nivel A (básico), nivel B (intermedio) y nivel C (avanzado). La división es bastante homogénea: 5 estudiantes son de nivel A, 4 de nivel B y 5 de nivel C.

Antes de analizar los datos, es necesario recordar que al nivel A pertenecen sobre todos estudiantes que no tenían conocimientos previos de español antes de llegar a Oviedo y algún estudiante que había estudiado ELE en sus estudios secundarios. En cambio, a los niveles B y C pertenecen estudiantes que llevan dos o más años estudiando español.

Nivel	Palabras	Vocablos	Promedio palabras	Promedio vocablos
A (5)	191	135	38,2	27
B (4)	223	165	55,75	41,25
C (5)	272	187	54,4	37,4

Tabla 3.15. Índices generales según la variable Nivel de español

Por lo general hay un aumento tanto en las palabras como en los vocablos entre los tres niveles: el nivel inicial (A) es el que produjo menos palabras y vocablos y el nivel avanzado (C) es el que fue más productivo y más rico. Esta distribución cambia al considerar los promedios: el grupo de nivel inicial tiene un promedio de 38,2 palabras por informantes, seguido por el grupo C con 54,4 (+ 29,77 %), y el más productivo es el grupo intermedio con 55,57 palabras por informante (+ 2,42% respecto al anterior). Analizando los promedios de los vocablos los porcentajes cambian ligeramente: el grupo de nivel C produjo 37,4 vocablos, es decir, 27,80% más que el grupo de nivel A (27 unidades por informante), mientras que el grupo B, que resulta ser el más rico, produjo 41,25 vocablos, 9,33% más que el grupo C.

Nivel A						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	53	3	10,6	37	3	7,4
05. ALI	83	1	16,6	58	1	11,6
10. CIU	55	2	11	40	2	8

Tabla 3.16. Índices según la variante Nivel A

En el grupo de estudiantes de nivel básico, el centro menos productivo es 01. Partes del cuerpo, seguido por 10. La ciudad (+ 3,64%), mientras que el centro más productivo es 05. Alimentos y bebidas (+ 33,73%). Analizando la variedad léxica encontramos el mismo orden de rango, aunque con incrementos diferentes: el centro 10. La ciudad es 7,5% más rico que 01. Partes del cuerpo, y el centro más rico, 05. Alimentos y bebidas, es 31,03% más rico que su antecedente.

Nivel B						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	81	1	20,25	43	3	10,75
05. ALI	71	2	17,75	60	2	15
10. CIU	71	2	17,75	62	1	15,5

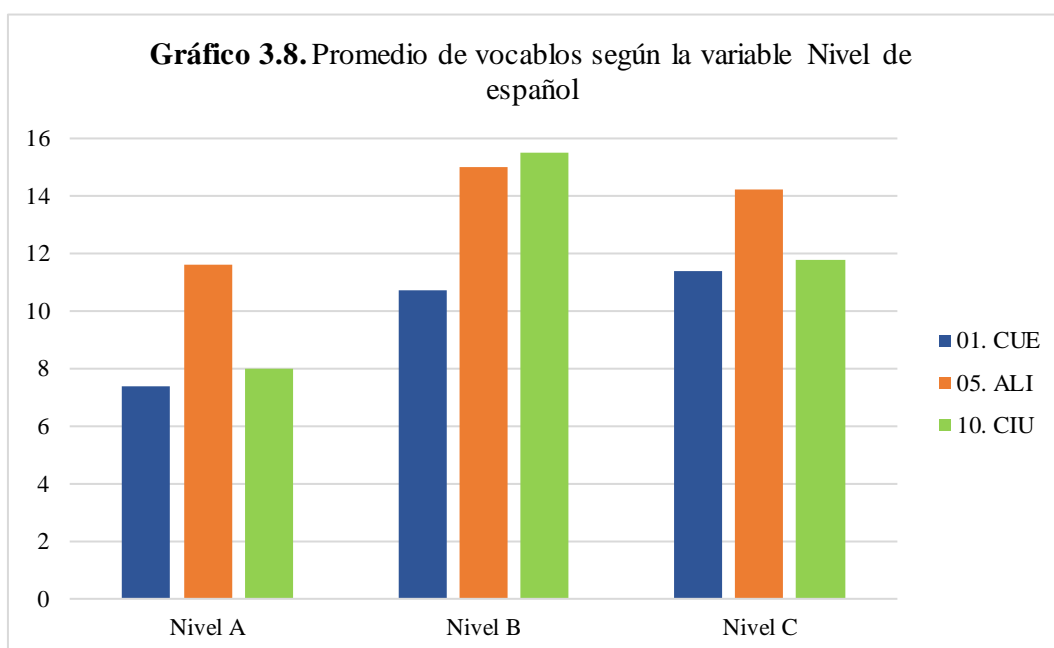
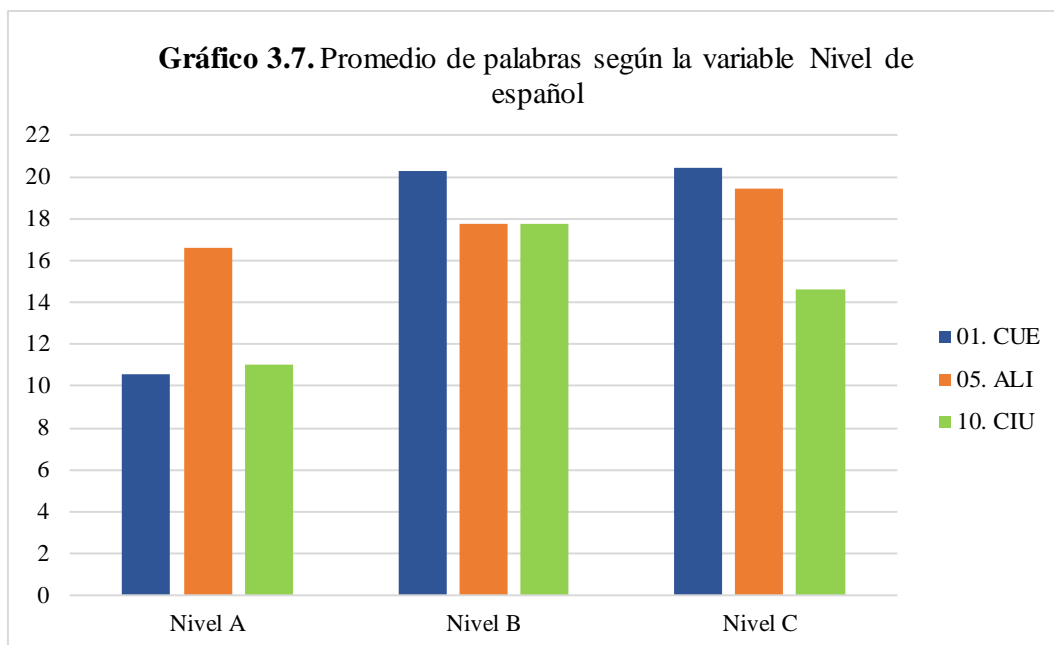
Tabla 3.17. Índices según la variante Nivel B

Analizando la productividad del grupo de nivel intermedio, nos encontramos con 01. Partes del cuerpo como el centro más productivo, seguido por los otros dos centros igualmente productivos. Por lo que se refiere a la variedad, el orden de rango cambia considerablemente, de hecho, el centro 01. Partes de cuerpo, que es el más productivo, es el menos rico. Segundo en variedad hay el centro 05. Alimentos y bebidas (+ 28,33% respecto a su antecedente) y el más rico es 10. La ciudad (+ 3,23%).

Nivel C						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	102	1	20,4	57	3	11,4
05. ALI	97	2	19,4	71	1	14,2
10. CIU	73	3	14,6	59	2	11,8

Tabla 3.18. Índices según la variante Nivel C

El grupo de nivel avanzado resulta más productivo en el centro 01. Partes del cuerpo con 102 palabras, seguido por 05. Alimentos y bebidas con 97 y finalmente por 10. La ciudad con 73 (24,74% de palabras menos que el anterior). En ese caso también, analizando la variedad de estos centros, el orden de rango cambia: el centro menos rico es 01. Partes del cuerpo, el segundo es 10. La ciudad (+ 3,39%), mientras que el más rico es 05. Alimentos y bebidas (+ 16,90%).



Comparando los promedios de palabras actualizadas por centro de interés por cada nivel, representados en el gráfico 3.7., podemos observar como el centro 01. Partes del cuerpo es el menos productivo a nivel inicial, mientras que es el más productivo tanto a nivel intermedio, como a nivel avanzado. De hecho, los estudiantes de nivel intermedio y avanzado escriben casi el doble de las palabras producidas por sus compañeros de nivel inicial, respectivamente hay un aumento de 47,65% y de 48,03% en los promedios por informante. En el caso del centro de interés 05. Alimentos y bebidas también se da este aumento, aunque más reducido: a nivel A los estudiantes producen mediamente 16,6 palabras cada uno, a nivel B hay

un incremento del 6,48% y un ulterior aumento a nivel C de 8,50%. Por lo que se refiere al centro 10. La ciudad también hay una variación ligera entre los promedios de los tres grupos: el grupo de nivel A tiene el promedio menor, seguido por el grupo de nivel C (+ 24,65%) y por el grupo B (+ 17,74% con respecto al promedio por informante del grupo C).

Analizando la variedad léxica, mostrada en el gráfico 3.8., es posible ver como el promedio de vocablos producidos por el centro de interés 01. Partes del cuerpo, sube con el nivel. De hecho, los estudiantes de nivel básico producen mediamente 7,4 vocablos por persona, los de nivel intermedio 10,75 (+ 31,16%) y los de nivel avanzado 11,4 (+ 5,70%). Los promedios de los centros 05. Alimentos y bebidas y 10. La ciudad tienen una distribución distinta: aumentan entre el nivel básico y el intermedio, pero descienden entre el intermedio y el avanzado. En el caso del primero hay un aumento de 22,67% de los vocablos entre el nivel A y el B, y una disminución de 5,33% entre el B y el C. Las diferencias son mayores en los promedios del centro 10. La ciudad: los estudiantes de nivel intermedio escriben casi el doble de los vocablos respecto a sus compañeros de nivel básico (+ 48,38%), mientras que los de nivel avanzado escriben 23,87% de vocablos menos que los de nivel intermedio.

Tanto en el promedio de palabras como en lo de los vocablos, las diferencias mayores se dan entre el nivel básico y el nivel intermedio, mientras que entre el intermedio y el avanzado los valores no siempre aumentan, incluso en algunos casos disminuyen.

CI	Densidad léxica			Índice de cohesión		
	A	B	C	A	B	C
01. CUE	1,43	1,88	1,78	0,29	0,47	0,36
05. ALI	1,43	1,18	1,37	0,29	0,30	0,27
10. CIU	1,38	1,15	1,24	0,28	0,29	0,25

Tabla 3.19. Densidad léxica e índice de cohesión por centro de interés por nivel de español

En la tabla 3.19. analizamos la densidad y el índice de cohesión de los tres grupos en cada centro de interés. El grupo de estudiantes de nivel básico tiene unos índices de cohesión y una densidad léxica bastante constantes en los tres centros de interés, debido también a la limitada cantidad de palabras y vocablos que escribieron. Hay más variación en los grupos de nivel intermedio y avanzado: en el primer caso, el centro de interés que tiene un índice de cohesión y una densidad más altos es 01. Partes del cuerpo, eso significa que hay una considerable repetición de las mismas unidades entre todos los miembros, y en el segundo, encontramos la misma situación, aunque con un índice de cohesión y una densidad ligeramente

inferiores. En los tres grupos el centro de interés con cohesión y densidad más bajos es 10. La ciudad, eso significa que hay mayor variación en las unidades aportadas por cada informante.

3.2.4. Variable Conocimiento de otras lenguas

La última variable que consideramos en nuestro estudio es el conocimiento de otras lenguas. Como explicamos al presentar las variables seleccionadas (cf. apartado 2.1.2.), partimos de la hipótesis de que conocer otras lenguas romances, además de la lengua materna que tienen todos nuestros informantes, puede influir en la producción de palabras y vocablos en español.

Los participantes tienen una distribución poco uniforme, por eso nos centraremos en una presentación de los datos generales y luego analizaremos los promedios de palabras y vocablos. El grupo de informantes que conoce otras lenguas, pero ninguna romance (LENoRom) está formado por 11 personas, mientras que el grupo que conoce otras lenguas romances (LERom) está formado por 3 personas.

Otras lenguas	Palabras	Vocablos	Promedio palabras	Promedio vocablos
LENoRom (11)	517	292	47	26,55
LERom (3)	169	127	56,33	42,33

Tabla 3.20. Índices generales según la variable Conocimiento de otras lenguas

A partir de la tabla 20, analizamos los promedios de palabras y vocablos. Por lo general los informantes que conocen también otras lenguas romances han producido más palabras (+ 16,56%) que sus compañeros. La misma situación se da también en el caso de los vocablos, con una mayor riqueza de 37,28% con respecto a los que conocen otras lenguas, pero ninguna romance.

LENoRom						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	172	2	15,64	75	3	6,82
05. ALI	190	1	17,27	115	1	10,45
10. CIU	155	3	14,09	102	2	9,27

Tabla 3.21. Índices según la variante LENOrom

El centro más productivo del grupo de estudiantes que conoce otras lenguas, pero ninguna romance, es 05. Alimentos y bebidas con 190 palabras, seguido por 01. Partes del cuerpo con 172 y el menos productivo es 10. La ciudad con 155. Analizando la variedad léxica

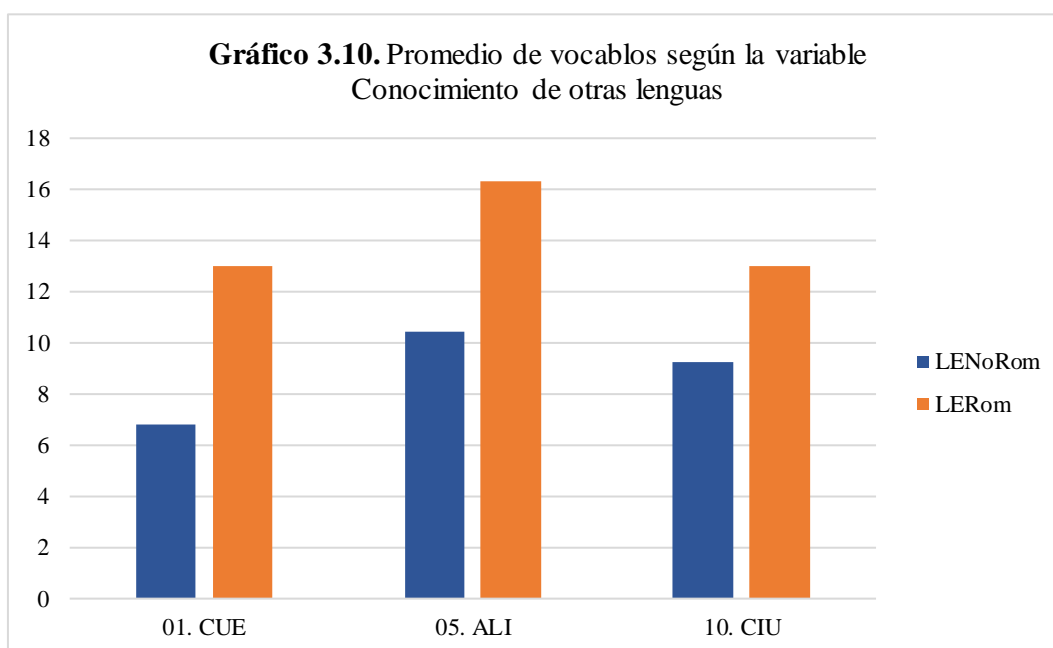
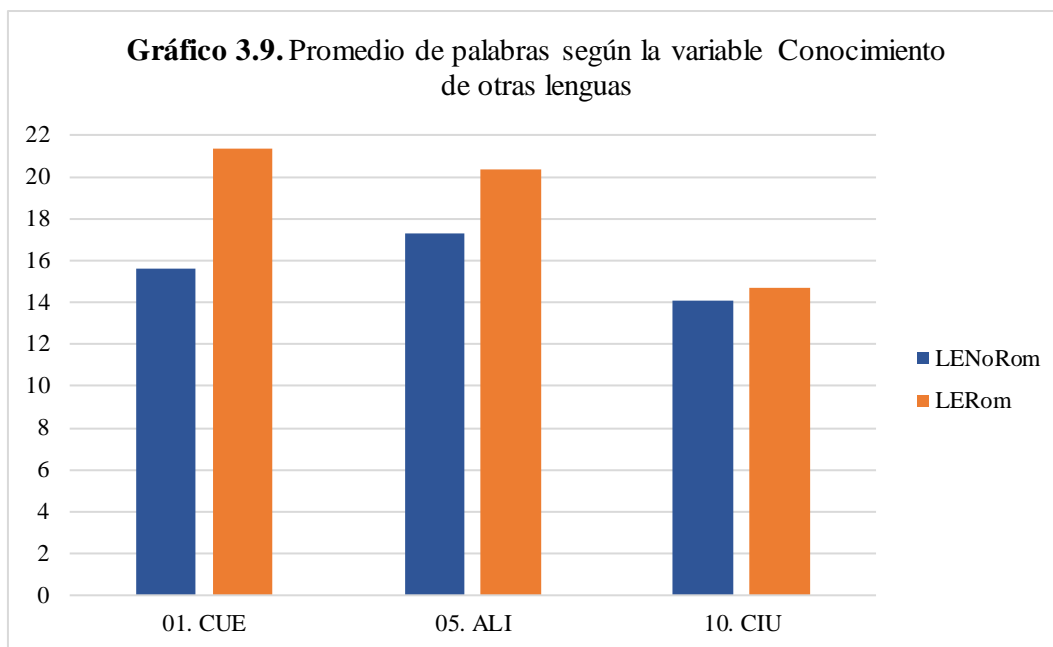
de este grupo, el orden de rango cambia: el centro más rico sigue siendo el mismo, pero en segunda posición encontramos 10. La ciudad con 102 vocablos, y por último 01. Partes del cuerpo con 75. Las diferencias entre los vocablos producidos en cada uno de ellos son notables, sobre todo entre los centros de rango 2 y 3 (26,47%).

LERom						
CI	Palabras	Rango	Promedio	Vocablos	Rango	Promedio
01. CUE	64	1	21,33	39	2	13
05. ALI	61	2	20,33	49	1	16,33
10. CIU	44	3	14,67	39	2	13

Tabla 3.22. Índices según la variante LERom

En el caso de los estudiantes que conocen también otra lengua romance además de su lengua materna y del español, el centro más productivo es 01. Partes del cuerpo con 64 palabras, seguido por 05. Alimentos y bebidas con 61 y finalmente por 10. La ciudad con 44. Entre el primero y el segundo hay una diferencia de 4,69%, mientras que entre el segundo y el tercero de 27,87%. Si consideramos el número de vocablos en cambio, encontramos que el centro más rico es 05. Alimentos y bebidas con 49 unidades, seguido por 01. Partes del cuerpo y 10. La ciudad igualmente ricos (39). Entre el primero y los otros dos hay una diferencia de 20,40%.

Como ya adelantamos no podemos comparar directamente los dos grupos ya que la productividad y variedad dependen del número de los componentes, por eso confrontamos los rangos sin considerar los valores en sí mismos. Por lo que se refiere a la productividad léxica, los dos grupos coinciden en el centro menos productivo, es decir 10. La ciudad, mientras que el centro 05. Alimentos y bebidas, que es el más productivo por el grupo LENOrom, se coloca en el rango 2 del grupo LERom. La situación contraria se da con el centro 01. Partes del cuerpo. En el caso de la variedad léxica, los dos grupos coinciden en el centro más rico, 05. Alimentos y bebidas. Los otros dos centros son igualmente ricos en el grupo LERom, mientras que en el grupo LENOrom 10. La ciudad se coloca en rango 2 y 01. Partes del cuerpo en rango 3.



Analizamos ahora los promedios de los dos grupos a través de los gráficos 3.9 y 3.10. Por lo general los estudiantes del grupo LERom actualizan más palabras en media en comparación con el otro grupo. La diferencia mayor se da en el centro 01. Partes del cuerpo (+ 26,68%), seguido por 05. Alimentos y bebidas, y finalmente el centro en el que tiene menos variación es 10. La ciudad (+ 3,95%). Por lo que se refiere al promedio de vocablos por informante, el grupo de LERom se reafirma como el más rico con diferencias aún mayores. Los dos grupos actualizan su mayor número de vocablos en 05. Alimentos y bebidas, aunque el grupo LERom escribe mediamente 16,33 unidades y el grupo LENOrom 10,45, con una variación del 36%. El centro en el que hay más diferencia es 01. Partes del cuerpo en el cual

hay una variación de 47,54%. Finalmente, el centro con menos diferencia en el promedio de vocablos es 10. La ciudad con 28,69%.

3.3. Conclusiones parciales

Al terminar este apartado de análisis cuantitativo, exponemos algunas breves conclusiones sobre los aspectos más relevantes que hemos detectado. Los dos centros que tanto en la evaluación de los datos generales como de los datos divididos por variable han sido los más productivos son 01. Partes del cuerpo y 05. Alimentos y bebidas, mientras que el centro menos productivo es 10. La ciudad. En el caso de la variedad léxica, encontramos mayor concordancia entre los datos generales y los desglosados, puesto que el centro que constantemente resulta ser el más rico es 05. Alimentos y bebidas, en cambio, el centro con el menor número de vocablos es 01. Partes del cuerpo. El análisis del índice de cohesión y de densidad léxica destaca que generalmente el centro más compacto y cerrado es 01. Partes del cuerpo, mientras que el más variado y abierto es 10. La ciudad.

Como ha ocurrido también en muchos otros estudios de disponibilidad llevados a cabo tanto con hablantes nativos de español como con estudiantes de ELE, el factor sexo incide en la productividad y en la variedad léxica, de hecho, el grupo femenino ha sido el más productivo y el más rico en comparación con el grupo masculino en todos los centros de interés analizados.

Por lo que se refiere al factor carrera, hemos registrado la influencia de los conocimientos científicos en el campo médico en la producción y variedad léxica del estudiante de medicina en el centro 01. Partes del cuerpo, que realiza un mayor número de palabras y vocablos con respecto a sus compañeros. Por lo general, entre los otros tres grupos los estudiantes de lenguas extranjeras son los que producen más unidades y más unidades distintas, de ahí que podamos inferir que también el estudio académico de ELE influye positivamente en el léxico disponible.

Seguidamente, hemos analizado el factor nivel de ELE que supone un aumento notable del promedio de palabras entre el nivel básico y el intermedio en todos los centros de interés investigados, mientras que entre el nivel intermedio y el avanzado incrementa de manera más contenida solamente en dos de ellos. En el caso del promedio de vocablos, hemos encontrado una situación distinta: el valor aumenta entre los primeros dos niveles y disminuye entre el intermedio y el avanzado en todos los tres centros de interés. Sánchez-Saus Laserna (2011)

también había encontrado esta simetría entre los niveles intermedio y avanzado, y una diferencia mayor entre el básico y el intermedio.

Finalmente hemos identificado la influencia del factor conocimiento de otras lenguas romances tanto en la producción de palabras como de vocablos. De hecho, el grupo de informantes que conoce otras lenguas romances actualizó un número de unidades mayor con respecto al grupo de estudiantes que conocen otras lenguas, pero ninguna romance.

CAPÍTULO IV

ANÁLISIS CUALITATIVO DE LOS RESULTADOS

Después del análisis cuantitativo de los datos obtenidos, presentamos en este cuarto capítulo un análisis cualitativo de los resultados. Inicialmente pondremos en evidencia los aspectos generales, examinando el léxico disponible activado en cada centro de interés y analizando los vocablos que alcanzan un índice de disponibilidad (ID) superior o igual a 0,1. De esta manera averiguamos cuáles son las temáticas predominantes y cuál es la forma flexiva más disponible³². Sucesivamente presentaremos un análisis más detallado de los resultados según las cuatro variables que consideramos en nuestro estudio. Finalmente exponemos un resumen de las conclusiones que pudimos sacar de este análisis cualitativo.

4.1. Resultados generales

En esta primera fase de análisis presentamos el léxico activado en cada centro de interés que tiene un ID superior o igual a 0,1, y un breve comentario de las temáticas y categorías gramaticales más presentes.

01. Partes del cuerpo

Rango	Vocablo	ID
1	ojos	0,622486
2	cabeza	0,552789
3	dedos	0,435316
4	piernas	0,426121
5	pelo	0,41932
6	pies	0,388433
7	manos	0,356616
8	cuello	0,33123
9	nariz	0,32594
10	mano	0,305233
11	boca	0,290203
12	brazos	0,276092
13	corazón	0,203378
14	barriga	0,186177
15	espalda	0,172591
16	orejas	0,171095

³² Recordamos que en la fase de edición de los datos hemos mantenido, cuando ambas se actualizan, tanto las formas en singular que en plural de los vocablos precisamente con el objetivo de averiguar cuál de las dos resulta más disponible.

17	lengua	0,154607
18	uñas	0,145121
19	ojo	0,135891
20	rodillas	0,130616
21	dientes	0,123317
22	muñeca	0,121011
23	labios	0,114571
24	uña	0,112962
25	garganta	0,10789
26	hombros	0,107118
27	pierna	0,100967

Tabla 4.1. Vocablos más disponibles CI 01. CUE

En la tabla 4.1. enseñamos los vocablos más disponibles, con un ID superior o igual a 0,1 del centro de interés 01. Partes del cuerpo. Las palabras en las primeras posiciones pueden considerarse como las prototípicas, estrechamente asociadas con el estímulo. La mayoría se relaciona con la parte superior del cuerpo (*ojos, cabeza, pelo, cuello, nariz, boca, espalda, orejas, lengua, dientes, labios, hombros*), con las extremidades (*dedos, piernas, pies, manos, brazos, uñas, rodillas, muñeca*) o con partes internas (*corazón, barriga, garganta*).

Todas las palabras actualizadas por los informantes son sustantivos y entre ellos encontramos tres compuestos: *globo ocular, columna vertebral y palma de la mano*. Señalamos también la presencia de la pareja *culo-trasero*, como disfemismo y eufemismo de la misma entidad.

Confrontando las diferencias en el índice de disponibilidad de las unidades que se han producido tanto en la forma singular como en plural, es evidente que la forma plural tiene una disponibilidad más alta en casi todos los casos: *ojos* es la palabra con el ID mayor de todas, y su forma singular está en la posición 19; lo mismo pasa por *piernas* (rango 4/ rango 27)³³, *pies* (rango 6/ rango 48), *manos* (rango 7/ rango 10), *brazos* (rango 12/ rango 33), *orejas* (rango 16/ rango 29), *uñas* (rango 18/ rango 24), *rodillas* (rango 20/ rango 30), *hombros* (rango 26/ rango 69) y *codos* (rango 43/ rango 46) mientras que tenemos la situación inversa en el caso de *muñeca* (rango 22/ rango 62), *pecho* (rango 32/ rango 58) y *cadera* (rango 50/ rango 53). Es curioso que esta diferencia no la hay con la palabra *dedos*, que aparece solamente en la forma plural o con *tobillo, mejilla* o *glúteo* que aparecen solamente al singular.

³³ Entre paréntesis indicamos el rango que alcanzan las dos formas. El primer rango se refiere a la palabra escrita en el texto y el segundo a la forma opuesta. Por ejemplo, en este caso *piernas* alcanza el rango 4 mientras que *pierna* rango 27.

05. Alimentos y bebidas

Rango	Vocablo	ID
1	agua	0,582957
2	leche	0,343344
3	pasta	0,320409
4	pollo	0,255224
5	arroz	0,214101
6	cerveza	0,204891
7	cachopo	0,203928
8	tomate	0,203056
9	sidra	0,185461
10	pan	0,176861
11	cebolla	0,175436
12	vino	0,169155
13	manzana	0,168132
14	queso	0,159284
15	jamón	0,157211
16	huevos	0,152979
17	carne	0,150741
18	comida	0,142857
19	chocolate	0,140828
20	piña	0,118919
21	azúcar	0,105743

Tabla 4.2. Vocablos más disponibles CI 05. ALI

El segundo centro de interés que analizamos es 05. Alimentos y bebidas. En este caso las palabras más disponibles se refieren a las bebidas (*agua, leche, cerveza, sidra, vino*), a cereales (*pasta, arroz, pan*), a carne (*pollo, jamón*), a frutas y verduras (*tomate, cebolla, manzana, piña*) y a otros productos como *queso, huevos, chocolate* y *azúcar*. En estas primeras 21 palabras encontramos varios platos y bebidas típicas de Asturias (*cachopo* y *sidra*), y en posiciones más bajas otros productos asturianos o españoles (*queso, jamón, fabada, calimocho, churros, tortilla, sangría, chorizo, morcilla, pincho, croquetas, cabrales, paella*). Señalamos también la presencia de dos términos de variación diatópica: *papas* y *frijoles*.

Es posible individuar el grupo de condimentos y aliños (*azúcar, sal, aceite, mantequilla, albahaca, vinagre, manteca de cerdo, pimentón, ajo*) y varios hiperónimos (*comida, verduras, refresco, pescado, fruta*). En posiciones más bajas encontramos también palabras que se relacionan con el centro de interés de una manera más amplia (*dieta, nutrición, salud, platos compuestos, colorantes*). Finalmente encontramos dos marcas comerciales, *Coca-Cola* y *San Martin*.

También hemos encontrado varios sustantivos compuestos o locuciones que, en la mayoría de los casos, son una especificación de un término más general: *zum de naranja*, *vino tinto*, *té caliente*, *bebidas alcohólicas*, *bebidas energizantes*, *bebidas nutrientes*, *bebidas sin azúcar*.

Por lo que se refiere a la presencia de las formas flexivas, el singular alcanza un ID considerablemente mayor al plural en siete palabras: *cerveza* (rango 6/ rango 62), *manzana* (rango 13/ rango 117), *naranja* (rango 23/ rango 102), *zanahoria* (rango 24/ rango 105), *tarta* (rango 32/ rango 126), *plátano* (rango 50/ rango 60) y *limón* (rango 95/ rango 100), mientras que solamente en tres la forma plural es la más disponible: *galletas* (rango 25/ rango 45), *patatas* (rango 27/ rango 68) y *pimiento* (rango 39/ rango 58).

10. La ciudad

Rango	Vocablo	ID
1	calle	0,44286
2	parque	0,395487
3	tiendas	0,304969
4	casas	0,271211
5	catedral	0,250812
6	plaza	0,217964
7	luces	0,190737
8	calles	0,188975
9	iglesia	0,143158
10	iglesias	0,13606
11	bar	0,126987
12	supermercado	0,123347
13	coches	0,122118
14	universidad	0,112619
15	ciudadanos	0,101804
16	ayuntamiento	0,100696
17	piso	0,100102

Tabla 4.3. Vocablos más disponibles CI 10. CIU

Analizando las palabras con un índice de disponibilidad $\geq 0,1$ es posible individuar los principales grupos de vocablos del centro 10. La ciudad. La primera palabra que aparece es *calle*, y junto con *plaza*, ejemplifican todas aquellas palabras que tienen que ver con los espacios de la ciudad y la circulación (en posiciones inferiores tenemos *acera*, *carreteras*, *camino*, *paso de cebra*, *rotonda*, *esquina*); un segundo grupo está formado por los elementos naturales de las ciudades, como *parque*, *árbol*, *flores*, *jardín*, *prados*. Otro subcampo tiene que ver con las tiendas y los servicios (*tiendas*, *supermercado*, *ayuntamiento*, *sidrería*, *tienda*,

centro comercial, restaurante, negocios, peluquería, esteticista, oficinas, pescadería, palacios, carnicería, frutería, farmacias, hospitales, centros de salud, librerías, mercado, almacén, talleres), hay también los edificios, de los cuales cuatro están incluidos en las palabras más disponibles (*casas, catedral, iglesia, piso, edificio, habitación*), el grupo de los espacios de ocio o deporte (*bar, cafeterías, teatro, gimnasio, centro deportivo, discotecas, pub, centros de ocio, cine*) o relacionados con la educación (*universidad, colegio, escuela, campus, bibliotecas*). Finalmente encontramos el grupo de los medios de transporte y sus espacios (*coches, puerto, carro, automóvil, autobús, taxi, aparcamiento, bus, parking, aeropuerto, estación de trenes, Uber, andén*) y dos grupos menos numerosos: otros elementos de la ciudad (*farola, luces, semáforo*) y las personas que viven en ella (*ciudadanos, gente, personas, población*).

Encontramos también un gran número de palabras que no están estrechamente relacionadas con el campo de la ciudad, pero que sí encajan en ello de una manera más amplia (*capital, ruido, economía, comunidades autónomas, ventanas, comida típica, compras, historia, manera de moverse, calidad de vida, calidad del aire, frío, ciudad rural, lluvia, fiestas y arte*). Por lo que se refiere a la variación diatópica del español sobresale *carro*, palabra típica de la variedad hispanoamericana.

La mayoría de las palabras que los informantes han actualizado son sustantivos, todavía hemos encontrado algunos adjetivos en combinación con uno sustantivo, como en el caso de *centro comercial, casco antiguo, comunidades autónomas, comida típica, centro deportivo, casco histórico*. Es posible individuar también varias locuciones: *manera de moverse, calidad de vida, calidad del aire, paso de cebra, estación de trenes, centros de salud y centros de ocio*.

Como ocurrió también en los otros dos centros de interés encontramos palabras tanto en singular como en plural, y en este caso las palabras más disponibles están en forma singular. De hecho, encontramos quince ejemplos en los que prevalece el singular: *calle* (rango 1/ rango 8), *parque* (rango 2/ rango 19), *plaza* (rango 6/ rango 117), *iglesia* (rango 9/ rango 10), *supermercado* (rango 12/ rango 28), *universidad* (rango 14/ rango 39), *piso* (rango 17/ rango 57), *edificio* (rango 25/ rango 56), *centro comercial* (rango 29/ rango 118), *colegio* (rango 37/ rango 64), *restaurante* (rango 40/ rango 65), *autobús* (rango 50/ rango 81), *paso de cebra* (rango 72/ rango 102), *jardín* (rango 73/ rango 80) y *cine* (rango 113/ rango 116). En cambio, solo en ocho casos la forma al plural es más disponible que la singular: *tiendas* (rango 3/ rango 27), *casas* (rango 4/ rango 18), *luces* (rango 7/ rango 115), *coches* (rango 13/ rango 21),

ciudadanos (rango 15/ rango 95), carreteras (rango 24/ rango 61), semáforos (rango 33/ rango 99) y gimnasios (rango 74/ rango 78).

4.2. Resultados por variable

En esta segunda parte del análisis cualitativo, analizamos los vocablos más disponibles (que alcanzan un ID $\geq 0,1$) según las cuatro variables. En cada tabla marcamos en gris las unidades compartidas entre las dos o más variantes, mientras que las que quedan no marcadas no aparecen entre las más disponibles de los grupos o no han sido actualizadas por todos ellos.

4.2.1. Variable Sexo

01. Partes del cuerpo

Rango	HOMBRE	MUJER
1	ojos	cuello
2	pelo	ojos
3	cabeza	cabeza
4	mano	dedos
5	pies	piernas
6	dedos	manos
7	piernas	boca
8	nariz	nariz
9	manos	pies
10	ojo	orejas
11	brazos	brazos
12	corazón	pelo
13	lengua	uñas
14	boca	espalda
15	barriga	muñeca
16	brazo	garganta
17	labios	hombros
18	espaldas	barriga
19	pierna	dientes
20	uña	rodillas
21	piel	corazón
22	cara	pestañas
23		pecho
24		mano
25		cejas
26		rodilla
27		oreja
28		culo
29		uña

Tabla 4.4. Vocablos más disponibles del CI 01. CUE según la variable Sexo

A partir de la tabla 4.4. podemos ver como las primeras 9 palabras más disponibles coinciden casi exactamente entre los dos grupos, a excepción de *pelo* (segunda más disponible entre los hombres y decimosegunda entre las mujeres) y *cuello* (palabra más disponible entre las mujeres y que, en cambio, ningún hombre ha actualizado). Este alto grado de correspondencia entre los dos grupos es debido a que, como analizamos en el capítulo anterior, este centro de interés es el que tiene un índice de cohesión mayor, lo cual implica que las unidades actualizadas por los informantes coinciden en mayor medida con respecto a los otros dos centros.

El total de palabras que alcanzan un $ID \geq 0,1$ compartidas entre los dos grupos es de 14 unidades. Entre estas palabras más disponibles, las mujeres han escrito algunas que los hombres no han escrito (*cuello, orejas, uñas, muñeca, garganta, hombros, pestañas, pecho, cejas, culo*) o que no llegan al ID requerido (*espalda, dientes, rodillas, rodilla, oreja*).

Por lo que se refiere a las diferencias entre las formas flexivas, en el grupo de los hombres predomina el plural en las palabras *ojos* (rango 1/ rango 10), *piernas* (rango 7/ rango 19), *brazos* (rango 11/ rango 16), *rodillas* (rango 26/ rango 27), mientras que solo en el caso de *mano*, la forma singular precede la plural (rango 4/ rango 9). En el caso de las mujeres hemos obtenido la misma situación: la forma plural es más disponible en *ojos* (rango 2/ rango 52), *piernas* (rango 5/ rango 39), *pies* (rango 9/ rango 40), *manos* (rango 6/ rango 24), *orejas* (rango 10/ rango 27), *brazos* (rango 11/ rango 68), *uñas* (rango 13/ rango 29), *hombros* (rango 17/ rango 59), *rodillas* (rango 20/ rango 26) y *codos* (rango 35/ rango 37), mientras que el singular prevalece en *muñeca* (rango 15/ rango 53), *pecho* (rango 23/ rango 51) y *cadera* (rango 45/ rango 47).

05. Alimentos y bebidas

Rango	HOMBRE	MUJER
1	agua	leche
2	sidra	agua
3	cachopo	pollo
4	pasta	tomate
5	vino	pasta
6	cerveza	carne
7	queso	comida
8	cebolla	arroz
9	huevos	chocolate
10	manzana	pan
11	arroz	piña

12	azúcar	café
13	jamón	bocadillo
14	pan	cerveza
15	leche	patatas
16	sal	ensalada
17	zanahoria	dieta
18	refresco	galletas
19	calabacín	pizza
20	cabracho	tarta
21	pimientos	lasaña
22	pollo	lechuga
23	calimocho	galleta
24	calabaza	espaguetis
25	plátanos	nutrición
26	chipirones	plátano
27	fabada	pimiento
28		bizcocho
29		salud
30		cerdo
31		jamón
32		vino tinto
33		hamburguesa
34		platos compuestos
35		cebolla

Tabla 4.5. Vocablos más disponibles del CI 05. ALI según la variable Sexo

Analizando las palabras más disponibles del centro de interés 05. Alimentos y bebidas, aumenta la diferencia entre las unidades de los hombres y de las mujeres. Los dos grupos comparten nueve palabras, aunque alcanzan ID muy distintos como por ejemplo *leche*, palabra más disponible entre las mujeres que se sitúa en posición 15 entre los hombres, o *pollo* (rango 3 entre las mujeres y rango 22 entre los hombres), o en el caso contrario, *cerveza* (rango 6 entre los hombres, rango 14 entre las mujeres), o *cebolla* (rango 8 y rango 35).

En los casos en los cuales los informantes hayan escrito tanto la forma plural como la singular de la misma palabra, en el grupo de los hombres el mayor índice de disponibilidad lo alcanza la forma singular en los cuatro casos: *cerveza* (rango 6/ rango 28), *zanahoria* (rango 17/ rango 58), *patata* (rango 44/ rango 74) y *limón* (rango 54/ rango 57), mientras que en el grupo de las mujeres dos palabras en su forma singular son más disponibles, *tarta* (rango 20/ rango 85) y *naranja* (rango 37/ rango 67), y dos en su forma plural: *patatas* (rango 15/ rango 78) y *galletas* (rango 18/ rango 23). Es curioso cómo, dentro de las palabras que alcanzan un $ID \geq 0,1$, los hombres actualizan las palabras *pimiento* y *plátano* en su forma plural, y casi en

las mismas posiciones encontramos las mismas palabras en forma singular en el grupo de las mujeres.

10. La ciudad

Rango	HOMBRE	MUJER
1	calle	casas
2	parque	calles
3	catedral	tiendas
4	plaza	iglesia
5	tiendas	parque
6	luces	calle
7	piso	parques
8	bar	coches
9	casa	coche
10	iglesias	luces
11	edificio	población
12	sidrería	carreteras
13	supermercado	acera
14	monumento	carro
15	puerto	prados
16	universidad	capital
17	estadio	iglesias
18	campus	semáforos
19	casas	universidades
20		ruido
21		ciudadanos
22		estatuas
23		economía
24		ayuntamiento
25		farola
26		automóvil
27		comunidades autónomas

Tabla 4.6. Vocablos más disponibles del CI 10. CIU según la variable Sexo

Por lo que se refiere al centro de interés 10. La ciudad, los dos grupos comparten seis palabras entre las más disponibles. Esta baja coincidencia es debida a que este centro era el que tenía el índice de cohesión más bajo entre los tres analizados y por lo tanto las unidades actualizadas difieren mayormente entre sí.

Entre las primeras seis palabras se concentran la mayor parte de las correspondencias, aunque cabe destacar que la palabra más disponible entre las mujeres, *casas*, está en la última posición entre las palabras que alcanzan un $ID \geq 0,1$ entre los hombres (rango 19).

Centrándonos en las formas flexivas, cuando los hombres actualizan ambas formas, prevalece la forma singular en *casa* (rango 9/ rango 19), *supermercado* (rango 13/ rango 48), *restaurante* (rango 21/ rango 30), y la plural en *tiendas* (rango 5/ rango 37) e *iglesias* (rango 10/ rango 38). En cambio, entre las mujeres prevalecen los plurales de *calles* (rango 2/ rango 6), *tiendas* (rango 3/ rango 58), *coches* (rango 8/ rango 9), *luces* (rango 10/ rango 87), *carreteras* (rango 12/ rango 39), *semáforos* (rango 18/ rango 71), *universidades* (rango 19/ rango 41), *ciudadanos* (rango 21/ rango 63), *colegios* (rango 43/ rango 90), y los singulares de *iglesia* (rango 4/ rango 17), *parque* (rango 5/ rango 7), *autobús* (rango 30/ rango 56), *supermercado* (rango 35/ rango 38), *plaza* (rango 44/ rango 83), *jardín* (rango 47/ rango 53), *centro comercial* (rango 65/ rango 88) y *cine* (rango 85/ rango 86).

4.2.2. Variable Carrera

01. Partes del cuerpo

Rango	LENGUAS	HUMANIDADES	MEDICINA	FÍSICA
1	cabeza	manos	brazo	ojos
2	ojos	ojos	pierna	piernas
3	dedos	cuello	cabeza	pelo
4	cuello	cabeza	cara	brazos
5	nariz	pelo	mano	dedos
6	boca	pies	pies	mano
7	piernas	piernas	ojo	ojo
8	pies	dedos	oreja	espaldas
9	orejas	muñeca	nariz	cabeza
10	manos	brazos	boca	manos
11	mano	garganta	labios	piel
12	barriga	oreja	pelo	
13	espalda	uñas	rodilla	
14	pelo	barbita	corazón	
15	uña	lengua	hígado	
16	dientes	hombros	intestino	
17	rodillas	corazón	pulmón	
18	brazos	boca	bazo	
19	corazón	espalda	páncreas	
20	uñas	pie	riñón	
21	pestañas	rodillas	vejiga	
22	lengua	piel	cerebro	
23	cejas		nervios	
24	rodilla		arterias	
25	hombros		huesos	
26	pecho		cartílago	
27	culo		músculos	
28	muñeca		tórax	

29	pulgar		tiroides
30	garganta		

Tabla 4.7. Vocablos más disponibles del CI 01. CUE según la variable Carrera

A partir de la tabla 4.7. es posible analizar los vocablos más disponibles ($ID \geq 0,1$) de las cuatro variantes de la variable Carrera. Entre las primeras unidades es posible individuar las palabras más prototípicas con respecto al centro de interés investigado, mientras que en las posiciones más bajas encontramos palabras más específicas y menos prototípicas.

Entre las palabras más disponibles, los cuatro grupos comparten solamente *cabeza* y *pelo*. Todavía señalamos que las palabras *corazón* y *nariz* están incluidas en la lista de palabras de cada uno de ellos, aunque en el caso de *corazón*, solamente en tres grupos alcanza el ID requerido, y en el caso de *nariz* solamente en dos.

En la tabla hemos marcado en azul aquellas palabras que aparecen en tres grupos en una forma (singular o plural) y que en el cuarto aparecen solamente en la forma opuesta. Es el caso de *ojos*, *piernas* y *brazos* que aparecen en forma plural en tres grupos y en el grupo de medicina en forma singular, o de *mano*, que aparece en tres grupos en forma singular y solamente en el grupo de humanidades en su forma plural.

Por lo que se refiere a las variantes flexivas al interior de cada grupo, entre los estudiantes de lenguas extranjeras en siete casos tiene mayor disponibilidad la forma plural: *ojos* (rango 2/ rango 44), *piernas* (rango 7/ rango 35), *manos* (rango 10/ rango 11), *rodillas* (rango 17/ rango 24), *brazos* (rango 18/ rango 58), *hombros* (rango 25/ rango 47), *caderas* (rango 38/ rango 55) y en cuatro la forma singular: *uña* (rango 15/ rango 20), *pecho* (rango 26/ rango 41), *muñeca* (rango 28/ rango 42) y *codo* (rango 32/ rango 45). En el grupo de estudiantes de humanidades encontramos solo una palabra actualizada en las dos formas y la más disponible es la plural: *pies* (rango 6/ rango 20), mientras que en el grupo de estudiantes de física hay dos ejemplos, uno en el cual prevalece la forma plural (*ojos* (rango 1/ rango 7)) y uno en el cual prevalece la forma singular (*mano* (rango 6/ rango 10)).

05. Alimentos y bebidas

Rango	LENGUAS	HUMANIDADES	MEDICINA	FÍSICA
1	agua	leche	refresco	agua
2	leche	agua	cerveza	vino
3	pollo	pasta	agua	pasta
4	tomate	comida	huevos	queso
5	piña	cerveza	pasta	sidra
6	cachopo	sidra	calabacín	cebolla

7	carne	ensalada	sopa	cachopo
8	huevos	manzana	pollo	pan
9	arroz	galleta	pavo	jamón
10	pan	bizcocho	cerdo	zanahoria
11	chocolate	plátanos	verduras	arroz
12	pasta	arroz	tomate	cabracho
13	café	tarta	aguacate	cerveza
14	comida	pizza	galletas	calimocho
15	bocadillo	cervezas	pan	manzana
16	zumo de naranja	tomate	tarta	chipirones
17	sal	azúcar	pimientos	calabaza
18	dieta	palomitas	espinacas	mandarina
19	galletas	chocolate	fabada	calamares
20	lasaña	té caliente	vacuno	gambas
21	espaguetis	carne	sidra	melocotón
22	azúcar	pollo	leche	naranja
23	lechuga	albahaca	café	fabada
24	nutrición	vino	ternera	morcilla
25	churros	cerdo		
26	salud	patatas		
27	pimiento	tinto		
28	platos compuestos	mantequilla		
29	hamburguesa	manteca de cerdo		
30	pimientos			
31	cebolla			
32	jamón			
33	vino tinto			
34	manzana			
35	verduras			
36	melocotón			
37	sangría			
38	plátano			

Tabla 4.8. Vocablos más disponibles del CI 05. ALI según la variable Carrera

Analizando el centro de interés 05. Alimentos y bebidas, es posible individuar dos palabras comunes a los cuatro grupos: *agua* y *pasta*. En los cuatro casos *agua* se encuentra siempre entre las tres más disponibles, mientras que *pasta* está en posición 2 en el grupo de humanidades y medicina, en posición 5 en medicina y en 12 en el grupo de lenguas. Además de estas dos palabras, señalamos *leche* y *pollo* que fueron actualizadas por los cuatros grupos, pero solamente en tres alcanzan un ID $\geq 0,1$.

Por lo que se refiere a las formas flexivas, en el grupo de estudiantes de lenguas extranjeras prevalece la forma singular en tres casos: *pimiento* (rango 27/ rango 30), *manzana* (rango 34/ rango 70) y *naranja* (rango 52/ rango 62), y la forma plural en uno: *tartas* (rango

76/ rango 86). En el grupo de estudiantes de humanidades se da la situación contraria: en dos casos la forma plural es más disponible (*plátanos* (rango 11/ rango 40) y *patatas* (rango 26/ rango 38)) y sólo en un caso el singular (*cerveza* (rango 5/ rango 15)). Finalmente, en el grupo de estudiantes de física sólo hay dos palabras repetidas en las dos formas y en ambos casos el singular alcanza un ID mayor con respecto al plural (*zanahoria* (rango 10/ rango 38) y *patata* (rango 28/ rango 49)).

10. La ciudad

Rango	LENGUAS	HUMANIDADES	MEDICINA	FÍSICA
1	casas	calle	catedral	calle
2	tiendas	tiendas	bar	parque
3	luces	iglesias	calle	catedral
4	calles	casas	plaza	plaza
5	supermercado	coches	parque	edificio
6	acera	piso	casco antiguo	sidrería
7	población	prados	centro comercial	puerto
8	carreteras	iglesia	gimnasios	monumento
9	semáforos	parque	parking	universidad
10	capital	ayuntamiento	estación de trenes	campus
11	carro	casa	ayuntamiento	estadio
12	iglesias	parques	supermercados	tiendas
13	estatuas	calles	oficinas	camino
14	automóvil	negocios	bibliotecas	restaurante
15	farola	ventanas		luces
16	ciudadanos	distritos		cafeterías
17	economía	ciudadanos		colegio
18	parque	plaza		gente
19	comunidades autónomas	restaurantes		casa
20	coche	aparcamiento		piso
21	pisos	personas		frío
22	peluquería	fiestas		tienda
23	comida típica	bar		teatro
24	autobús	escuela		
25	edificios			
26	compras			
27	esteticista			
28	manera de moverse			
29	coches			
30	carretera			
31	supermercados			
32	catedral			

Tabla 4.9. Vocablos más disponibles del CI 10. CIU según la variable Carrera

A partir de la tabla 4.9. hemos analizado los vocablos más disponibles del centro 10. La ciudad y la única palabra que los cuatro grupos tiene en común es *parque*, aunque tenga ID bastante diferentes; de hecho, alcanza el rango 2 en el grupo de estudiantes de física, el rango 5 en medicina, rango 9 en humanidades y rango 18 entre los estudiantes de lenguas extranjeras. Señalamos también la palabra *calle*, que se posiciona entre las tres más disponibles en tres grupos, mientras que, en el cuarto, compuesto por los estudiantes de lenguas, no alcanza el ID requerido y se sitúa en posición 66 (su forma plural está posición 4). Esta heterogeneidad es debida a que, como analizamos en el capítulo anterior, este centro de interés es el que, por lo general, alcanza el índice de cohesión bajo (0,12).

Por lo que se refiere a las variantes flexivas, en el grupo de estudiantes de lenguas extranjeras en siete casos la forma plural es más disponible que la singular: *tiendas* (rango 2/ rango 45), *calles* (rango 4/ rango 66), *carreteras* (rango 8/ rango 30), *semáforos* (rango 9/ rango 59), *iglesias* (rango 12/ rango 33), *colegios* (rango 36/ rango 72) y *universidades* (rango 38/ rango 73), mientras que en seis casos tenemos la situación opuesta: *supermercado* (rango 5/ rango 31), *parque* (rango 18/ rango 51), *coche* (rango 20/ rango 29), *autobús* (rango 24/ rango 48), *jardín* (rango 39/ rango 44) y *paso de cebra* (rango 40/ rango 62). En el grupo de estudiantes de humanidades encontramos cuatro ejemplos en los cuales el singular es más disponible: *calle* (rango 1/ rango 13), *parque* (rango 9/ rango 12), *plaza* (rango 18/ rango 39) y *universidad* (rango 25/ rango 27), y cuatro ejemplos en los cuales el plural es más disponible: *iglesias* (rango 3/ rango 8), *casas* (rango 4/ rango 11), *coches* (rango 5/ rango 31) y *ciudadanos* (rango 17/ rango 28). Finalmente, en el grupo de estudiantes de física solo se da un caso en el cual han actualizado las dos formas y el plural resulta más disponible: *tiendas* (rango 12/ rango 22).

4.2.3. Variable Nivel de español

01. Partes del cuerpo

Rango	A	B	C
1	pelo	piernas	cabeza
2	cabeza	brazos	ojos
3	ojos	ojos	dedos
4	pies	manos	cuello
5	mano	pelo	nariz
6	manos	cuello	boca
7	ojo	nariz	mano
8	dedos	garganta	piernas
9	lengua	dedos	pies

10	piernas	cabeza	orejas
11	brazo	pies	uña
12	labios	uñas	muñeca
13	corazón	piel	espalda
14	pierna	boca	barriga
15	espaldas	barriga	corazón
16	cara	rodillas	pelo
17	estómago	hombros	dientes
18	brazos	espalda	pecho
19	oreja	orejas	rodillas
20	nariz	dientes	oreja
21		corazón	rodilla
22		lengua	barbita
23		pestañas	pulgar
24		mano	labios
25			hombros
26			uñas
27			cejas
28			codo
29			manos
30			ombbligo
31			tobillo
32			pestañas

Tabla 4.10. Vocablos más disponibles del CI 01. CUE según la variable Nivel de español

A partir de la tabla 4.10., podemos analizar los vocablos más disponibles del centro de interés 01. Partes del cuerpo según el nivel de español de los informantes. Los tres grupos comparten 10 palabras entre las más disponibles, cuya mayoría está concentrada en las primeras posiciones. De hecho, si confrontamos las primeras once palabras de los tres grupos, encontramos muchas similitudes: los tres comparten cinco palabras (*cabeza, ojos, pies, dedos y piernas*), dos de ellos comparten también *pelo, mano y manos* que en el tercer grupo se posicionan más abajo entre las palabras con un $ID \geq 0,1$. Hay también tres casos (*boca, brazos y lengua*) que alcanzan el ID requerido en uno o más grupos, pero en el otro tienen un ID más bajo.

En azul hemos señalado las palabras que dos grupos actualizan en una forma flexiva (singular o plural) y el otro en la forma opuesta. Esto es el caso de *espalda, orejas y rodillas*. Por lo que se refiere a la mayor o menor disponibilidad de las formas flexivas, cuando ambas actualizadas, en el grupo de informantes de nivel inicial encontramos dos ejemplos en los cuales la forma al plural es más disponible (*ojos*, rango 3/ rango 7 y *piernas*, rango 10/ rango 14) y dos en los cuales la forma singular es más disponible (*mano*, rango 5/ rango 6 y *brazo*,

rango 11/ rango 18), mientras que en el grupo de nivel intermedio solo encontramos dos ejemplos y en ambos la forma plural prevalece (*manos*, rango 4/ rango 24 y *rodillas*, rango 16/ rango 37). La mayor diferencia la hay en el grupo de nivel avanzado en el cual parece prevalecer la forma al plural: *ojos* (rango 2/ rango 49), *piernas* (rango 8/ rango 34), *pies* (rango 9/ rango 33), *orejas* (rango 10/ rango 20), *rodillas* (rango 19/ rango 21), *hombros* (rango 25/ rango 52), *brazos* (rango 38/ rango 56), mientras que la forma singular es más disponible solamente en los casos de *mano* (rango 7/ rango 29), *uña* (rango 11/ rango 26), *pecho* (rango 18/ rango 47) y *codo* (rango 28/ rango 43).

05. Alimentos y bebidas

Rango	A	B	C
1	agua	agua	leche
2	sidra	pasta	tomate
3	pasta	leche	pollo
4	cerveza	pan	chocolate
5	queso	arroz	carne
6	cebolla	bocadillo	agua
7	vino	café	arroz
8	jamón	comida	piña
9	cachopo	ensalada	huevos
10	zanahoria	patatas	comida
11	refresco	lasaña	pan
12	calabacín	manzana	cerveza
13	cabracho	pizza	cachopo
14	pollo	calabaza	sal
15	leche	pollo	dieta
16	manzana	mandarina	tarta
17	calimocho	hamburguesa	nutrición
18	chipirones	vino tinto	espaguetis
19	huevos	palomitas	galleta
20	plátanos	naranja	lechuga
21	fabada	sangría	azúcar
22	cervezas	melocotón	plátano
23	gambas	piña	churros
24	calamares	té caliente	mantequilla
25	azúcar	carne	bizcocho
26	melocotón	vino	salud
27	arroz	cachopo	pimiento
28	sopa	chorizo	pimientos
29	morcilla	pescado	platos compuestos
30	albahaca	Coca-Cola	jamón
31		vinagre	cebolla
32		tomate	manzana

33		vodka	naranja
34		sal	verduras
35		galletas	zumos
36			zumos de naranja
37			bebidas alcohólicas
38			papas

Tabla 4.11. Vocablos más disponibles del CI 05. ALI según la variable Nivel de español

Entre las palabras con un ID $\geq 0,1$ los tres grupos comparten seis unidades (*agua, arroz, cachopo, leche, manzana y pollo*) aunque, como se puede ver de la tabla 4.11., alcanzan ID bastante variados puesto que se encuentran en posiciones muy distintas en las listas.

Por lo que se refiere a las variantes flexivas, encontramos dos ejemplos en cada grupo en los que prevalece la forma singular. En el grupo de nivel inicial *cerveza* (rango 4/ rango 22) y *patata* (rango 34/ rango 55), en el grupo intermedio *manzana* (rango 12/ rango 53) y *naranja* (rango 20/ rango 42) y en el grupo de nivel avanzado *galleta* (rango 19 / rango 44) y *pimiento* (rango 27/ rango 28).

10. La ciudad

Rango	A	B	C
1	calle	casas	luces
2	parque	calles	tiendas
3	catedral	tiendas	casas
4	plaza	parques	iglesia
5	piso	coches	supermercado
6	bar	carreteras	parque
7	iglesias	edificio	calle
8	puerto	iglesias	coche
9	sidrería	semáforos	población
10	estadio	prados	acera
11	tiendas	monumento	capital
12	campus	estatuas	carro
13	casa	calle	calles
14	universidad	ciudadanos	automóvil
15	restaurante	universidades	farola
16	ventanas	camino	economía
17	luces	distritos	ayuntamiento
18		edificios	comunidades autónomas
19		compras	negocios
20		parque	peluquería
21		supermercados	comida típica
22		plaza	pisos
23		colegios	autobús

24		casa	escuela
25		ruido	manera de moverse
26		tienda	carretera
27		personas	esteticista
28			catedral
29			universidad
30			calidad de vida
31			plaza

Tabla 4.12. Vocablos más disponibles del CI 10. CIU según la variable Nivel de español

En el CI 10. La ciudad, los tres grupos de informantes comparten cuatro palabras entre las más disponibles: *calle*, *parque*, *plaza* y *tiendas*. Pese a que alcancen un $ID \geq 0,1$, estas palabras tienen una posición de rango muy distinta en los grupos: *calle* es la más disponible del grupo de nivel inicial, en el grupo de nivel intermedio se sitúa decimotercera y en el grupo de nivel avanzado trigésima primera. Una situación parecida la encontramos también con las otras tres unidades.

Por lo que se refiere a las variantes flexivas, en el grupo de estudiantes de nivel inicial la palabra *iglesias* es más disponible que su forma al singular (rango 7/ rango 35), en cambio *restaurante* es más disponible que *restaurantes* (rango 15/ rango 23). Entre los informantes del grupo intermedio parecen prevalecer las formas plurales: *casas* (rango 1/ rango 24), *calles* (rango 2/ rango 13), *tiendas* (rango 3/ rango 26), *parques* (rango 4/ rango 20), *iglesias* (rango 8/ rango 60), *universidades* (rango 15/ rango 52), *supermercados* (rango 21/ rango 42), mientras que sólo en dos casos prevalece la singular: *monumento* (rango 11/ rango 58) y *plaza* (rango 22/ rango 54). Finalmente, en el grupo de nivel avanzado, resultan más disponibles los plurales: *luces* (rango 1/ rango 53) y *tiendas* (rango 2/ rango 37), mientras que la situación opuesta se da con *calle* (rango 7/ rango 13) y *paso de cebra* (rango 32/ rango 47).

4.2.4. Variable Conocimiento de otras lenguas

01. Partes del cuerpo

Rango	LENoRom	LERom
1	ojos	ojos
2	cabeza	cabeza
3	pies	nariz
4	pelo	piernas
5	dedos	dedos
6	manos	pelo
7	piernas	orejas
8	cuello	rodillas
9	mano	boca

10	corazón	espalda
11	boca	brazos
12	brazos	mano
13	nariz	barriga
14	ojo	pies
15	lengua	cejas
16	muñeca	pestañas
17	barriga	uña
18	dientes	manos
19	pierna	cuello
20	oreja	uñas
21	uñas	labios
22	garganta	ombligo
23	piel	glúteo
24	hombros	
25	pecho	
26	brazo	
27	rodilla	
28	espalda	

Tabla 4.13. Vocablos más disponibles del CI 01. CUE según la variable Conocimiento de otras lenguas

Los dos grupos comparten quince palabras y de estas, la mayoría se concentra en las primeras posiciones. Los primeros siete vocablos de los dos grupos son prácticamente los mismos, excepto *pies*, *nariz* y *manos* que se sitúan más abajo entre las más disponibles y *orejas* que ha sido actualizada sólo por el grupo LERom.

Por lo que se refiere a las variantes flexivas, cuando actualizadas ambas, es más disponible la forma plural en el grupo de LENOrom en seis casos: *ojos* (rango 1/ rango 14), *pies* (rango 3/ rango 42), *manos* (rango 6/ rango 9), *piernas* (rango 7/ rango 19), *brazos* (rango 12/ rango 26) y *uñas* (rango 21/ rango 36), mientras que en cuatro casos encontramos la situación opuesta: *oreja* (rango 20/ rango 30), *rodilla* (rango 27/ rango 43), *codo* (rango 41/ rango 51) y *cadera* (rango 45/ rango 48). En el grupo LERom, en cambio, encontramos dos palabras cuya forma más disponible es la plural: *rodillas* (rango 8/ rango 31) y *hombros* (rango 25/ rango 32), y dos palabras cuya forma más disponible es la singular: *mano* (rango 12/ rango 18) y *uña* (rango 17/ rango 20).

05. Alimentos y bebidas

Rango	LENoRom	LERom
1	agua	leche
2	pasta	piña
3	leche	agua
4	cerveza	huevos
5	arroz	pan
6	sidra	pollo
7	pollo	cachopo
8	cebolla	café
9	vino	zumو de naranja
10	queso	tomate
11	comida	azúcar
12	manzana	lechuga
13	tomate	pimiento
14	cachopo	sal
15	carne	hamburguesa
16	jamón	pimientos
17	pan	jamón
18	chocolate	chocolate
19	cerdo	melocotón
20	zanahoria	zumو
21	tarta	chorizo
22		melocotón
23		patatas
24		palomitas
25		galletas
26		naranja
27		helado
28		churros
29		manzana
30		limón
31		plátano
32		naranjas
33		carne

Tabla 4.14. Vocablos más disponibles del CI 05. ALI según la variable Conocimiento de otras lenguas

En el CI 05. Alimentos y bebidas, los dos grupos comparten 10 palabras entre las más disponibles. Contrariamente a lo que ocurría en el centro de interés precedente, las posiciones en las que se sitúan estos vocablos son bastante distintas, excepto por *agua* (rango 1 y rango 3), *leche* (rango 3 y 1), *pollo* (rango 7 y 6), *jamón* (rango 16 y 17) y *chocolate* (rango 18).

Analizando las variantes flexivas en el grupo de estudiantes LENOrom son más disponibles las formas al singular de cuatro palabras: *cerveza* (rango 4/ rango 49), *zanahoria*

(rango 20/ rango 91), *galleta* (rango 31/ rango 38) y *patata* (rango 52/ rango 53), mientras solo en el caso de *plátanos* el plural es más disponible (rango 43/ rango 76). En el grupo de estudiantes LERom los singulares alcanzan un ID mayor con respecto a los plurales: *pimiento* (rango 13/ rango 16), *naranja* (rango 26/ rango 32) y *manzana* (rango 29/ rango 40).

10. La ciudad

Rango	LENoRom	LERom
1	calle	casas
2	parque	tiendas
3	plaza	carreteras
4	catedral	acera
5	tiendas	luces
6	bar	supermercado
7	calles	iglesias
8	coches	calles
9	luces	farola
10	casas	estatuas
11	iglesia	pisos
12	ciudadanos	edificios
13	ayuntamiento	peluquería
14	piso	esteticista
15	universidad	catedral
16	casa	supermercados
17	sidrería	colegios
18		iglesia
19		paso de cebra
20		universidades
21		rotonda
22		jardines
23		esquina
24		parques
25		farmacias
26		teatro

Tabla 4.15. Vocablos más disponibles del CI 10.CIU según la variable Conocimiento de otras lenguas

Entre las palabras con un $ID \geq 0,1$ los dos grupos comparten seis unidades. Excepto *calles* que está en rango 7 y 8, las demás palabras se sitúan en posiciones muy distintas al considerar su índice de disponibilidad.

Por lo que se refiere a las formas flexivas, en el grupo LENOrom resulta más disponible la forma singular en nueve casos: *calle* (rango 1/ rango 7), *parque* (rango 2/ rango 37), *plaza* (rango 3/ rango 95), *iglesia* (rango 11/ rango 21), *universidad* (rango 15/ rango 80), *monumento* (rango 25/ rango 98), *restaurante* (rango 32/ rango 54), *supermercado* (rango 42/ rango 81),

autobús (rango 44/ rango 67), mientras que la forma plural prevalece en siete: *tiendas* (rango 5/ rango 22), *coches* (rango 8/ rango 18), *casas* (rango 10/ rango 16), *ciudadanos* (rango 12/ rango 77), *semáforos* (rango 29/ rango 83) y *gimnasios* (rango 57/ rango 100). En el grupo de estudiantes LERom en cambio, los plurales son más disponibles en tres casos: *iglesias* (rango 7/ rango 18), *universidades* (rango 20/ rango 37) y *parques* (rango 24/ rango 33), y sólo en el caso de *supermercado* la forma singular es más disponible que la plural (rango 6/ rango 16).

4.3. Conclusiones parciales

De este análisis cualitativo podemos concluir, como ya adelantamos anteriormente, que la casi totalidad de palabras actualizadas por los informantes son sustantivos. Solo se registran algunos casos de palabras compuestas o locuciones. En general todas las unidades que se sitúan en las posiciones más altas, según su índice de disponibilidad, son las más prototípicas, mientras que en los rangos inferiores encontramos palabras meno prototípicas y que están relacionadas con el centro de interés mediante asociaciones secundarias. Por lo que se refiere a las variantes flexivas, en los datos generales prevalece la forma plural de las palabras en el centro de interés 01. Partes del cuerpo, mientras que, en el caso de los otros dos, 05. Alimentos y bebidas y 10. La ciudad, la forma singular es más disponible que la plural.

Analizando las palabras con un índice de disponibilidad $\geq 0,1$ según las cuatro variables hemos encontrado que en el grupo de las mujeres un número mayor de palabras alcanza este valor en los tres centros de interés investigados. Según la variable carrera, los estudiantes de lenguas extranjeras son los que producen el mayor número de palabras con un ID $\geq 0,1$ en todos los centros de interés, mientras que la situación cambia en las otras posiciones. De hecho, el estudiante de medicina es el que produce más palabras con este requisito, después de los de lenguas extranjeras, en el centro 01. Partes del cuerpo, seguido por los estudiantes de humanidades y finalmente por los de física. En los otros dos centros de interés el grupo que produce el segundo mayor número de palabras con el ID requerido es lo de humanidades. Por lo que se refiere al factor nivel de ELE, hay un aumento en el número de palabras con un ID $\geq 0,1$ con el incremento del nivel en todos los centros de interés. Finalmente, el grupo de estudiantes que conoce también otras lenguas romances produce más palabras que alcanzan el ID requerido en los centros de interés 05. Alimentos y bebidas y 10. La ciudad, mientras que en 01. Partes del cuerpo se da la situación contraria.

CAPÍTULO V

ANÁLISIS DE LAS RELACIONES LÉXICAS

En este quinto capítulo presentamos un análisis de las relaciones léxicas que se establecen entre las palabras producidas en los tres centros de interés objeto de este estudio. En la primera parte explicaremos qué son las relaciones léxicas y cómo se pueden dividir, tanto en la dimensión vertical como horizontal. En la segunda parte nos centraremos en esta última dimensión, es decir en los vínculos que se establecen entre las palabras actualizadas, puesto que su organización es un indicador de la estructura del lexicón mental de los hablantes. En concreto analizaremos las relaciones entre los pares léxicos, es decir entre las parejas de palabras consecutivas que en cada centro de interés se repiten más de una vez; sucesivamente analizaremos todos los pares producidos en cada centro distinguiendo además por nivel de conocimiento de español. De esta manera averiguaremos si aumentando el nivel de competencia lingüística, cambia el tipo de relación que se establece entre las palabras.

5.1. Las relaciones léxico-semánticas

Las relaciones que se establecen entre las palabras producidas a partir de un estímulo categorial, como el título del centro de interés, pueden ser de distinta naturaleza. Como subraya Sánchez-Saus Laserna (2011: 326) “es necesario distinguir entre las relaciones que se dan dentro de la lengua (*relaciones semánticas* en general y, particularmente, en el nivel léxico, *relaciones léxicas*) y las que tienen su origen en la vinculación entre el uso de la lengua en el hablar y lo extralingüístico (*relaciones “significativas”*)”. En este sentido, las *relaciones léxicas* son aquellas que se basan en los significados de las palabras, mientras que las *relaciones “significativas”* se fundan en los vínculos lingüísticos o extralingüísticos. En este caso se pueden distinguir las *relaciones significativas designativas* que se establecen a partir de la experiencia individual del mundo (Sánchez-Saus Laserna, 2022: 172-173).

Estos tipos de relaciones se establecen entre las palabras de un centro de interés dependiendo del tipo de centro y de su índice de cohesión. Siguiendo la distinción que hizo Tomé Cornejo (2015) es posible dividir los centros de interés según dos criterios: la estructura interna y el nivel de inclusividad. Por lo que se refiere a la estructura interna pueden ser:

- a. bien definidos, que tienen límites claros que permiten determinar inequívocamente la pertinencia o la no pertinencia. Un ejemplo es 01. CUE³⁴: hay consenso en que una mano es una parte del cuerpo, pero no ocurre así con unas tijeras.
- b. categorías naturales, que se organizan en torno a un prototipo o ejemplar. Ocho centros se basan en estas categorías: 02. ROP, 04. MUE, 05. ALI, 09. ILU, 12. TRAN, 13. TRAB, 14. ANI y 16. PRO. Un perro es un ejemplar de animal más prototípico que un murciélago.
- c. categorías derivadas de metas, que se relacionan con un objetivo o finalidad específica. Un ejemplo es 06. MES: un plato tiene la finalidad de contener la comida.
- d. categorías radiales, basadas en palabras que activan varias redes conceptuales que, a su vez, activan otros conceptos que no necesariamente están relacionados con el primero. Un ejemplo es 15. JUE: se pueden identificar varios subgrupos, unos más centrales que otros, como juegos de mesa, juegos de pelota, juegos infantiles, etc.
- e. esquemas, relacionados con el mundo real, para los cuales se evocan escenas o espacios físicos, acciones, sentimientos y/o agentes. Por ejemplo: 03. CAS, 07. COC, 08. ESC, 10. CIU y 11. CAM.

Por lo que se refiere al nivel de inclusividad, es decir, a la relación entre la categoría representada por la denominación del centro de interés y las palabras actualizadas, se puede distinguir entre los centros:

1. inclusivos, en los que “X es Y” o “X es un tipo de Y” como por ejemplo “la boca es una parte del cuerpo” o “gato es un tipo de animal”. A esta categoría pertenecen: 01. CUE, 02. ROP, 03. CAS, 04. MUE, 05. ALI, 06. MES, 07. COC, 12. TRAN, 13. TRAB, 14. ANI, 15. JUE y 16. PRO.
2. relacionales, en los que “X es parte de Y” o “X está relacionado con Y”, como por ejemplo “las calles son parte de la ciudad” o “los profesores están relacionados con la escuela”. A esta categoría pertenecen: 08. ESC, 09. ILU, 10. CIU y 11. CAM.

³⁴ Para evitar sobrecargar de informaciones estos párrafos, empleamos las abreviaturas en lugar de la denominación completa de cada centro de interés: 01. CUE: Partes del cuerpo, 02. ROP: La ropa, 03. CAS: Partes de la casa (sin muebles), 04. MUE: Los muebles de la casa, 05. ALI: Alimentos y bebidas, 06. MES: Objetos colocados en la mesa para la comida, 07. COC: La cocina y sus utensilios, 08. ESC: La escuela: muebles y materiales, 09. ILU: Calefacción e iluminación, 10. CIU: La ciudad, 11. CAM: El campo, 12. TRAN: Medios de transporte, 13. TRAB: Trabajos del campo y del jardín, 14. ANI: Los animales, 15. JUE: Juegos y distracciones, 16. PRO: Profesiones y oficios.

Basándonos en estas dos divisiones, los centros de interés objeto de este estudio pertenecen a distintas categorías: 01. Partes del cuerpo es un centro bien definido e inclusivo, 05. Alimentos y bebidas es una categoría natural e inclusiva, y 10. La ciudad es un esquema y relacional.

Además de la dimensión vertical, entre el título del centro de interés y las palabras actualizadas, es posible identificar también relaciones en la dimensión horizontal, es decir, entre las palabras producidas por los informantes. Estas relaciones dependen de la estructura interna del lexicón mental.

Así como las divide Sánchez-Saus Laserna (2022), se puede distinguir entre las relaciones léxicas en sentido estricto y las relaciones significativas, que pueden a su vez dividirse en lingüísticas y designativas. Las relaciones léxicas en sentido estricto se basan en los significados de las palabras y son hiperonimia-hiponimia (*flor-tulipán*), antonimia (*calor-frío*) y parasonimia (*profesor-maestro*). En cambio, las relaciones significativas pueden ser lingüísticas cuando “los miembros de la relación forman parte del signo lingüístico” (p. 172), o designativas. En estas últimas el vínculo se establece entre las cosas designadas y se basan en las asociaciones derivadas de la experiencia personal del mundo. Estas pueden instaurarse por similitud (relaciones metafóricas, como *cabeza-melón*), por contigüidad (metonimia, *árbol-flor*), por inclusión (meronimia, *cara-ojo*) o por oposición de tipo cultural, basadas en oposiciones mentales que parten de los conocimientos de la realidad (“antonimia” del hablar, como *perro-gato* o *carne-pescado*). En algunos casos, como veremos también en el análisis, la relación entre un par puede interpretarse como más de un tipo de relación, por ejemplo, *brazos-manos* o *manos-dedos* se relacionan tanto por contigüidad, o metonimia, como por inclusión, o meronimia.

En este trabajo hemos encontrado, y tenido en cuenta, las relaciones fonológicas o gráficas, basadas en semejanzas en los segmentos de las palabras, en el cuerpo fónico o en su escritura. Por lo que se refiere a las relaciones léxico-semánticas consideramos las relaciones de hiperonimia-hiponimia, de parasonimia, de contigüidad, de meronimia, de antonimia del hablar y las relaciones contextuales, es decir todas esas asociaciones que dependen de la experiencia individual del mundo, que tienen que ver con la cultura o las experiencias comunes.

5.2. Análisis general

Antes de analizar los pares léxicos producidos en cada centro de interés por los estudiantes de cada uno de los tres niveles de competencia lingüística, analizamos los datos de manera general, sin distinguir entre los tres grupos. En este caso, contrariamente a lo que presentaremos en la segunda parte, nos centramos únicamente en los pares que han sido repetidos más de una vez.

Para este análisis hemos unificado las formas flexivas que habíamos mantenido distinguidas para averiguar el índice de disponibilidad de cada una. En este caso preferimos unificarlas puesto que la distinción entre singular y plural no es tan relevante en este tipo de evaluación. Así que, por ejemplo, los pares *manos-pies*, *mano-pie*, *manos-pie* han sido unificados en *manos-pies*. Cabe destacar también que los juicios acerca del tipo de relación establecida entre los pares léxicos han sido emitidos por la autora y acordados por un informante externo (R.C.) con fin de obtener una mayor objetividad en las evaluaciones.

01. Partes del cuerpo

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
manos	pies	4	manos	brazos	2
ojos	boca	4	manos	dedos	2
brazos	piernas	3	muñecas	codos	2
pelo	nariz	3	nariz	boca	2
barriga	espalda	2	ojos	orejas	2
boca	labios	2	ojos	pelo	2
cabeza	ojos	2	pelo	ojos	2
cabeza	pelo	2	piernas	pies	2
dedos	uñas	2	piernas	rodillas	2
hombros	espalda	2	pies	dedos	2

Tabla 5.1. Pares léxicos CI 01. CUE

La tabla 5.1. ilustra los pares léxicos presentes en el centro de interés 01. Partes del cuerpo que han sido repetidos más de dos veces. Todas las palabras actualizadas son cohipónimos de la denominación del centro de interés, es decir, son todas partes del cuerpo humano.

La mayoría de estas relaciones son de tipo designativo y se establecen por contigüidad (*ojos-boca, pelo-nariz, hombros-espalda, manos-brazos, muñeca-codos, nariz-boca, ojos-orejas, ojos-pelo, pelo-ojos y piernas-pies*). Además, aunque se podrían considerar relacionados por contigüidad, hay otros pares relacionados también por meronimia: *boca-labios, cabeza-ojos, cabeza-pelo, dedos-uñas, manos-dedos, piernas-rodillas y pies-dedos*. Algunas de estos pares se podrían también asociar por razones fonológicas o gráficas, además de semánticas, como, por ejemplo: *ojos-orejas y piernas-pies*.

Finalmente hay tres pares que, a pesar de ser cohipónimos, como partes del cuerpo, son elementos contrarios y se podrían considerar como antónimos a partir del conocimiento de la realidad: *manos-pies, brazos-piernas y barriga-espalda*.

05. Alimentos y bebidas

Palabra 1	Palabra 2	N
manzana	plátano	3
agua	leche	2
carne	pollo	2
chocolate	churros	2
pollo	pavo	2

Tabla 5.2. Pares léxicos CI 05. ALI

En el centro de interés 05. Alimentos y bebidas, los informantes han repetido solamente cinco pares más de dos veces. Entre ellos se establecen distintas relaciones: esas diez palabras son cohipónimos de las categorías alimentos o bebidas. Además, se pueden identificar otros cohipónimos de hiperónimos subordinados a alimentos: *manzana-plátano* y *pollo-pavo*, además de ser cohipónimos de alimentos son también cohipónimos de dos categorías intermedias, la fruta y la carne. Por lo que se refiere a esta última subcategoría hay también expresada la relación de subordinación en *carne-pollo*.

La misma relación de hiponimia se da con el par *agua-leche*, ambos cohipónimos de bebidas. Finalmente, entre *chocolate* y *churros* hay una relación más bien contextual que depende de la experiencia real y de esquemas culturales puesto que se suelen comer juntos.

10. La ciudad

Palabra 1	Palabra 2	N
calle	plaza	2
coche	ruido	2
colegio	universidad	2
escuela	universidad	2
parque	plaza	2

Tabla 5.3. Pares léxicos CI 10. CIU

En la tabla 5.3. se ilustran los cinco pares léxicos que han sido repetidos más de dos veces en el centro de interés 10. La ciudad. En este caso el vínculo que se establece entre la denominación del centro y estas palabras es de tipo relacional, es decir, todas las palabras designan elementos relacionados con la ciudad, contrariamente a lo que ocurría en los centros de interés anteriores donde el vínculo es de inclusión.

Entre estas diez palabras encontramos dos pares, *calle-plaza* y *parque-plaza*, que pertenecen a un subgrupo que podría ser “los espacios de una ciudad”. De la misma manera también *colegio-universidad* representan dos espacios de la ciudad relacionados con la educación. Entre *escuela-universidad* hay una relación de lingüística de hiperonimia, puesto que se puede considerar la universidad como un tipo de escuela. Finalmente, el par *coche-ruido* se asocia a la ciudad de una manera más extendida, por contigüidad y también por una relación contextual dado que dependen de la experiencia del mundo.

5.3. Análisis según la variable Nivel de español

Presentamos a continuación un análisis de los pares encontrados en cada centro de interés dividiéndolos entre los tres niveles de competencia de español de los informantes. En este caso, analizamos todos los pares que pudimos encontrar, también los que han sido creados por un único estudiante.

5.3.1. CI 01. Partes del cuerpo

Nivel básico

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
manos	pies	2	labios	pelo	1
bazo	páncreas	1	nariz	boca	1
boca	labios	1	ojo	lengua	1

brazos	piernas	1	ojos	oreja	1
cabeza	cara	1	oreja	nariz	1
cabeza	pelo	1	pelo	ojos	1
cabeza	pies	1	pierna	cabeza	1
cerebro	nervios	1	pies	dedos	1
cartílago	músculos	1	pies	manos	1
hígado	intestino	1	pulmón	bazo	1
huesos	cartílago	1	riñón	vejiga	1
labios	lengua	1	tórax	tiroides	1

Tabla 5.4. Pares léxicos CI 01. CUE nivel básico

En la tabla 5.4. se ilustran los 24 pares léxicos entre las palabras actualizadas por el grupo de informantes de nivel básico en el centro de interés 01. Partes del cuerpo. Como se puede ver, todas las palabras se relacionan con el campo representado por el centro de interés, de hecho, son todos cohipónimos de las partes del cuerpo humano. Hay solamente un par que se repite dos veces: *manos-pies*, mientras que todos los demás son únicos.

Se pueden identificar tres pares que se relacionan, además que por contigüidad, también fonográficamente: *cabeza-cara* (comparten los sonidos /ka/ iniciales y los grafos ca-), *ojo-oreja* (comparten el sonido /x/ y los grafos o-, -j-) y *tórax-tiroides* (comparten varios sonidos en el cuerpo de la palabra así como letras t-, -o- y -r-). Estos tres representan el 12% de las relaciones.

La mayoría de las relaciones se establecen por contigüidad, como *cerebro-nervios*. Además, es posible identificar varios subgrupos, como el de la *cabeza*, que contiene varios merónimos: *pelo, cara, ojos, lengua, oreja, nariz, boca y labios* o el subcentro de los órganos internos: *hígado, intestino, bazo, páncreas, riñón y vejiga*. Hay también un subcentro de las extremidades, cuyas relaciones se basan sobre todo en la configuración real del cuerpo humano, concretamente en la oposición entre las extremidades superiores y las inferiores: *manos-pies, brazos-piernas, pies-manos*. De la misma manera se relacionan: *pierna-cabeza* o *cabeza-pies*, que podrían considerarse antónimos. Como ya adelantamos anteriormente, en algunos casos las relaciones se pueden clasificar tanto por contigüidad como por meronimia, eso es el caso de *cabeza-cara* o *pies-dedos*.

Nivel intermedio

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
manos	brazos	2	espalda	hombros	1
ojos	boca	2	hombros	espalda	1
pelo	nariz	2	manos	pies	1
barbilla	patillas	1	mente	cerebro	1
barriga	espalda	1	muñecas	codos	1
boca	lengua	1	nariz	pestañas	1
brazos	muñeca	1	nariz	orejas	1
brazos	piernas	1	ojos	pelo	1
cabeza	nariz	1	orejas	cabeza	1
cabeza	pelo	1	patillas	pelo	1
cadera	culo	1	pestañas	cejas	1
corazón	barriga	1	pestañas	orejas	1
cuello	cabeza	1	piel	pelo	1
cuello	garganta	1	piernas	rodillas	1
culo	trasero	1	rodilla	talón	1
dedos	pies	1	uñas	pestañas	1
dedos	uñas	1			

Tabla 5.5. Pares léxicos CI 01. CUE nivel intermedio

A partir de la tabla 5.5. se pueden ver los 36 pares de palabras actualizadas por los informantes de nivel intermedio. En este caso, como ocurría con los estudiantes de nivel básico, todas las palabras son cohipónimos del mismo hiperónimo, las partes del cuerpo. Las relaciones que alcanzan un número mayor de repeticiones son *manos-brazos* (2), *ojos-boca* (2) y *pelo-nariz* (2), mientras que todas las demás son relaciones individuales, es decir, creadas por un único informante.

Aunque en este trabajo consideramos los pares léxicos y no las agrupaciones (o *cluster*), inevitablemente, por la estructura de la prueba de disponibilidad léxica, los pares que se individualizan en muchas ocasiones forman parte de grupos más amplios. Por esta razón se explican pares como *ojos-boca*, *pelo-nariz*, *nariz-pestañas*, *nariz-orejas*, *ojos-pelo* y *pestañas-orejas*, que considerados individualmente podrían parecer una relación forzosa, pero al tener

en cuenta la agrupación en la que se insertan³⁵, es evidente la relación de contigüidad e incluso meronímica, como partes de la cabeza. La misma relación acerca de la cabeza la mantienen también *cabeza-nariz*, *cabeza-pelo* y *orejas-cabeza*.

Otras relaciones meronímicas se dan entre *boca-lengua*, *brazos-muñecas*, *dedos-pies*, *dedos-uñas* y *piernas-rodillas*. Por supuesto estas podrían ser clasificadas también como relaciones de contigüidad, y en el caso de las últimas cuatro podrían considerarse parte del subcentro de las extremidades, juntas con *manos-brazos*, *muñecas-codos* y *rodilla-talón*. Otras relaciones de contigüidad son: *cadera-culo*, *corazón-barriga*, *cuello-cabeza*, *cuello-garganta*, *espalda-hombros*, *hombros-espalda*, *patillas-pelo* y *pestaña-cejas*. Se da también dos parejas de palabras vinculadas por parasinonimia: *culo-cerebro* y *mente-cerebro*. Es posible también individualizar pares que se relacionan por oposición, o antonimia, con respecto a la conformación real del cuerpo humano: *barriga-espalda*, *brazos-piernas* y *manos-pies*.

Además de estas relaciones semánticas es posible individuar algunas relaciones fonológicas o gráficas, como *barbilla-patillas*, *piel-pelo* y *uñas-pestañas*, que comparten respectivamente los sonidos /k/ (y gráficamente -ll-), las letras p-, -el- y /j/ (gráficamente -ñ-). Estas relaciones representan el 8,33% de todos los pares.

Nivel avanzado

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
cabeza	ojos	2	muñeca	codo	1
mano	dedos	2	muñeca	dedos	1
ojos	boca	2	nariz	boca	1
piernas	pies	2	nariz	ojo	1
barriga	espada	1	ojos	pelo	1
barriga	pubis	1	ojos	orejas	1
boca	labios	1	ombligo	barriga	1
boca	nariz	1	orejas	ojos	1
boca	orejas	1	orejas	pelo	1
brazos	piernas	1	pecho	cintura	1
cabeza	cuello	1	pelo	nariz	1
cadera	cintura	1	piernas	rodillas	1

³⁵ Las agrupaciones que contienen estos pares son: *cabeza-pelo-ojos-boca*, *ojos-pelo-nariz-orejas-cabeza* y *cabeza-nariz-pestañas-orejas*.

cadera	pelvis	1	pies	dedos	1
cejas	pestañas	1	pies	piernas	1
cintura	cadera	1	pies	rodillas	1
codo	rodilla	1	pies	tobillo	1
corazón	cabeza	1	pubis	cadera	1
costillas	huesos	1	pulgar	muñeca	1
culo	trasero	1	rodilla	mejilla	1
dedos	uña	1	rodillas	piernas	1
hombros	espalda	1	rodillas	pies	1
lengua	dientes	1	tobillo	cuello	1
manos	pies	1	uña	pulgar	1
mente	corazón	1	vagina	barriga	1

Tabla 5.6. Pares léxicos CI 01. CUE nivel avanzado

Todas las palabras producidas por los informantes de nivel avanzado en el centro de interés 01. Partes del cuerpo mantienen una relación de hiperonimia-hiponimia con la denominación del centro mismo. Entre ellas encontramos cuatro pares repetidos dos veces: *cabeza-ojos*, *mano-dedos*, *ojos-boca* y *piernas-pies*. Los cuatro están relacionados por contigüidad, y se podrían clasificar en dos subgrupos distintos: el subgrupo construido acerca de cabeza y el de las extremidades.

En el primero de estos subgrupos encontramos pares que se relacionan por contigüidad y meronimia con la cabeza: *boca-nariz*, *boca-orejas*, *boca-labios*, *cabeza-cuello*, *cejas-pestañas*, *lengua-dientes*, *nariz-ojo*, *nariz-boca*, *ojo-pelo*, *ojos-orejas*, *orejas-ojos*, *orejas-pelo* y *pelo-nariz*³⁶. Además, entre *ojos-orejas* y *oreja-ojos* también se puede apreciar una relación fonológica puesto que comparten los sonidos /x/ y /o/.

Otro subgrupo es el de las extremidades en el cual encontramos relaciones meronímicas: *dedos-uña*, *mano-dedos*, *pies-dedos*, o relaciones por contigüidad: *muñeca-codo*, *muñeca-dedos*, *piernas-pies*, *piernas-rodillas*, *pies-tobillo*, *pies-rodillas*, *pies-piernas*, *pulgar-muñeca*, *rodillas-piernas*, *rodillas-pies* y *uña-pulgar*. En este subgrupo es posible individuar

³⁶ Puede parecer que entre algunos de estos pares se haya forzado la relación. Pero, como explicamos anteriormente, hay que considerar que las palabras están producidas en secuencias más largas, de ahí que algunos pares podrían parecer desconectados, pero en el conjunto no lo son. Las agrupaciones que se han creado en este caso, y que explican estos pares son: *nariz-ojo-pelo*, *cabeza-ojos-boca-nariz*, *oreja-ojos-barbita*, *cabeza-ojos-orejas-pelo-nariz-boca-labios-cejas-pestañas* y *cabeza-cuello-ojos-boca-orejas-dientes*.

también relaciones de oposición o antonimia con respecto a la configuración real del cuerpo humano: *brazos-piernas*, *codo-rodilla* y *manos-pies*. Entre estos pares del subgrupo de las extremidades es posible individuar dos pares conectados, además que por contigüidad, también por razones fonológicas y gráficas: *piernas-pies* y *pies-piernas*. De la misma manera se vinculan otros dos pares compuestos por partes del cuerpo muy lejanas, así que su única relación es fonológica y gráfica: *rodillas-mejilla* y *tobillo-cuello*, que comparten el sonido /l/ (y gráficamente -ll-). Entre los 54 pares creados por los informantes de nivel avanzado, las relaciones fonológicas se establecen (únicamente o también por contigüidad) en el 12,96 % de los casos.

Otras relaciones de contigüidad se encuentran en palabras que designan partes del tronco: *barriga-pubis*, *cadera-cintura*, *cadera-pelvis*, *cintura-cadera*, *hombros-espalda*, *ombligo-barriga*, *pecho-cintura*, *pubis-cadera* y *vagina-barriga*. Hay también el par *barriga-espalda*, que podría conectarse por contigüidad o por oposición, puesto que la barriga está en la parte anterior del cuerpo y la espalda en la posterior. Finalmente, es posible también apreciar un caso de relación de hiponimia-hiperonimia (*costillas-huesos*), uno de parasonimia (*culo-trasero*) y dos casos de oposición derivadas de la polaridad entre las dos palabras (*corazón-cabeza* y *mente-corazón*).

5.3.2. CI 05. Alimentos y bebidas

Nivel básico

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
agua	calimocho	1	pavo	cerdo	1
bocadillos	pincho	1	pimientos	espinacas	1
cachopo	cabracho	1	pollo	lomo	1
calabacín	zanahoria	1	pollo	pavo	1
calimocho	cerveza	1	queso	bocadillos	1
cebolla	zanahoria	1	queso	jamón	1
cebolla	tortilla	1	refresco	cerveza	1
cerveza	agua	1	sidra	agua	1
chipirones	calamares	1	sidra	leche	1
fabada	morcilla	1	sopa	pollo	1
galletas	pan	1	tinto	vino	1
leche	arroz con leche	1	tomate	aguacate	1

leche	café	1	tortilla	patatas	1
manzana	plátano	1	verduras	tomate	1
pan	tarta	1	vino	sidra	1
patata	hamburguesa	1			

Tabla 5.7. Pares léxicos CI 05. ALI nivel básico

En la tabla 5.7. se pueden ver los pares léxicos individualizables entre las palabras actualizadas por el grupo de informante de nivel básico en el centro de interés 05. Alimentos y bebidas. Todos los pares son únicos, es decir producidos solamente por un informante. Todos los elementos se relacionan con uno de los dos subcampos que componen el CI: “alimentos” y “bebidas”. De hecho, todas las palabras mantienen una relación de hiperonimia-hiponimia con estos dos subgrupos, aunque a su interior es posible identificar ulteriores distinciones.

Dentro del subcampo “alimentos” es posible individuar el subgrupo de la carne y de los platos típicos españoles de carne (*cachopo-cabracho, fabada-morcilla, pavo-cerdo, pollo-pavo* y *pollo-lomo*), el grupo del pescado, con una relación de contextual de plato-ingrediente (*chipirones-calamares*), de la fruta (*manzana-plátano*) y de la verdura (*calabacín-zanahoria, cebolla-zanahoria* y *pimientos-espinacas*). Hay también el subgrupo de cereales y productos con harina (*galletas-pan, pan-tarta*) y finalmente un ejemplo de relación de hiperónimo-hipónimo: *verduras-tomate*³⁷.

Por lo que se refiere al subcampo de las “bebidas” hay un par compuesto por lo que se podría considerar un hiperónimo, *refresco*, con un hipónimo, *cerveza*. La mayoría de los otros pares mantienen una relación de cohipónimos del subcampo: *agua-calimocho, calimocho-cerveza, cerveza-agua, leche-café, sidra-leche, sidra-agua, tinto-vino* y *vino-sidra*.

Además, se pueden identificar cuatro relaciones fonológicas o gráficas (*cachopo-cabracho, cebolla-tortilla, leche-arroz con leche* y *tomate-aguacate*) que representan el 12,9% de las relaciones y vínculos establecidos por el conocimiento del mundo o el contexto: *cebolla-tortilla, fabada-morcilla, sopa-pollo, queso-bocadillos, tortilla-patatas*, como ingredientes y plato, y *bocadillos-pincho, leche-café, patata-hamburguesa, queso-jamón* puesto que suelen aparecer juntos.

³⁷ Nótese que el tomate en realidad es un fruto, aunque en la cultura occidental se suele consumir y considerar como verdura. A la luz de esta costumbre aceptamos este par léxico.

Nivel intermedio

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
agua	leche	1	pasta	lasaña	1
agua	vino	1	pescado	dorada	1
agua	vino tinto	1	pesto	macarrones	1
cabra	cerdo	1	pizza	palomitas	1
café	agua	1	pizza	pasta	1
carne	cachopo	1	pollo	hamburguesa	1
Coca-Cola	vodka	1	pollo asado	fabada	1
comida	ensalada	1	sal	aceite	1
mandarina	naranja	1	tartas	pan	1
manzana	calabaza	1	té caliente	leche	1
manzanas	kiwis	1	tomate frito	pesto	1
melocotón	piña	1	vinagre	sal	1
pan	arroz	1	vino	vinagre	1
pan	pizza	1	vino tinto	sangría	1
pasta	arroz	1	zum de naranja	naranjas	1

Tabla 5.8. Pares léxicos CI 05. ALI nivel intermedio

Como ocurría también en las palabras actualizadas por los informantes de nivel básico, los estudiantes de nivel intermedio actualizan palabras que son cohipónimos de los dos grupos “alimentos” y “bebidas”. Por lo que se refiere a los primeros encontramos varios subgrupos: la carne (*cabra-cerdo*, *carne-cachopo*, *pollo-hamburguesa*), el pescado (*pescado-dorada*), las frutas (*mandarina-naranja*, *manzanas-kiwis* y *melocotón-piña*) y los cereales o productos con harina (*pan-pizza*, *pan-arroz*, *pasta-arroz*, *pasta-lasaña*, *pizza-pasta*, y *tartas-pan*), los condimentos y aliños (*sal-aceite*, *tomate frito-pesto* y *vinagre-sal*). Los elementos de algunos de estos pares no son solamente cohipónimos de “alimentos” y cohipónimos a los varios subgrupos, sino también se establece una relación de subordinación entre las dos palabras como en el caso de *carne-cachopo*, *comida-ensalada*, *pasta-lasaña* y *pescado-dorada*.

En el subgrupo de las “bebidas” encontramos varios pares: *agua-leche*, *agua-vino*, *agua-vino tinto*, *café-agua*, *Coca-Cola-vodka*, *té caliente-leche*, *vino tinto-sangría*, *zum de naranja-naranjas*. También hay tres ejemplos de pares influenciados por razones fonológicas o gráficas: *manzana-calabaza*, *vino-vinagre* y *zum de naranja-naranjas*, que constituyen el

10% de todas las relaciones. Finalmente, es posible distinguir pares conectados también por razones contextuales como *Coca-Cola-vodka*, *pesto-macarrones*, *pizza-palomitas*, *pizza-pasta*, *pollo asado-fabada* y *té caliente-leche*.

Nivel avanzado

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
carne	pollo	2	mantequilla	manteca de cerdo	1
chocolate	churros	2	melocotón	naranja	1
manzana	plátano	2	naranja	manzana	1
aceite	aceitunas	1	naranja	plátano	1
agua	leche	1	nutrición	salud	1
agua	zumو de naranja	1	paella	arroz	1
almendras	turrón	1	pan	tomate	1
arroz	espagueti	1	pasta	tomate	1
azúcar	sal	1	pimentón	sal	1
bebidas alcohólicas	bebidas energizantes	1	piña	melocotón	1
bebidas energizantes	bebidas nutrientes	1	plátano	uva	1
bebidas nutrientes	bebidas sin azúcar	1	pollo	arroz	1
bizcocho	tarta	1	pollo	cerdo	1
cacahuete	chocolate	1	pollo	pavo	1
cerveza	leche	1	sal	ajo	1
churros	dulces	1	tarta	chocolate	1
comida	dieta	1	tarta	queso	1
dieta	nutrición	1	tomate	cebolla	1
dulces	caramelos	1	tomate	lechuga	1
espagueti	tomate	1	tortilla	huevos	1
fruta	manzana	1	uva	fresa	1
galleta	bizcocho	1	verduras	carne	1
galletas	chocolate	1	vino	agua	1
leche	galleta	1	yogur	galletas	1
leche	zumو	1	zumو	piña	1
lechuga	pimiento	1			

Tabla 5.9. Pares léxicos CI 05.ALI nivel avanzado

En la tabla 5.9. se ilustran los pares léxicos producidos por los informantes de nivel avanzado en el centro de interés 05. Alimentos y bebidas. Hay tres pares de palabras que se repiten dos veces: *carne-pollo*, *chocolate-churros* y *manzana-plátano*. Todos los demás pares representan relaciones producidas por un solo informante.

Contrariamente a lo que ocurría con los otros dos grupos de estudiantes, no todas las palabras actualizadas se relacionan directamente con la denominación del centro, o, mejor dicho, no todas son cohipónimos de “alimentos” o de “bebidas”. De hecho, hay unos pares que tienen que ver más bien con la alimentación: *comida-dieta*, *dieta-nutrición* y *nutrición-salud*.

Todos los demás elementos actualizados son cohipónimos de uno de los dos macro grupos “alimentos” y “bebidas”, al interior de los cuales es posible distinguir otros subgrupos. Entre los alimentos hay: la carne (*carne-pollo*, *pollo-cerdo* y *pollo-pavo*), las verduras (*lechuga-pimiento*, *tomate-cebolla* y *tomate-lechuga*), la fruta (*fruta-manzana*, *manzana-plátano*, *melocotón-naranja*, *naranja-manzana*, *naranja-plátano*, *piña-melocotón*, *plátano-uva* y *uva-fresa*), los dulces (*biscocho-tarta*, *cacahuete-chocolate*, *chocolate-churros*, *churros-dulces*, *dulces-caramelos*, *galleta-biscocho*, *galletas-chocolate*, *tarta-chocolate*, *tarta-queso* y *yogur-galletas*), los cereales (*arroz-espagueti*) y los condimentos y aliños (*azúcar-sal*, *mantequilla-manteca de cerdo*, *pimentón-sal* y *sal-ajo*).

Por lo que se refiere a las bebidas, es posible distinguir entre el grupo de las bebidas sin alcohol (*agua-leche*, *agua-zumo de naranja*, *leche-zumo* y *zumo-piña*) y los pares compuestos también por bebidas alcohólicas (*cerveza-leche* y *vino-agua*). Además, hay también un grupo de cohipónimos que ejemplifican un tipo general de bebidas (*bebidas alcohólicas-bebidas energizantes*, *bebidas energizantes-bebidas nutrientes*, *bebidas nutrientes-bebidas sin azúcar*).

Entre los pares actualizados hay también relaciones de hiperonimia-hiponimia y al revés: *biscocho-tarta*, *carne-pollo* y *fruta-manzana*. Y relaciones instauradas a partir de razones fonológicas o gráficas: *aceite-aceitunas*, *bebidas alcohólicas-bebidas energizantes*, *bebidas energizantes-bebidas nutrientes*, *bebidas nutrientes-bebidas sin azúcar*, *cacahuete-chocolate*, *chocolate-churros*, y *mantequilla-manteca de cerdo*, que representan el 12,96% de las relaciones.

Finalmente se dan también relaciones contextuales puesto que estos alimentos o bebidas suelen aparecer juntos o porque son un ingrediente del plato: *espagueti-tomate*, *leche-galleta*, *pan-tomate*, *pasta-tomate*, *pollo-arroz* y *almendras-turrón*, *paella-arroz*, *tarta-queso*, *tarta-*

chocolate, tortilla-huevos. También hay un par que representa una antonimia del hablar, basada en la experiencia del mundo que tenemos los hablantes y es: *verdura-carne*.

5.3.3. CI 10. La ciudad

Nivel básico

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
calle	plaza	2	piso	casa	1
cafeterías	gente	1	piso	habitación	1
calle	catedral	1	plaza	parque	1
campus	universidad	1	plaza	sidrería	1
catedral	calle	1	plaza	teatro	1
iglesia	historia	1	sidrería	plaza	1
frío	lluvia	1			

Tabla 5.10. Pares léxicos CI 10.CIU nivel básico

El centro de interés 10. La ciudad ha sido el menos rico entre los estudiantes de nivel básico, de ahí que tengamos pocos pares entre las palabras actualizadas. Hay solamente una relación repetida dos veces: *calle-plaza*, mientras que todas las demás son únicas.

Todas las palabras se relacionan con la denominación del centro de interés, siendo partes de la ciudad o relacionándose con la realidad de la ciudad. Entre las partes de la ciudad encontramos: *calle-catedral, calle-plaza, campus-universidad, catedral-calle, iglesia-historia, plaza-parque, plaza-sidrería, plaza-teatro y sidrería-plaza*.

A partir de la palabra piso hay dos pares: *piso-casa* en el cual se establece una relación de sinonimia o de hiponimia-hiperonimia dado que un piso es un tipo de casa, y *piso-habitación* relacionados por meronimia. Finalmente hay dos relaciones contextuales que dependen de la experiencia del contexto ciudad de cada persona: *cafeterías-gente* y *frío-lluvia*.

Nivel intermedio

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
colegio	universidad	2	frutería	supermercado	1
autobuses	taxi	1	historia	casco antiguo	1
bar	pub	1	hospital	centro de salud	1
calle	camino	1	jardines	parques	1

calles	semáforos	1	museos	librería	1
carnicería	frutería	1	oficinas	talleres	1
casco antiguo	casco histórico	1	parques	calles	1
centro comercial	tiendas	1	parque	plaza	1
centros de salud	centros de ocio	1	personas	fiestas	1
cines	centros comerciales	1	pescadería	carnicería	1
coches	ruido	1	plazas	monumentos	1
escuela	colegios	1	taxi	Uber	1
farmacias	hospitales	1	tiendas	compras	1

Tabla 5.11. Pares léxicos CI 10. CIU nivel intermedio

En la tabla 5.11. se ilustran los pares léxicos de los estudiantes de nivel intermedio en el centro de interés 10. La ciudad. Todos los pares han sido actualizados una vez, excepto *colegio-universidad*, actualizado por dos informantes.

En este caso también es posible individuar varios subgrupos de palabras que tienen que ver de manera distinta con la ciudad: hay los medios de transporte (*autobuses-taxi* y *taxi-Uber*), los espacios educativos (*colegio-universidad* y *escuela-colegios*), los espacios conectados con la salud (*farmacias-hospitales* y *hospital-centro de salud*) y otros espacios urbanos o elementos de la ciudad (*calles-semáforos*, *casco antiguo-casco histórico*, *coches-ruido*, *historia-casco antiguo*, *museos-librerías*, *parques-calles*, *parque-plaza*, *personas-fiestas*, *plazas-monumentos*, *tiendas-compras*), todos conectados contextualmente con la denominación del centro de interés.

Otros tres pares que se podrían considerar relacionados contextualmente o por meronimia *centro comercial-tiendas*, *cines-centros comerciales* y *frutería-supermercado*. Entre *carnicería-frutería*, *casco antiguo-casco histórico*, *centros de salud-centros de ocio*, *jardines-parques* y *pescadería-carnicería* se establecen relaciones fonológicas y/o gráficas (18,51%). Finalmente se pueden apreciar pares de sinónimos: *bar-pub*, *calle-camino*, *casco antiguo-casco histórico*, *jardines-parques* y *oficinas-talleres*.

Nivel avanzado

Palabra 1	Palabra 2	N	Palabra 1	Palabra 2	N
escuela	universidad	2	coche	autobús	1
acera	calles	1	coche	ruido	1

autobús	carretera	1	flores	árbol	1
automóvil	coche	1	luces	tiendas	1
bar	restaurante	1	parque	jardín	1
calidad de vida	calidad del aire	1	parque	plaza	1
calles	farola	1	peluquería	esteticista	1
carro	automóvil	1	semáforo	paso de cebra	1
casas	pisos	1	supermercado	centro comercial	1
catedral	iglesia	1	tienda	supermercado	1
comunidades autónomas	comida típica	1	tiendas	negocios	1

Tabla 5.12. Pares léxicos CI 10. CIU nivel avanzado

En la tabla 5.12. se presentan los pares léxicos del centro de interés 10. La ciudad del grupo de estudiantes de nivel avanzado. La única relación que se repite es *escuela-universidad* que podría considerarse de hiperonimia-hiponimia dado que escuela es un término más general que universidad. El mismo tipo de relación se da también entre *catedral-iglesia*, puesto que una catedral un tipo de iglesia.

Todos los pares léxicos actualizados se vinculan al tema “ciudad” contextualmente y por ser parte de ella o relacionados con ella. A su interior, se pueden identificar varios subgrupos: los medios de transporte (*automóvil-choche*, *carro-automóvil* y *coche-autobús*) y las vías públicas con sus elementos (*autobús-carretera*, *acera-calles*, *calles-farola*, *semáforo-paso de cebra*), otros espacios urbanos (*bar-restaurante*, *casas-pisos*, *peluquería-esteticista*, *supermercado-centro comercial*, *tienda-supermercado*, *tiendas-negocios*) o a los espacios verdes en la ciudad (*flores-árbol*, *parque-jardín* y *parque-plaza*). Finalmente encontramos pares que se relacionan de manera aún más amplia con el tema (*calidad de vida-calidad del aire*, *comunidades autónomas-comida típica*, *coche-ruido*, *luces-tiendas*).

El único par que podría considerarse relacionado por razones fonológicas o gráficas es *calidad de vida-calidad del aire* (4,34%); entre los demás hay varios ejemplos de meronimia: *acera-calles*, *calles-farola*, *supermercado-centro comercial* y *tienda-centro comercial*, y de sinonimia: *automóvil-choche*, *carro-automóvil*, *casas-pisos*, *parque-jardín* y *tiendas-negocios*.

5.4. Conclusiones parciales

Después de este análisis exponemos ahora unas breves conclusiones. Generalmente, como esperábamos el tipo de centro de interés influye en las relaciones que se establecen entre las palabras. De hecho, en el centro 01. Partes del cuerpo, que al principio definimos como bien definido e inclusivo, encontramos todas palabras que se relacionan con su denominación por ser “partes del cuerpo humano”. A su interior las relaciones que prevalecen son de contigüidad y de meronimia, dada, evidentemente, la conformación real del cuerpo humano al que hace referencia. En el centro 05. Alimentos y bebidas, categoría natural e inclusiva, encontramos sobre todo palabras prototípicas, aunque no son limitados los ejemplos de palabras más complejas. Dentro de este campo, todas las palabras actualizadas (excepto cuatro) pueden considerarse cohipónimos de los dos grupos que forman este centro, los alimentos y las bebidas. Además, son también numerosas las relaciones establecidas por razones contextuales. En el centro 10. La ciudad, que describimos como centro esquema y relacional, todas las palabras actualizadas se pueden explicar por la evocación del espacio urbano, de sus elementos y sus agentes, y todas están relacionadas con ellos. Las relaciones que predominan son de tipo contextual, que dependen de la experiencia personal del mundo, o por ser elementos que forman parte de la ciudad.

Entre las palabras actualizadas por los tres grupos en cada centro de interés es siempre posible individuar varios subgrupos a su interior. Así por ejemplo en el centro 01. Partes del cuerpo predominan cohipónimos de la cabeza, de las extremidades y de los órganos internos; en 05. Alimentos y bebidas, además de dividirse en los dos subgrupos, hay numerosos cohipónimos de los subgrupos carne, pescado, fruta, verdura, cereales y aliños; y en el centro 10. La ciudad hay subgrupos relativos a los espacios urbanos, a los medios de transporte, a los espacios educativos y a los servicios.

Una diferencia que se puede apreciar a partir del análisis de los pares léxicos divididos por nivel de español, es que tanto en el centro 01. Partes del cuerpo como en 05. Alimentos y bebidas, los estudiantes de nivel avanzado crean casi el doble de pares léxicos que sus compañeros de los niveles básico e intermedio. Esto nos demuestra que, avanzando en el nivel de competencia lingüística hay también un desarrollo en la organización del léxico. Solamente en el centro 10. La ciudad los estudiantes de nivel intermedio producen más pares que los de nivel avanzado (27 frente a 23).

Uno de los objetivos de este análisis era comprobar si podíamos detectar la presencia de relaciones distintas en función del distinto nivel de conocimiento de español. De hecho, nos esperaríamos que, aumentando el nivel lingüístico disminuyan las relaciones basadas en la fonología o en la grafía y aumenten las relaciones semánticas. Comparando los resultados en el centro 01. Partes del cuerpo, en los tres grupos predominan las relaciones de contigüidad y de meronimia. Por lo que se refiere a las relaciones fonológicas y gráficas, tanto el grupo de nivel básico como el avanzado producen más pares de este tipo con respecto a sus compañeros de nivel intermedio. Aunque es necesario señalar que los tres pares conectados fonológicamente o gráficamente del grupo de nivel básico podrían también considerarse relacionados por contigüidad, lo mismo pasa con cuatro de los seis pares del grupo de nivel avanzado, pero solamente con uno de los tres de nivel intermedio.

En los tres grupos las relaciones que predominan en el centro de interés 05. Alimentos y bebidas son de hiponimia, frente a los dos hiperónimos representados por la denominación del centro, y contextuales, es decir que dependen de la experiencia real de cada persona. En este caso también el grupo de nivel básico y avanzado producen un porcentaje de pares relacionados fonéticamente o gráficamente más alto que los estudiantes de nivel intermedio. Sin embargo, por el grupo de nivel básico tres de los cuatro pares tienen también una explicación contextual, por el grupo de nivel intermedio solamente uno de los tres tiene otra justificación, y en el grupo de nivel avanzado solamente tres de los siete.

Finalmente, por lo que se refiere al centro de interés 10. La ciudad, la mayoría de los pares están conectados con la denominación del centro por ser relacionados o por ser partes de la ciudad. Por lo que se refiere a las relaciones fonológicas o gráficas el grupo de nivel intermedio supera los otros con cinco pares, aunque todos se podrían asociar también contextualmente, el grupo de nivel avanzado produce solamente un par conectado fonológicamente y el grupo de nivel básico ninguno.

De este resumen, podemos concluir que no hay mucha diferencia entre el tipo de relaciones que pudimos encontrar entre los pares léxicos producidos por los estudiantes de los tres niveles puesto que en la mayoría de los casos las relaciones basadas en razones fonológicas o gráficas se podrían también justificar por razones semánticas. No obstante, reconocemos que los resultados obtenidos no se puedan generalizar puesto que los tres grupos estaban compuestos por 4 o 5 personas y que la mayoría de las relaciones que encontramos no alcanzaban más de una repetición.

CONCLUSIONES

La investigación que presentamos ha tenido como objetivos los de presentar nuevos datos del léxico disponible de los estudiantes italianos de ELE en contexto inmersivo y analizar los pares léxicos individualizables en las listas de palabras obtenidas. Como explicamos en el primer capítulo, la disponibilidad léxica es un campo de la lingüística que ha tenido –y sigue teniendo– su mayor desarrollo en campo hispánico con numerosos estudios con hablantes nativos y no nativos. Sin embargo, hasta hoy en día sigue habiendo pocas publicaciones de investigaciones llevadas a cabo enteramente con aprendices italianos. Por eso elegimos acercarnos a esta disciplina e investigar el léxico disponible de este grupo de estudiantes.

Tras un primer capítulo teórico dedicado a la descripción de los orígenes de la disponibilidad léxica, de su suceso en el mundo hispánico y de su reciente aplicación a los estudiantes de ELE, presentamos en el segundo capítulo las características de nuestro trabajo. El grupo de informantes, compuestos por 14 estudiantes de español, se encontraba en Oviedo para participar al programa Erasmus. Les administramos una prueba de disponibilidad para obtener las palabras más disponibles en tres centros de interés: 01. Partes del cuerpo, 05. Alimentos y bebidas y 10. La ciudad, y un cuestionario sociológico del cual hemos obtenido las variables, con las cuales hemos analizado los datos, y una serie de informaciones acerca de sus usos, opiniones y actitudes hacia el español. Los resultados más interesantes del cuestionario sociológico son sobre las actitudes: generalmente nuestros informantes están de acuerdo o muy de acuerdo con el hecho de que el español es fácil de aprender y que es más fácil aprenderlo en España; la mayoría piensa que el mejor español se habla en España, mientras que cuatro personas están poco de acuerdo o en desacuerdo con esta afirmación, y finalmente la mayoría piensa que sus conocimientos previos eran suficientes para vivir en España y para entender las clases.

Por lo que se refiere a los resultados cuantitativos, destacan como más productivos los centros 01. Partes del cuerpo y 05. Alimentos y bebidas, mientras que 10. La ciudad es generalmente el menos productivo. Analizando la variedad léxica, es decir los vocablos, el centro más rico es 05. Alimentos y bebidas mientras que el con menos palabras diferentes es 01. Partes del cuerpo. Esto se explica también con los índices de cohesión y de densidad léxica: el centro 01. Partes del cuerpo es el más compacto y cerrado, debido a que los vocablos producidos por los informantes coinciden en mayor medida, mientras que el centro más variado y abierto es 10. La ciudad.

A partir del análisis de estos valores según las cuatro variables que consideramos, es posible identificar una mayor producción y variedad léxica en las mujeres que en los hombres en todos los centros de interés investigados. Debido a la heterogeneidad numérica de los grupos divididos según las otras tres variables, hemos elegido analizar los promedios en lugar de las cantidades totales. Por lo que se refiere a la carrera, generalmente los estudiantes de lenguas extranjeras son los que alcanzan un promedio mayor en la productividad y la variedad léxica, aunque hemos detectado una notable influencia de los conocimientos científicos en el campo médico en la producción y riqueza léxica del estudiante de medicina en el centro de interés 01. Partes del cuerpo. El factor nivel lingüístico de ELE incide mayormente en la producción de palabras entre el nivel básico y el intermedio en los tres centros de interés, mientras que solamente en dos hay un incremento entre el nivel intermedio y avanzado. Por lo que se refiere a la producción de vocablos, encontramos un aumento entre el nivel básico y el intermedio y una disminución entre el intermedio y el avanzado. Finalmente, hemos detectado la influencia del conocimiento de otras lenguas romances ya que estos estudiantes han producido un mayor número de palabras y de vocablos con respecto a sus compañeros que no conocen otras lenguas romances.

En el análisis cuantitativo nos hemos centrado en la presentación de las palabras más disponibles de cada centro de interés, y como esperábamos las palabras en los rangos más altos corresponden a las más prototípicas. Todas las palabras que componen nuestro corpus son sustantivos, y solo registramos algunos casos de palabras compuestas o locuciones. Contrariamente a lo que suele ocurrir en los estudios de disponibilidad léxica, nosotros hemos elegido mantener las formas flexivas distintas para evaluar cuál es la más disponible. De los datos generales, cuando actualizadas ambas, la forma plural alcanza un índice de disponibilidad mayor en el centro de interés 01. Partes del cuerpo, mientras que en los otros dos centros la forma singular es más disponible. Por lo que se refiere al análisis cualitativo del corpus de palabras entre las cuatro variables, es posible notar la supremacía en las palabras que alcanzan un índice de disponibilidad $\geq 0,1$ de los grupos de las mujeres frente a los hombres, de los estudiantes de lenguas extranjeras frente a los demás y de los estudiantes de nivel avanzado en todos los centros de interés, mientras que en el caso de los estudiantes que conocen también otras lenguas romances, ellos priman en dos centros frente a sus compañeros que conocen otras lenguas pero ninguna romance.

El último análisis que llevamos a cabo, ha tenido como objeto las relaciones léxico-semánticas que pudimos encontrar entre los pares léxicos. Después de una breve introducción

teórica hemos presentado los pares que alcanzan más de dos repeticiones en los tres centros de interés. De este primer análisis es posible apreciar como la distinta tipología de centro de interés influye en las relaciones que se establecen. De hecho, en el centro 01. Partes del cuerpo que definimos como bien definido e inclusivo, puesto que tiene límites claros de lo que pertenece o no a él y tiene asociaciones de tipo “X es Y”, las relaciones que prevalecen son de contigüidad y de meronimia; en el centro 05. Alimentos y bebidas, que es una categoría natural organizada entorno a un prototipo e inclusiva, las palabras actualizadas en muchos casos son prototípicas y generalmente se pueden asociar como cohipónimos de alimentos o de bebidas; el centro 10. La ciudad, centro esquema y relacional, para el cual los informantes evocan escenas reales y la asociación es “X está relacionado con Y”, las relaciones que predominan son de tipo contextual y dependen de la experiencia personal. En la segunda parte del análisis nos centramos en los pares léxicos producidos por los estudiantes de los tres niveles de conocimiento de español. Generalmente pudimos destacar que el nivel influye en la organización del léxico puesto que los estudiantes de nivel avanzado han producido más pares léxicos que sus compañeros de niveles inferiores en los centros 01. Partes del cuerpo y 05. Alimentos y bebidas. Solamente en el centro 10. La ciudad, los estudiantes de nivel intermedio superan los de nivel avanzado. Por lo que se refiere al tipo de relaciones individualizables, no encontramos muchas diferencias puesto que la mayoría de las relaciones fonológicas tenían también una explicación semántica. Todavía, reconocemos que estos resultados no se puedan generalizar dado el número limitado de informantes que componen nuestra muestra.

En conclusión, consideramos que nuestra investigación ha cumplido los dos objetivos que nos propusimos al principio: presentar nuevos datos sobre el léxico disponible de estudiantes italianos de ELE en contexto inmersivo y analizar las relaciones léxico-semánticas del corpus de palabras obtenidas. No obstante, reconocemos que serán necesarios ulteriores estudios con muestras más numerosas para poder alcanzar una visión más exhaustiva del léxico disponible de los aprendices italianos de ELE y comprobar las conclusiones que hemos sacado de este primer acercamiento.

REFERENCIAS BIBLIOGRÁFICAS

- Alba, O. (1995). *Léxico disponible de la República Dominicana*. Pontificia Universidad Católica Madre y Maestra.
- Alba, O. (2000). Densidad de anglicismos en el léxico disponible. En *Nuevos aspectos del español en Santo Domingo* Librería La Trinitaria y Brigham Young University. pp. 73-98.
- Álvarez, C. J., Alameda, R., & Domínguez, A. (1999). El reconocimiento de las palabras. Procesamiento ortográfico y silábico. En M. De Vega & F. Cuetos (Eds.), *Psicolingüística del español*. Editorial Trotta. pp. 89-130.
- Areta Lara, M. (2008). *Análisis sociolingüístico de la disponibilidad léxica de los estudiantes preuniversitarios de Navarra*. Universidad de Navarra.
- Arnal, M. L. (2004). *El léxico disponible de Aragón*. Libros Pórtico.
- Ayora Esteban, C. (2006). *Disponibilidad léxica en Ceuta: aspectos sociolingüísticos*. Universidad de Cádiz.
- Azurmendi Ayerbe, M. J. (1983). *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial en la comarca de San Sebastián*. Ediciones de la Caja de Ahorros Provincial de Guipúzcoa.
- Bailey Victory, J. (1971). *A study of lexical availability among monolingual-bilingual speakers of Spanish and English*. Rice University.
- Bartol Hernández, J. A. (2004). *Léxico disponible de Soria. Estudio y diccionarios*. Instituto Castellano y Leonés de la Lengua.
- Bartol Hernández, J. A. (2006). La disponibilidad léxica. *Revista Española de Lingüística*, 36. pp. 379-384.
- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En R. M. Castañer Martín & V. Lagüéns Gracia (Eds.), *De moneda nunca usada. Estudios filológicos dedicados a José M. Enguita Utrilla*. Institución Fernando El Católico. pp. 85-107.
- Benítez Pérez, P. (1992). Disponibilidad léxica en la zona metropolitana de Madrid. *Boletín de la Academia Puertorriqueña de la Lengua Española*, 1, 1. pp. 71-102.
- Benítez Pérez, P. (1994). Léxico real/irreal en los manuales de español para extranjeros. En S. Montesa Peydró & A. Garrido Moraga (Eds.), *Actas del Segundo Congreso Nacional de ASELE. Español para extranjeros: didáctica e investigación*. ASELE. pp. 325-333.
- Benítez Pérez, P., & Zebrowski, J. (1993). El léxico español en los manuales polacos. En S. Montesa Peydró & A. Garrido Moraga (Eds.), *Actas del III Congreso Nacional de la Asociación para la enseñanza del español como lengua extranjera (ASELE) "El español como lengua extranjera: de la teoría al aula"*. ASELE. pp. 223-230.
- Buchanan, M. A. (1927). *A Graded Spanish Word Book*. University Press.

- Caggiula, S. (2013). *El español como lengua extranjera: un estudio de disponibilidad léxica y su aplicación a la enseñanza*. Lulu Press.
- Cañizal Arévalo, A. (1987). *Disponibilidad léxica en escolares de primaria terminada. Análisis de seis centros de interés* [Memoria de licenciatura inédita]. Universidad Autónoma de México.
- Cañizal Arévalo, A. (1991). Redes semánticas y disponibilidad léxica en el español de escolares mexicanos. En C., Hernández Alonso, G. P., Granda, C., Hoyos, V., Fernández, D. Dietrick, & Y. Carballera (Eds.), *El español de América. Actas del III Congreso Internacional de "El español de América"*. Junta de Castilla y León. pp. 631-641
- Carcedo González, A. (1998a). Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE. *RILCE: Revista de Filología Hispánica*, 14.2, pp. 205-224.
- Carcedo González, A. (1998b). Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica. *Lingüística*, 10, pp. 5-68.
- Carcedo González, A. (1999). Análisis de errores léxicos del español en la interlingua de los finlandeses. En M. C., Losada Aldrey, J. F. Márquez Caneda, & T. E. Jiménez Juliá (Eds.), *Español como lengua extranjera: enfoque comunicativo y gramática: actas del IX Congreso Internacional de ASELE, Santiago de Compostela, 23- 26 de septiembre de 1998*. Universidad de Santiago de Compostela. pp. 465-472.
- 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)*. Turun Yliopisto.
- Carcedo González, A. (2001). *Léxico disponible de Asturias*. Universidad de Turku.
- Carcedo González, A. (2003). Unidad y variedad diatópica de la disponibilidad léxica del español: comparación de los inventarios de Puerto Rico, Cádiz y Asturias. En F., Moreno Fernández, J. A., Samper Padilla, M., Vaquero, M. L. Gutiérrez Araus, & F. Gimeno Menéndez (Eds.), *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales: Vol. I*. Arco/Libros. pp. 199-225.
- Casanova Ávalos, M. (2017). *Léxico disponible de Castellón. Estudio y diccionarios* [Tesis doctoral]. Universitat Jaume I.
- Cruz Alonso, R. (2012). *El léxico disponible de Castilla y León* [Tesis doctoral]. Universidad de Salamanca.
- Delgado Fernández, R. (2024). *Edad y disponibilidad léxica. La variación social en el léxico disponible adulto de Salamanca* [Tesis doctoral]. Universidad de Salamanca.
- Del Barrio de la Rosa, F. (2016). Algunas observaciones sobre la disponibilidad léxica en estudiantes italofonos de español. En E. Sainz González, I. Solís García, F. Del Barrio de la Rosa, & I. Arroyo Hernández (Eds.), *Geométrica explosión. Estudios de lengua y literatura en homenaje a René Lenarduzzi*. Edizioni Ca' Foscari. pp. 127-143.

- Del Barrio de la Rosa, F. (2017). Los estudiantes italo-fonos de ELE y los estudios de disponibilidad léxica en español. *Revista Nebrija de Lingüística Aplicada*, 22, pp. 52-57.
- Del Barrio de la Rosa, F. (2018). Pares léxicos en el léxico disponible de estudiantes italianos. *Lingue e linguaggi*, 26, pp. 173-196.
- Del Barrio de la Rosa, F., & Mae Vann, M. (2018). *Disponibilidad léxica de los estudiantes de español en Italia. Estudio y diccionarios*. Aracne.
- Dimitrijević, N. (1969). *Lexical Availability. A new aspect of the lexical availability of secondary school children*. Julius Gross Verlag.
- Echeverría, M. S., Herrera, M. O., Moreno, P., & Pradenas, F. (1987). Disponibilidad léxica en Educación Media. *Revista de Lingüística Teórica y Aplicada*, 25, pp. 55-115.
- Etxebarria Arostegui, M. (1996). Disponibilidad léxica en escolares del País Vasco: variación sociolingüística y modelos de enseñanza bilingüe. *Revista Española de Lingüística*, 26, 2, pp. 301-325.
- Fernández Juncal, C. (2008). *Léxico disponible de Burgos*. Fundación Instituto Castellano y Leonés de la Lengua.
- Fernández Juncal, C. (2013). *El léxico disponible de Cantabria. Estudio sociolingüístico*. Editorial Universidad de Cantabria.
- Fernández Smith, G., Rico Martín, A. M., Molina García, M. ^a J., & Jiménez Jiménez, M. Á. (2008). *Léxico disponible de Melilla: estudio sociolingüístico y repertorios léxicos*. Arco/Libros.
- Galoso Camacho, M. V. (2003). *El léxico disponible de Ávila, Salamanca y Zamora*. Instituto Castellano y Leonés de la Lengua.
- Gómez Devis, M. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos* [Tesis doctoral]. Universitat de València.
- Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1964). *L'élaboration du Français Fondamental (1er degré)*. Dider.
- Guerra Salas, L., Paredes García, F., & Gómez Sánchez, E. (2004). Disponibilidad léxica en la Comunidad de Madrid. *XXXIV Simposio de la Sociedad Española de Lingüística*.
- Henmon, V. A. C. (1924). *A French Word Book Based on a Count of 400,000 Running Words*. Bureau of Educational Research. University of Wisconsin.
- Hernández Cabrera, C. E., & Samper Hernández, M. (2001). Léxico disponible y norma culta. En A. Valencia (Ed.), *Actas del Congreso Internacional 'El español culto en el mundo hispánico'*. Universidad Bolivariana. pp. 79-89.
- Hernández Muñoz, N. (2004). *El léxico disponible de los estudiantes conquenses*. Universidad de Salamanca.

- Hernández Muñoz, N. (2005a). *Hacia una teoría cognitiva integrada de la disponibilidad léxica. El léxico de los estudiantes castellano-manchegos* [Tesis doctoral]. Universidad de Salamanca.
- Hernández Muñoz, N. (2005b). La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología. En M. A. Mayor, B. Zubianz, & E. Díez (Eds.), *Estudios sobre adquisición del lenguaje* (pp. 942-954). Ediciones Universidad de Salamanca.
- Hernández Muñoz, N., & Tomé Cornejo C. (2017). Léxico disponible en primera y segunda lengua: bases cognitivas. En F. Del Barrio de la Rosa (Ed.), *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía*. Edizioni Ca' Foscari. pp. 99-122.
- Hernández Muñoz, N., Tomé Cornejo, C., López García, M. & Bartol Hernández, J. A. (2023): *Manual de LexPro*. <https://dispogram.usal.es/> [Consulta: abril 2024]
- Hoz García, M. V. (1952). *Vocabulario usual, común y fundamental*. C.S.I.S. Instituto San José de Calasanz.
- Hugo, N. M. (2003). *Disponibilidad léxica en español como lengua extranjera* [Trabajo de investigación inédito]. Universität des Saarlandes.
- Izquierdo Gil, M. C. (2003). *La selección de léxico en la enseñanza del español como lengua extranjera: su aplicación al nivel elemental en estudiantes francófonos*. Universitat de Valencia.
- Jiménez Berrio, F. (2013). *Léxico disponible de inmigrantes escolares no hispanohablantes*. Servicio de Publicaciones de la Universidad de Navarra.
- Käding, J. W. (1897). *Häufigkeitwörterbuch der deutschen Sprache*. Steglitz bei Berlin. Der Herausgeber.
- Llopis Rodrigo, F. (2008). *El lèxic del valencià central: un estudi de disponibilitat léxica*. Universidad de Valencia.
- Llopis Rodrigo, F., & Gómez Devís, M. B. (2010). *El lèxic disponible de València*. Denes.
- López Chávez, J. (1992). Alcances panhispánicos del léxico disponible. *Lingüística*, 4, pp. 26-124.
- López Chávez, J. (1994). Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica. *Revista de Estudios de Adquisición de la Lengua Española*, 1, pp. 67-84.
- López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Arco/Libros.
- López Morales, H., & García Marcos, F. J. (1995). Disponibilidad léxica en Andalucía. Proyecto de investigación. *Revista de Estudios de Adquisición de la Lengua Española*, 3, pp. 65-76.
- Mackey, W. C. (1971). *Le vocabulaire disponible du français: Vol. 2*. Didier.

- Martínez Olmos, E. (2007). *Disponibilidad léxica en las comunidades de habla alicantinas* [Tesis doctoral]. Universidad de Alicante.
- Mateo García, M. V. (1994). Disponibilidad léxica: posibles aplicaciones. En J. D. Luque & A. Pamies (Eds.), *Actas de las Primeras Jornadas sobre estudio y enseñanza del léxico*. Universidad de Granada, pp. 141-147.
- Meirama López, B. (2008). *Léxico disponible en español de Galicia*. Universidade de Santiago de Compostela.
- Mena Osorio, M. (1986). *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Universidad de Concepción.
- Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les Langues Modernes*, 47, pp. 338-344.
- Murillo Rojas, M. (1993). Disponibilidad léxica en los preescolares: estudio de cinco campos semánticos. *Káñina (Universidad de Costa Rica)*, 17/2, pp. 117-127.
- Murillo Rojas, M. (1994). Comidas y bebidas: estudio de la disponibilidad léxica en preescolares. *Káñina (Universidad de Costa Rica)*, 18/2, pp. 117-133.
- Nalesso, G. (2018). El desarrollo de la competencia léxica de estudiantes italianos universitarios de ELE. *Orillas: Rivista d'ispanistica*, 7, pp. 381-394.
- Nalesso, G. (2019). *Disponibilidad y riqueza léxica en un grupo de estudiantes universitarios italianos de ELE*. Università degli studi di Padova.
- Nalesso, G. (2020). Disponibilidad léxica terminológica en ELE: una propuesta de análisis. *Orillas: Rivista d'ispanistica*, 9, pp. 609-631.
- Njock, P. E. (1979). *L'univers familier de l'enfant africain*. CIRB. Université Laval.
- 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.
- Pérez Jiménez, E. (2016). *El léxico disponible de los estudiantes preuniversitarios en La Rioja* [Tesis doctoral]. Universidad de La Rioja.
- Prado Aragonés, J., & Galloso Camacho, M. V. (2015). *El léxico disponible de Extremadura y comparación con el de Andalucía*. Universidad de Huelva.
- Prado Aragonés, J., Galloso Camacho, M. V., & Conceição, M. C. (2009). *La disponibilidad léxica en situación de contacto de lenguas en las zonas limítrofes de Andalucía y Extremadura (España) y Algarve y Alentejo (Portugal)*. Universidad de Huelva.
- Rodríguez Bou, I. (1952). *Recuento de Vocabulario español*. Universidad de Puerto Rico.
- Rubio Sánchez, R. (2015). Estudio de disponibilidad léxica en aprendices italianos de español: análisis cuantitativo. En A. Gordejuela Senosiáin, D. Izquierdo Alegría, F. Jiménez Berrio, A. de Lucas Vicente, & M. Casado Velarde (Eds.), *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General*. Servicio de Publicaciones de la Universidad de Navarra. pp. 429-441.

- Rubio Sánchez, R. (2017). Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de Español como Lengua Extranjera. En F. del Barrio de la Rosa (Ed.), *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía*. Edizioni Ca' Foscari. pp. 143-161.
- Samper Hernández, M. (2001). Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de selección. En J. A. Bartol, S. Crespo, C. Fernández Juncal, C. Pensado, E. Prieto, & N. Sánchez (Eds.), *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*. Luso-Española de Ediciones. pp. 277-286.
- Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. ASELE.
- Samper Hernández, M. (2008). Datos comparativos entre dos léxicos disponibles: Aragón y Gran Canaria. En M. L. Arnal Purroy (Ed.), *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*. Institución Fernando el Católico. pp. 245-282.
- Samper Padilla, J. A. (1998). Criterios de edición del léxico disponible. *Lingüística*, 10, pp. 311-333.
- Samper Padilla, J. A. (1999). Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria. En E., Forastieri Braschi, J., Cardona, H. López Morales, & A. Morales de Walters (Eds.), *Estudios de Lingüística Hispánica (Homenaje a María Vaquero)*. Editorial de la Universidad de Puerto Rico. pp. 551-573.
- Samper Padilla, J. A., Hernández Cabrera, C. E., & Bellón, J. J. (2003). Léxico disponible y variación dialectal: datos de Canarias y Córdoba. En *Actas del I Congreso Internacional sobre el español de Canarias.: Vol. II*. Academia Canaria de la Lengua. pp. 1077-1099.
- Samper Padilla, J. A., & Hernández Cabrera, C. E. (1997). El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística. En J. Dorta & M. Almeida (Eds.), *Contribuciones al estudio de la lingüística hispánica. Homenaje al profesor Ramón Trujillo*. Montesinos. pp. 229-239.
- Sánchez Corrales, M., & Murillo Rojas, V. M. (2006). *Disponibilidad léxica de los niños preescolares costarricenses*. Universidad de Costa Rica.
- Sánchez-Saus Laserna, M. (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas* [Tesis doctoral]. Universidad de Cádiz.
- Sánchez-Saus Laserna, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas*. Universidad de Sevilla.
- Sánchez-Saus Laserna, M. (2022). Redes semánticas, léxico disponible y didáctica del vocabulario en ELE: un análisis por niveles de español. *Tejuelo*, 35.3, pp. 167-204.
- Santiago Guervós, J. de. (2008). *Léxico disponible de Segovia*. Fundación Instituto Castellano y Leonés de la Lengua.
- Santos Díaz, I. C. (2015). La influencia del conocimiento de varias lenguas extranjeras en el léxico disponible. En A. Gordejuela Senosiáin, D. Izquierdo Alegría, F. Jiménez Berrio,

- A. De Lucas Vicente, & M. Casado Velarde (Eds.), *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General*. Servicio de Publicaciones de la Universidad de Navarra. pp. 447-484.
- Serfati, M. (2017). *Cálculo de disponibilidad léxica en ELE de universitarios marroquíes de Agadir: el programa LexiDisp*.
- Serrano Zapata, M. (2014). *Disponibilidad léxica en la provincia de Lleida: estudio comparado de dos lenguas en contacto* [Tesis doctoral]. Universitat de Lleida.
- Taberner, C. (2008). El léxico disponible como fuente de aproximación al estudio de los regionalismos. En I., Olza Moreno, M. Casado Velarde, & R. González Ruiz (Eds.), *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL)*. Universidad de Navarra. pp. 811-824.
- Thorndike, E. L. (1921). *The teacher's word book*. Teachers College Press.
- Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. Universidad de Salamanca.
- Tomé Cornejo, C., & Recio Diego, Á. (2022). De la disponibilidad léxica a la disponibilidad gramatical. Desarrollos metodológicos y aplicación didáctica. *Tejuelo*, 35.3, pp. 299-336.
- Valencia, A., & Echeverría W., M. S. (1999). *El léxico disponible en estudiantes chilenos*. Ediciones Universidad de Chile.

ANEXO I – MODELO PRUEBA DE DISPONIBILIDAD Y CUESTIONARIO SOCIOLÓGICO

Escribe todas las palabras que se te ocurran durante 2 minutos sobre el siguiente tema: ALIMENTOS Y BEBIDAS

1.	51.
2.	52.
3.	53.
4.	54.
5.	55.
6.	56.
7.	57.
8.	58.
9.	59.
10.	60.
11.	61.
12.	62.
13.	63.
14.	64.
15.	65.
16.	66.
17.	67.
18.	68.
19.	69.
20.	70.
21.	71.
22.	72.
23.	73.
24.	74.
25.	75.
26.	76.
27.	77.
28.	78.
29.	79.
30.	80.
31.	81.
32.	82.
33.	83.
34.	84.
35.	85.
36.	86.
37.	87.
38.	88.
39.	89.
40.	90.
41.	91.
42.	92.
43.	93.
44.	94.
45.	95.
46.	96.
47.	97.
48.	98.
49.	99.
50.	100.

Escribe todas las palabras que se te ocurran durante 2 minutos sobre el siguiente tema: LAS PARTES DEL CUERPO

1.	51.
2.	52.
3.	53.
4.	54.
5.	55.
6.	56.
7.	57.
8.	58.
9.	59.
10.	60.
11.	61.
12.	62.
13.	63.
14.	64.
15.	65.
16.	66.
17.	67.
18.	68.
19.	69.
20.	70.
21.	71.
22.	72.
23.	73.
24.	74.
25.	75.
26.	76.
27.	77.
28.	78.
29.	79.
30.	80.
31.	81.
32.	82.
33.	83.
34.	84.
35.	85.
36.	86.
37.	87.
38.	88.
39.	89.
40.	90.
41.	91.
42.	92.
43.	93.
44.	94.
45.	95.
46.	96.
47.	97.
48.	98.
49.	99.
50.	100.

Escribe todas las palabras que se te ocurran durante 2 minutos sobre el siguiente tema: LA CIUDAD

1.	51.
2.	52.
3.	53.
4.	54.
5.	55.
6.	56.
7.	57.
8.	58.
9.	59.
10.	60.
11.	61.
12.	62.
13.	63.
14.	64.
15.	65.
16.	66.
17.	67.
18.	68.
19.	69.
20.	70.
21.	71.
22.	72.
23.	73.
24.	74.
25.	75.
26.	76.
27.	77.
28.	78.
29.	79.
30.	80.
31.	81.
32.	82.
33.	83.
34.	84.
35.	85.
36.	86.
37.	87.
38.	88.
39.	89.
40.	90.
41.	91.
42.	92.
43.	93.
44.	94.
45.	95.
46.	96.
47.	97.
48.	98.
49.	99.
50.	100.

CUESTIONARIO SOCIOLOGIO

1. Edad: _____

2. Sexo:

<input type="checkbox"/>	Hombre
<input type="checkbox"/>	Mujer

3. ¿Cuál es tu carrera? _____

4. Además del italiano, ¿conoces otras lenguas?

<input type="checkbox"/>	Sí
<input type="checkbox"/>	No

 ¿Cuáles? _____

5. ¿Qué lenguas hablas en familia? Italiano Otras:

6. ¿Qué lenguas hablas aquí en casa? Italiano Español Otras:

7. ¿Desde cuándo estás en España? _____

8. ¿Has estudiado español antes de llegar aquí?

<input type="checkbox"/>	Sí
<input type="checkbox"/>	No

 ¿Dónde? _____
¿Cuánto tiempo? _____

9. ¿Has estado en España antes?

<input type="checkbox"/>	Sí
<input type="checkbox"/>	No

 ¿En qué año? _____
¿Cuánto tiempo? _____
¿Dónde? _____

10. ¿Has visitado otros países hispanohablantes?

<input type="checkbox"/>	Sí
<input type="checkbox"/>	No

 ¿En qué año? _____
¿Por cuánto tiempo? _____
¿Cuáles? _____

11. ¿Estás haciendo o has hecho un curso en la Casa de las Lenguas?

<input type="checkbox"/>	Sí
<input type="checkbox"/>	No

 ¿De qué nivel? _____

12. ¿Tienes una certificación DELE?

<input type="checkbox"/>	Sí
<input type="checkbox"/>	No

 ¿De qué nivel? _____

13. ¿Con que frecuencia tienes contacto con medios de comunicación, literatura, música, etc. en español ahora que estás en España?

	Todos los días	Varias veces por semana	Una vez por semana	Con menor frecuencia	Nunca
Televisión					
Radio					
Podcast					
Periódicos o revistas					
Social network					
Libros					
Música					
Películas/series					

14. ¿Cómo ves las películas/series?

<input type="checkbox"/>	Dobladas en italiano
<input type="checkbox"/>	Dobladas en español
<input type="checkbox"/>	En lengua original (español) y con subtítulos en italiano
<input type="checkbox"/>	En lengua original (español) y con subtítulos en español
<input type="checkbox"/>	En lengua original (español) SIN subtítulos
<input type="checkbox"/>	Otros:

¿Cuánto estás de acuerdo con estas afirmaciones?

1. El español me parece fácil de aprender

1	Muy de acuerdo	2	De acuerdo	3	Poco de acuerdo	4	En desacuerdo
---	----------------	---	------------	---	-----------------	---	---------------

2. Aprender español en España me parece más fácil

1	Muy de acuerdo	2	De acuerdo	3	Poco de acuerdo	4	En desacuerdo
---	----------------	---	------------	---	-----------------	---	---------------

3. El mejor español se habla en España

1	Muy de acuerdo	2	De acuerdo	3	Poco de acuerdo	4	En desacuerdo
---	----------------	---	------------	---	-----------------	---	---------------

4. Mis conocimientos previos de español eran suficientes para vivir aquí

1	Muy de acuerdo	2	De acuerdo	3	Poco de acuerdo	4	En desacuerdo
---	----------------	---	------------	---	-----------------	---	---------------

5. Mis conocimientos previos de español eran suficientes para entender las clases

1	Muy de acuerdo	2	De acuerdo	3	Poco de acuerdo	4	En desacuerdo
---	----------------	---	------------	---	-----------------	---	---------------

ANEXO II – DICCIONARIO DE DISPONIBILIDAD

01. PARTES DEL CUERPO

Nº	Token	Disponibilidad	Frecuencia relativa	% aparición	Frecuencia acumulada
1	ojos	0,622486	0,04661	0,785714	0,04661
2	cabeza	0,552789	0,042373	0,714286	0,088983
3	dedos	0,435316	0,038136	0,642857	0,127119
4	piernas	0,426121	0,038136	0,642857	0,165254
5	pelo	0,41932	0,038136	0,642857	0,20339
6	pies	0,388433	0,042373	0,714286	0,245763
7	manos	0,356616	0,025424	0,428571	0,271186
8	cuello	0,33123	0,029661	0,5	0,300847
9	nariz	0,32594	0,038136	0,642857	0,338983
10	mano	0,305233	0,021186	0,357143	0,360169
11	boca	0,290203	0,029661	0,5	0,389831
12	brazos	0,276092	0,025424	0,428571	0,415254
13	corazón	0,203378	0,025424	0,428571	0,440678
14	barriga	0,186177	0,029661	0,5	0,470339
15	espalda	0,172591	0,021186	0,357143	0,491525
16	orejas	0,171095	0,016949	0,285714	0,508475
17	lengua	0,154607	0,016949	0,285714	0,525424
18	uñas	0,145121	0,021186	0,357143	0,54661
19	ojo	0,135891	0,012712	0,214286	0,559322
20	rodillas	0,130616	0,012712	0,214286	0,572034
21	dientes	0,123317	0,016949	0,285714	0,588983
22	muñeca	0,121011	0,012712	0,214286	0,601695
23	labios	0,114571	0,012712	0,214286	0,614407
24	uña	0,112962	0,008475	0,142857	0,622881
25	garganta	0,10789	0,012712	0,214286	0,635593
26	hombros	0,107118	0,012712	0,214286	0,648305
27	pierna	0,100967	0,008475	0,142857	0,65678
28	piel	0,099268	0,012712	0,214286	0,669492
29	oreja	0,097302	0,008475	0,142857	0,677966
30	rodilla	0,097205	0,016949	0,285714	0,694915
31	pestañas	0,086274	0,012712	0,214286	0,707627
32	pecho	0,081948	0,012712	0,214286	0,720339
33	brazo	0,081263	0,008475	0,142857	0,728814
34	cejas	0,068575	0,008475	0,142857	0,737288
35	estómago	0,06612	0,008475	0,142857	0,745763
36	espaldas	0,065982	0,004237	0,071429	0,75
37	culo	0,057791	0,008475	0,142857	0,758475
38	cara	0,056304	0,004237	0,071429	0,762712
39	mente	0,053385	0,008475	0,142857	0,771186
40	barbita	0,048045	0,004237	0,071429	0,775424

41	pulgar	0,048045	0,004237	0,071429	0,779661
42	huesos	0,046555	0,012712	0,214286	0,792373
43	codos	0,044023	0,008475	0,142857	0,800847
44	ombbligo	0,042198	0,008475	0,142857	0,809322
45	cerebro	0,041083	0,008475	0,142857	0,817797
46	codo	0,040998	0,004237	0,071429	0,822034
47	tobillo	0,040096	0,008475	0,142857	0,830508
48	pie	0,034984	0,004237	0,071429	0,834746
49	cintura	0,030727	0,008475	0,142857	0,84322
50	cadera	0,030075	0,008475	0,142857	0,851695
51	mejilla	0,029853	0,004237	0,071429	0,855932
52	trasero	0,029853	0,004237	0,071429	0,860169
53	caderas	0,027577	0,004237	0,071429	0,864407
54	abdomen	0,027577	0,004237	0,071429	0,868644
55	glúteo	0,025474	0,004237	0,071429	0,872881
56	gemelos	0,023532	0,004237	0,071429	0,877119
57	hígado	0,023532	0,004237	0,071429	0,881356
58	pechos	0,021738	0,004237	0,071429	0,885593
59	intestino	0,021738	0,004237	0,071429	0,889831
60	axilas	0,02008	0,004237	0,071429	0,894068
61	pulmón	0,02008	0,004237	0,071429	0,898305
62	muñecas	0,02008	0,004237	0,071429	0,902542
63	bazo	0,018549	0,004237	0,071429	0,90678
64	pancreas	0,017135	0,004237	0,071429	0,911017
65	pelvis	0,017135	0,004237	0,071429	0,915254
66	globo ocular	0,017135	0,004237	0,071429	0,919492
67	costillas	0,017135	0,004237	0,071429	0,923729
68	riñón	0,015828	0,004237	0,071429	0,927966
69	hombro	0,015828	0,004237	0,071429	0,932203
70	venas	0,015828	0,004237	0,071429	0,936441
71	columna vertebral	0,014621	0,004237	0,071429	0,940678
72	vejiga	0,014621	0,004237	0,071429	0,944915
73	barbilla	0,013507	0,004237	0,071429	0,949153
74	talón	0,013507	0,004237	0,071429	0,95339
75	vagina	0,013507	0,004237	0,071429	0,957627
76	nervios	0,012477	0,004237	0,071429	0,961864
77	pubis	0,012477	0,004237	0,071429	0,966102
78	patillas	0,012477	0,004237	0,071429	0,970339
79	arterias	0,011525	0,004237	0,071429	0,974576
80	pelos	0,011525	0,004237	0,071429	0,978814
81	cartílago	0,009835	0,004237	0,071429	0,983051
82	músculos	0,009085	0,004237	0,071429	0,987288
83	palma de la mano	0,009085	0,004237	0,071429	0,991525
84	tórax	0,008392	0,004237	0,071429	0,995763
85	tiroides	0,007752	0,004237	0,071429	1

05. ALIMENTOS Y BEBIDAS

Nº	Token	Disponibilidad	Frecuencia relativa	% aparición	Frecuencia acumulada
1	agua	0,582957	0,043825	0,785714	0,043825
2	leche	0,343344	0,035857	0,642857	0,079681
3	pasta	0,320409	0,027888	0,5	0,10757
4	pollo	0,255224	0,023904	0,428571	0,131474
5	arroz	0,214101	0,023904	0,428571	0,155378
6	cerveza	0,204891	0,015936	0,285714	0,171315
7	cachopo	0,203928	0,015936	0,285714	0,187251
8	tomate	0,203056	0,01992	0,357143	0,207171
9	sidra	0,185461	0,015936	0,285714	0,223108
10	pan	0,176861	0,015936	0,285714	0,239044
11	cebolla	0,175436	0,015936	0,285714	0,25498
12	vino	0,169155	0,01992	0,357143	0,2749
13	manzana	0,168132	0,015936	0,285714	0,290837
14	queso	0,159284	0,01992	0,357143	0,310757
15	jamón	0,157211	0,015936	0,285714	0,326693
16	huevos	0,152979	0,015936	0,285714	0,342629
17	carne	0,150741	0,015936	0,285714	0,358566
18	comida	0,142857	0,007968	0,142857	0,366534
19	chocolate	0,140828	0,01992	0,357143	0,386454
20	piña	0,118919	0,011952	0,214286	0,398406
21	azúcar	0,105743	0,007968	0,142857	0,406375
22	sal	0,099174	0,011952	0,214286	0,418327
23	naranja	0,093992	0,011952	0,214286	0,430279
24	zanahoria	0,088826	0,011952	0,214286	0,442231
25	galletas	0,087054	0,011952	0,214286	0,454183
26	cerdo	0,083294	0,011952	0,214286	0,466135
27	patatas	0,08195	0,011952	0,214286	0,478088
28	café	0,08163	0,007968	0,142857	0,486056
29	palomitas	0,081095	0,007968	0,142857	0,494024
30	melocotón	0,081095	0,007968	0,142857	0,501992
31	hamburguesa	0,079633	0,007968	0,142857	0,50996
32	tarta	0,079293	0,011952	0,214286	0,521912
33	verduras	0,075387	0,007968	0,142857	0,52988
34	calabacín	0,072891	0,007968	0,142857	0,537849
35	bocadillo	0,071429	0,003984	0,071429	0,541833
36	refresco	0,071429	0,003984	0,071429	0,545817
37	fabada	0,069846	0,011952	0,214286	0,557769
38	zum de naranja	0,069005	0,007968	0,142857	0,565737
39	pimientos	0,067487	0,007968	0,142857	0,573705
40	cabracho	0,065381	0,003984	0,071429	0,577689
41	ensalada	0,065381	0,003984	0,071429	0,581673
42	dieta	0,065381	0,003984	0,071429	0,585657

43	pizza	0,062318	0,007968	0,142857	0,593625
44	lasaña	0,059846	0,003984	0,071429	0,59761
45	galleta	0,059846	0,003984	0,071429	0,601594
46	lechuga	0,059846	0,003984	0,071429	0,605578
47	nutrición	0,059846	0,003984	0,071429	0,609562
48	calimochó	0,059846	0,003984	0,071429	0,613546
49	espaguetis	0,059846	0,003984	0,071429	0,61753
50	plátano	0,058872	0,011952	0,214286	0,629482
51	vodka	0,056927	0,007968	0,142857	0,63745
52	churros	0,056485	0,007968	0,142857	0,645418
53	aceite	0,056485	0,007968	0,142857	0,653386
54	pavo	0,055901	0,007968	0,142857	0,661355
55	mantequilla	0,054836	0,007968	0,142857	0,669323
56	chipirones	0,054779	0,003984	0,071429	0,673307
57	bizcocho	0,054779	0,003984	0,071429	0,677291
58	pimiento	0,054779	0,003984	0,071429	0,681275
59	salud	0,054779	0,003984	0,071429	0,685259
60	plátanos	0,054779	0,003984	0,071429	0,689243
61	calabaza	0,054779	0,003984	0,071429	0,693227
62	cervezas	0,050142	0,003984	0,071429	0,697211
63	vino tinto	0,050142	0,003984	0,071429	0,701195
64	mandarina	0,050142	0,003984	0,071429	0,705179
65	calamares	0,050142	0,003984	0,071429	0,709163
66	platos compuestos	0,050142	0,003984	0,071429	0,713147
67	gambas	0,050142	0,003984	0,071429	0,717131
68	patata	0,049564	0,007968	0,142857	0,7251
69	tortilla	0,046344	0,007968	0,142857	0,733068
70	malocotón	0,045897	0,003984	0,071429	0,737052
71	sangría	0,045897	0,003984	0,071429	0,741036
72	zumó	0,042011	0,003984	0,071429	0,74502
73	té caliente	0,042011	0,003984	0,071429	0,749004
74	sopa	0,042011	0,003984	0,071429	0,752988
75	caramelos	0,041566	0,007968	0,142857	0,760956
76	pescado	0,041404	0,007968	0,142857	0,768924
77	uva	0,038494	0,007968	0,142857	0,776892
78	albahaca	0,038454	0,003984	0,071429	0,780876
79	papas	0,038454	0,003984	0,071429	0,784861
80	chorizo	0,038454	0,003984	0,071429	0,788845
81	morcilla	0,038454	0,003984	0,071429	0,792829
82	bebidas alcohólica	0,038454	0,003984	0,071429	0,796813
83	yogur	0,035199	0,003984	0,071429	0,800797
84	vinagre	0,035199	0,003984	0,071429	0,804781
85	bebidas energizante	0,035199	0,003984	0,071429	0,808765
86	Coca-Cola	0,035199	0,003984	0,071429	0,812749
87	bebidas nutrientes	0,032219	0,003984	0,071429	0,816733
88	tinto	0,032219	0,003984	0,071429	0,820717

89	manteca de cerdo	0,029491	0,003984	0,071429	0,824701
90	helado	0,029491	0,003984	0,071429	0,828685
91	bebidas sin azúcar	0,029491	0,003984	0,071429	0,832669
92	tomate frito	0,029491	0,003984	0,071429	0,836653
93	frijoles	0,026994	0,003984	0,071429	0,840637
94	colorantes	0,026994	0,003984	0,071429	0,844622
95	limón	0,026994	0,003984	0,071429	0,848606
96	lomo	0,026994	0,003984	0,071429	0,85259
97	pesto	0,026994	0,003984	0,071429	0,856574
98	aguacate	0,024709	0,003984	0,071429	0,860558
99	macarrones	0,024709	0,003984	0,071429	0,864542
100	limones	0,024709	0,003984	0,071429	0,868526
101	aceitunas	0,024709	0,003984	0,071429	0,87251
102	naranjas	0,024709	0,003984	0,071429	0,876494
103	dulces	0,024709	0,003984	0,071429	0,880478
104	tarta de queso	0,022617	0,003984	0,071429	0,884462
105	zanahorias	0,022617	0,003984	0,071429	0,888446
106	pepitas	0,020702	0,003984	0,071429	0,89243
107	salchichas	0,020702	0,003984	0,071429	0,896414
108	fruta	0,020702	0,003984	0,071429	0,900398
109	cacahuetes	0,018949	0,003984	0,071429	0,904382
110	dorada	0,018949	0,003984	0,071429	0,908367
111	arroz con leche	0,018949	0,003984	0,071429	0,912351
112	bocadillos	0,018949	0,003984	0,071429	0,916335
113	quesitos	0,018949	0,003984	0,071429	0,920319
114	San Martin	0,017345	0,003984	0,071429	0,924303
115	pincho	0,017345	0,003984	0,071429	0,928287
116	croquetas	0,017345	0,003984	0,071429	0,932271
117	manzanas	0,017345	0,003984	0,071429	0,936255
118	cabra	0,017345	0,003984	0,071429	0,940239
119	espinacas	0,015877	0,003984	0,071429	0,944223
120	kiwis	0,015877	0,003984	0,071429	0,948207
121	cabrales	0,015877	0,003984	0,071429	0,952191
122	almendras	0,015877	0,003984	0,071429	0,956175
123	fresa	0,014533	0,003984	0,071429	0,960159
124	pollo asado	0,014533	0,003984	0,071429	0,964143
125	turrón	0,014533	0,003984	0,071429	0,968127
126	tartas	0,014533	0,003984	0,071429	0,972112
127	vacuno	0,013302	0,003984	0,071429	0,976096
128	paella	0,013302	0,003984	0,071429	0,98008
129	pimentón	0,013302	0,003984	0,071429	0,984064
130	ajo	0,011145	0,003984	0,071429	0,988048
131	ternera	0,009338	0,003984	0,071429	0,992032
132	nata	0,008547	0,003984	0,071429	0,996016
133	sandía	0,007824	0,003984	0,071429	1

10.LA CIUDAD

Nº	Token	Disponibilidad	Frecuencia relativa	% aparición	Frecuencia acumulada
1	calle	0,44286	0,040201	0,571429	0,040201
2	parque	0,395487	0,045226	0,642857	0,085427
3	tiendas	0,304969	0,035176	0,5	0,120603
4	casas	0,271211	0,025126	0,357143	0,145729
5	catedral	0,250812	0,020101	0,285714	0,165829
6	plaza	0,217964	0,025126	0,357143	0,190955
7	luces	0,190737	0,015075	0,214286	0,20603
8	calles	0,188975	0,015075	0,214286	0,221106
9	iglesia	0,143158	0,020101	0,285714	0,241206
10	iglesias	0,13606	0,01005	0,142857	0,251256
11	bar	0,126987	0,020101	0,285714	0,271357
12	supermercado	0,123347	0,020101	0,285714	0,291457
13	coches	0,122118	0,015075	0,214286	0,306533
14	universidad	0,112619	0,020101	0,285714	0,326633
15	ciudadanos	0,101804	0,01005	0,142857	0,336683
16	ayuntamiento	0,100696	0,015075	0,214286	0,351759
17	piso	0,100102	0,01005	0,142857	0,361809
18	casa	0,097682	0,01005	0,142857	0,371859
19	parques	0,087521	0,01005	0,142857	0,38191
20	sidrería	0,082524	0,01005	0,142857	0,39196
21	coche	0,076692	0,01005	0,142857	0,40201
22	acera	0,071429	0,005025	0,071429	0,407035
23	población	0,071429	0,005025	0,071429	0,41206
24	carreteras	0,071429	0,005025	0,071429	0,417085
25	edificio	0,071429	0,005025	0,071429	0,422111
26	escuela	0,06803	0,015075	0,214286	0,437186
27	tienda	0,067565	0,01005	0,142857	0,447236
28	supermercados	0,0671	0,01005	0,142857	0,457286
29	centro comercial	0,065478	0,01005	0,142857	0,467337
30	casco antiguo	0,064837	0,01005	0,142857	0,477387
31	puerto	0,064631	0,005025	0,071429	0,482412
32	capital	0,064631	0,005025	0,071429	0,487437
33	semáforos	0,064631	0,005025	0,071429	0,492462
34	carro	0,064631	0,005025	0,071429	0,497487
35	prados	0,064631	0,005025	0,071429	0,502513
36	monumento	0,064631	0,005025	0,071429	0,507538
37	colegio	0,062928	0,015075	0,214286	0,522613
38	teatro	0,061747	0,01005	0,142857	0,532663
39	universidades	0,061747	0,01005	0,142857	0,542714
40	restaurante	0,06083	0,01005	0,142857	0,552764
41	ruido	0,060715	0,01005	0,142857	0,562814
42	estadio	0,058481	0,005025	0,071429	0,567839

43	estatuas	0,058481	0,005025	0,071429	0,572864
44	automóvil	0,058481	0,005025	0,071429	0,577889
45	campus	0,058481	0,005025	0,071429	0,582915
46	economía	0,058481	0,005025	0,071429	0,58794
47	farola	0,058481	0,005025	0,071429	0,592965
48	camino	0,052916	0,005025	0,071429	0,59799
49	comunidades autónomas	0,052916	0,005025	0,071429	0,603015
50	autobús	0,04788	0,005025	0,071429	0,60804
51	ventanas	0,04788	0,005025	0,071429	0,613065
52	negocios	0,04788	0,005025	0,071429	0,61809
53	comida típica	0,04788	0,005025	0,071429	0,623116
54	peluquería	0,04788	0,005025	0,071429	0,628141
55	distritos	0,04788	0,005025	0,071429	0,633166
56	edificios	0,04788	0,005025	0,071429	0,638191
57	pisos	0,04788	0,005025	0,071429	0,643216
58	compras	0,04788	0,005025	0,071429	0,648241
59	historia	0,047553	0,01005	0,142857	0,658291
60	carretera	0,043324	0,005025	0,071429	0,663317
61	manera de moverse	0,043324	0,005025	0,071429	0,668342
62	cafeterías	0,043324	0,005025	0,071429	0,673367
63	esteticista	0,043324	0,005025	0,071429	0,678392
64	colegios	0,039201	0,005025	0,071429	0,683417
65	restaurantes	0,039201	0,005025	0,071429	0,688442
66	gente	0,039201	0,005025	0,071429	0,693467
67	calidad de vida	0,039201	0,005025	0,071429	0,698492
68	taxi	0,038707	0,01005	0,142857	0,708543
69	calidad del aire	0,03547	0,005025	0,071429	0,713568
70	personas	0,03547	0,005025	0,071429	0,718593
71	frío	0,03547	0,005025	0,071429	0,723618
72	paso de cebra	0,03547	0,005025	0,071429	0,728643
73	jardín	0,03547	0,005025	0,071429	0,733668
74	gimnasios	0,03547	0,005025	0,071429	0,738693
75	aparcamiento	0,03547	0,005025	0,071429	0,743719
76	museos	0,035228	0,01005	0,142857	0,753769
77	oficinas	0,033321	0,01005	0,142857	0,763819
78	gimnasio	0,032515	0,01005	0,142857	0,773869
79	ciudad rural	0,032095	0,005025	0,071429	0,778894
80	jardines	0,032095	0,005025	0,071429	0,78392
81	autobuses	0,032095	0,005025	0,071429	0,788945
82	rotonda	0,032095	0,005025	0,071429	0,79397
83	habitación	0,032095	0,005025	0,071429	0,798995
84	bus	0,032095	0,005025	0,071429	0,80402
85	pescadería	0,032095	0,005025	0,071429	0,809045
86	lluvia	0,032095	0,005025	0,071429	0,81407
87	fiestas	0,032095	0,005025	0,071429	0,819095
88	parking	0,032095	0,005025	0,071429	0,824121

89	esquina	0,029041	0,005025	0,071429	0,829146
90	palacios	0,029041	0,005025	0,071429	0,834171
91	aeropuerto	0,029041	0,005025	0,071429	0,839196
92	estación de trenes	0,029041	0,005025	0,071429	0,844221
93	carnicería	0,029041	0,005025	0,071429	0,849246
94	frutería	0,026277	0,005025	0,071429	0,854271
95	ciudadano	0,026277	0,005025	0,071429	0,859296
96	farmacias	0,026277	0,005025	0,071429	0,864322
97	Uber	0,026277	0,005025	0,071429	0,869347
98	centro deportivo	0,026277	0,005025	0,071429	0,874372
99	semáforo	0,023777	0,005025	0,071429	0,879397
100	hospitales	0,023777	0,005025	0,071429	0,884422
101	discotecas	0,023777	0,005025	0,071429	0,889447
102	pasos de cebra	0,021514	0,005025	0,071429	0,894472
103	centros de salud	0,021514	0,005025	0,071429	0,899497
104	Fontán	0,021514	0,005025	0,071429	0,904523
105	flores	0,019467	0,005025	0,071429	0,909548
106	pub	0,019467	0,005025	0,071429	0,914573
107	centros de ocio	0,019467	0,005025	0,071429	0,919598
108	andén	0,019467	0,005025	0,071429	0,924623
109	casco histórico	0,019467	0,005025	0,071429	0,929648
110	bibliotecas	0,019467	0,005025	0,071429	0,934673
111	arte	0,017614	0,005025	0,071429	0,939698
112	árbol	0,017614	0,005025	0,071429	0,944724
113	cine	0,015938	0,005025	0,071429	0,949749
114	librería	0,015938	0,005025	0,071429	0,954774
115	luz	0,015938	0,005025	0,071429	0,959799
116	cines	0,015938	0,005025	0,071429	0,964824
117	plazas	0,015938	0,005025	0,071429	0,969849
118	centros comerciales	0,014421	0,005025	0,071429	0,974874
119	monumentos	0,014421	0,005025	0,071429	0,979899
120	mercado	0,014421	0,005025	0,071429	0,984925
121	fuente	0,011807	0,005025	0,071429	0,98995
122	almacén	0,010683	0,005025	0,071429	0,994975
123	talleres	0,010683	0,005025	0,071429	1