

UNIVERZITA PALACKÉHO V OLOMOUCI

Filozofická fakulta

Katedra romanistiky

**Disponibilidad léxica y selección léxica en un manual
checo de ELE**

**Lexical availability and lexical selection in a Czech
SFL textbook**

(Magisterská práce)

Autor: Bc. Monika Husárová

Vedoucí práce: doc. Mgr. Enrique Gutiérrez Rubio, Ph.D.

Olomouc 2024

Prehlasujem, že som túto magisterskú diplomovú prácu vypracovala samostatne pod odborným vedením doc. Mgr. Enrique Gutiérrez Rubio, Ph.D. a uviedla som v nej všetku literatúru a ostatné zdroje, ktoré som použila.

V Olomouci dňa.....

Monika Husárová

Quiero expresar mi sincero agradecimiento a doc. Mgr. Enrique Gutiérrez Rubio, Ph.D., el director de este trabajo de fin de Máster, por su dirección, consejos, comentarios y, sobre todo, por su infinita paciencia. También deseo agradecer a Mgr. Daniela Podhajská, Ph.D. por compartir su experiencia conmigo y ayudarme a orientarme mejor en el tema de mi investigación. Por último, si bien no menos importante, agradezco a los profesores José Antonio Bartol Hernández, Ph.D. y Julio Borrego Nieto, Ph.D. por su disposición a responder mis dudas acerca del tema.

Índice

1	Introducción	6
2	Disponibilidad léxica	7
2.1	Historia y definición de los conceptos clave	7
2.1.1	Léxico disponible frente al frecuente	7
2.1.2	Centros de interés	9
2.2	Las investigaciones del léxico disponible en el ámbito hispano.....	14
2.2.1	Proyecto Panhispánico de Disponibilidad Léxica	14
2.3	Aplicaciones	19
2.3.1	Español como lengua extranjera.....	19
2.3.2	Otras aplicaciones	21
3	La enseñanza y el aprendizaje del vocabulario de la lengua extranjera	22
3.1	Las peculiaridades del vocabulario en la lengua extranjera.....	22
3.2	Los manuales y la selección léxica.....	27
4	Metodología	31
4.1	Objetivo del análisis	31
4.2	Selección de las fuentes.....	31
4.2.1	Selección del manual.....	31
4.2.2	Selección del diccionario de DL	32
4.3	Procedimientos	33
4.3.1	Procedimientos para la obtención de los datos.....	33
4.3.2	Procedimientos del análisis	36
5	Resultados	39
5.1	Presencia de las palabras disponibles en el manual.....	40
5.2	Contenido del glosario.....	43
5.3	La presentación de las palabras disponibles en el manual.....	48
6	Conclusiones	56
7	Resumé	59
8	Referencias bibliográficas	60
9	Anexos	I
9.1	Lista de palabras con el $ID \geq 0,1$: “el cuerpo humano”	I
9.2	Lista de palabras con el $ID \geq 0,1$: “la ropa”	III
9.3	Lista de palabras con el $ID \geq 0,1$: “los animales”	V

Anotación	66
Annotation	67

1 Introducción

Según los datos de Eurostat del año 2022, el 98,8 % de los estudiantes checos en la educación secundaria general estudiaba dos o más lenguas extranjeras. Los mismos datos muestran que el segundo idioma extranjero más estudiado en este tipo de centro en la UE era el español, tan solo precedido por el inglés (2024). Partiendo de estos datos, resulta imprescindible la investigación en el campo de la instrucción de las lenguas extranjeras.

Dicho esto, los programas de la enseñanza de las lenguas extranjeras suelen estar predeterminados por los manuales utilizados en el aula, que dirigen los pasos del programa y, a la vez, proporcionan el material didáctico primario, es decir, en ellos se halla el *input* principal. Por lo tanto, el tema de la elaboración de los manuales es un elemento fundamental en la evaluación de la calidad de la enseñanza de las lenguas. Esta necesidad de avance y desarrollo en este sector dio lugar al empleo de diversas fuentes y corpus. Últimamente, se debate cada vez más sobre la necesidad de incluir *input* auténtico y basado en el lenguaje de los hablantes nativos, por lo cual se ha ido recurriendo cada vez más a las fuentes obtenidas de estudios tales como los listados de las palabras más frecuentes de la lengua. Una herramienta relativamente joven es la disponibilidad léxica, que pretende, aparte de otros objetivos no menos importantes, producir listas de vocabulario empleado por los hablantes nativos relacionado con la situación o el contexto comunicativos.

Así, este trabajo busca determinar si el léxico de un manual originado y de uso frecuente en la República Checa, *Nueva Aventura*, coincide con el presentado en los estudios de la disponibilidad léxica y en qué medida es así. Como referencia, se consulta el diccionario de léxico disponible presente en la investigación *El léxico disponible de Castilla y León* (Cruz Alonso, 2016). Los resultados de este trabajo abarcan tanto una dimensión cuantitativa como la cualitativa y se toma en cuenta la medida de coincidencia, la aparición de las unidades en el glosario del manual y su número de ocurrencias.

Los resultados proporcionados por este trabajo aportan nueva luz respecto al reflejo del léxico real de los hablantes nativos de español en este manual y, por lo tanto, en la enseñanza de español como lengua extranjera, especialmente en la República Checa. Además, el trabajo puede servir de guía a los profesores de este país a la hora de analizar el manual utilizado con el fin de compensar sus posibles deficiencias.

2 Disponibilidad léxica

2.1 Historia y definición de los conceptos clave

2.1.1 Léxico disponible frente al frecuente

Partiendo simplemente del nombre del concepto, “disponibilidad léxica” (en adelante, también DL) o “léxico disponible”, averiguamos que se trata de las palabras que están a disposición en la mente del hablante para ser usados en un momento dado. Pero para entenderlo mejor como un concepto de importancia científica (o pedagógica), recurramos primero a su historia y a sus antecedentes.

No es un concepto nuevo, de hecho, López Morales afirma que surgió en Francia en los años 50, cuando Michéa definió las llamadas “palabras disponibles” en contraste a las “palabras frecuentes”, lo que se debía a la ineficacia del uso del léxico frecuente en la enseñanza del francés a los extranjeros causada por la ausencia de vocablos cotidianos y familiares (1995, p. 245). Para ejemplificar esta falta, consultemos los lemas en el listado de frecuencias del Corpus de Referencia del Español Actual (CREA)¹: en las posiciones más altas, omitiendo los signos de puntuación, encontramos sobre todo los determinantes, preposiciones, pronombres personales y relativos, más abajo ya aparecen verbos (destacando los verbos poco específicos como “ir” [28], “tener” [31] o “estar” [30]), sustantivos (por ejemplo, “tiempo” [90] o “parte” [84]) y otros tipos de palabras (“ahora” [102], “cada” [110]) (Real Academia Española, CREA). En general, las posiciones más altas están ocupadas por las palabras menos específicas, más versátiles y que menos significado aportan, es decir, son las que más bien *modifican* o *añaden* el significado. Es lógico que las palabras de significado más específico aparezcan en posiciones más bajas y es indiscutible la importancia de las palabras más frecuentes a la hora de estudiar el idioma, pero es cuestionable si una lista como esta representa una buena muestra para el corpus en el que se base la enseñanza/el aprendizaje de un idioma². Samper Hernández menciona otra causa de una posible distorsión de los datos en los listados de frecuencias, la que surge del carácter de su fuente: parten de los textos escritos y, además, no se presta atención a las importantes diferencias entre los detalles de los

¹ CREA, <https://www.rae.es>. Fecha de la consulta: 29/1/2024.

² Dicho esto, esta conclusión ha sido hecha tras una observación breve y poco metódica. Además, se trata de unos datos brutos, basados en un corpus, los cuales es posible elaborar – por ejemplo, hay trabajos cuyo objetivo es presentar las palabras más frecuentes a los estudiantes del español como lengua extranjera y que proporcionan datos ya procesados para el objetivo del aprendizaje (como Davies y Hayward Davies, 2018).

textos, como los años de emisión o las procedencias distintas de los textos (2002, p. 7). Es importante avisar que Samper Hernández se refiere a los inconvenientes que tenían las listas de frecuencias en la época en que surgieron los trabajos de la disponibilidad léxica y que la ejecución de las listas de frecuencias también ha sufrido cambios y avances. Efectivamente, el CREA tiene “carácter natural” y los datos en las listas de frecuencias están extraídos de los textos “producidos en condiciones normales de comunicación entre seres humanos: novelas, biografías, ensayos, noticias de periódico, conversaciones informales, discursos en un parlamento, redacciones de estudiantes de secundaria, etc.” y la búsqueda se puede llevar a cabo de manera selectiva (sobre la base del origen, la fecha, el área temática, etc.) (Real Academia Española, CREA)³, pero en las listas de lemas frecuentes aparece todo tipo de datos (es decir, procedentes de todo tipo de textos) y es poco sorprendente que palabras como “gobierno” [101], “presidente” [119] y “político” [127], abundantes en la prensa, aparezcan en posiciones relativamente altas: para dar un poco de contexto, junto a aquellas se posicionan palabras como “después” [100], “siempre” [118] y “sino” [126], es decir, palabras con significado poco concreto, usadas en contextos muy variados (Real Academia Española, CREA).

Dicho esto, López Morales describe la situación en la que se encontraba la lingüística en Francia después de la observación de Michéa: resultaba que en la lengua había una gran cantidad de unidades léxicas que, por tener un significado muy concreto, no se actualizaban a menos que se produjera una situación comunicativa concreta, es decir, solo aparecen si el emisor pretende comunicar una información concreta. Por lo tanto, la frecuencia léxica no puede ser el mérito principal en la determinación de estas unidades (1995, p. 245). La singularidad del léxico disponible radica precisamente en esta particularidad. Samper Hernández formula una definición parecida y añade: “Una palabra es disponible si acude fácil y rápidamente a la mente del hablante cuando se trata un determinado tema” (2002, p. 8). Para resumir la definición de este concepto, el léxico disponible consta de palabras de significado muy concreto utilizadas tras cierto estímulo (en forma de una situación comunicativa dada).

A pesar de lo fácil que era investigar el léxico frecuente (López Morales, 2005, p. 141), considerado hasta entonces el núcleo del vocabulario para la enseñanza de una segunda lengua, faltaba una gran porción de los vocablos igualmente valiosos. Samper Hernández

³ Real Academia Española. *CREA*. <https://www.rae.es/crea-annotado/contenidos/descripcion>. Fecha de la consulta: 24/1/2024.

afirma que en las posiciones altas de las listas de frecuencias aparecen primero las unidades gramaticales, más abajo las siguen los verbos, adjetivos y adverbios (2002, p. 8), como he demostrado parcialmente en el ejemplo del CREA. Esta infrecuencia se debe a que cierta parte del léxico de una lengua consta de las palabras cuyo uso está condicionado por el tema del discurso (y esta peculiaridad se atribuye principalmente a los sustantivos), lo que está en contraste con las palabras que se emplean sin que importe el tema del discurso (López Morales, 2005, p. 141).

Samper Hernández afirma que Michéa fue el primero en establecer la oposición de las “palabras temáticas” y las “palabras atemáticas” (2002, p. 8) y esta oposición sigue siendo reforzada por otros autores hasta hoy (López Morales, 2005; Gallego Gallego, 2014, p. 99⁴).

2.1.2 Centros de interés

Dada la estrecha relación entre el léxico temático y el tema del discurso, Gallego Gallego enfatiza la necesidad de activación mental para que se actualicen estas palabras (2014, p. 99). Por lo tanto, el siguiente paso era dejar de obtener el léxico fundamental para la enseñanza recopilando palabras frecuentes y más bien fijarse en las asociaciones mentales producidas tras un estímulo. Según López Morales, no se tardó mucho en empezar a investigar el léxico disponible y estas investigaciones se ejecutaban a través de pruebas asociativas, de manera artificial: se observaban las listas de léxico producidas bajo la influencia de unos estímulos llamados “centros de interés” (CI) (1995, p. 246). González Fernández argumenta que la selección apropiada de estas áreas temáticas es una de las mayores complicaciones en el estudio de disponibilidad léxica y destaca que desde el principio de los estudios los CI prácticamente no han cambiado (2014, p. 2). Sin embargo, para presentar la lista de los CI, consultemos la acordada para las investigaciones del Proyecto Panhispánico de léxico disponible (del cual se tratará más en el apartado 2.2.1):

⁴ Gallego Gallego hasta acepta (indirectamente) los términos “atemáticas” y “frecuentes” como sinónimos (2014, p. 99), lo que es difícil de aceptar, dado el modesto experimento que he llevado a cabo antes: las palabras “gobierno”, “presidente” o “político” no se pueden considerar atemáticas, pero, de todos modos, según la lista del CREA han sido confirmadas como (relativamente) frecuentes. Para estar o no estar de acuerdo con esta supuesta sinonimia, tendríamos que aceptar primero una de estas opciones: por un lado, que las palabras frecuentes, en realidad, no son necesariamente atemáticas; o, por otro lado, que la lista de frecuencias CREA no refleja verdaderamente las frecuencias reales de la aparición de las palabras, y así declarar que la metodología del CREA no es eficaz.

01. 'Partes del cuerpo'
02. 'La ropa'
03. 'Partes de la casa (sin los 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. 'Iluminación, calefacción y medios de airear un recinto'
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'⁵ (Samper Padilla, Bellón Fernández y Samper Hernández, 2003, p. 47)

Algunas investigaciones introducen sus propios CI, por ejemplo, para reflejar las particularidades de la comunidad estudiada (Cruz Alonso, 2016, p. 25), y eso, aunque sigan la misma metodología. Mencionemos, por ejemplo, la categoría “La Rioja” incorporada en la investigación del léxico disponible de los preuniversitarios en La Rioja para analizar la percepción de los encuestados de su propia Comunidad Autónoma (Pérez Jiménez, 2016, p. 45); “Sensaciones y sentimientos” en Paredes García, Guerra Salas y Gómez Sánchez (2022) o el área temática “inteligencia” para estimular la producción de términos abstractos en

⁵ Es posible encontrar modificaciones de los nombres de los mismos centros interés, por ejemplo, Pérez Jiménez especifica: “09. Calefacción, iluminación y ventilación (medios de airear un recinto)” (2016, p. 45).

Hernández-Muñoz, Izura y Ellis (2006)⁶. En general, las áreas temáticas opcionalmente añadidas por los investigadores pueden ser específicas para la comunidad dada, como el caso del tema “La Rioja”; o no, como los otros dos CI que he mencionado.

En cuanto a la complejidad de la selección de los CI (mencionada por González Fernández, véase el párrafo anterior), la variación que acabo de presentar puede ser uno de los ejemplos. González Fernández señala que lo problemático es determinar las áreas temáticas adecuadas para asegurar la producción del léxico básico de la comunidad de habla dada (2014, p. 2) y en su trabajo se centra en la selección de los CI en el campo de la enseñanza del español como lengua extranjera (ELE), lo que va a ser el objeto de nuestro interés más adelante. Su observación se puede interpretar de la siguiente manera: es difícil determinar qué área temática es esencial para captar el léxico imprescindible a la hora de expresar su realidad.

¿Qué otros problemas metodológicos relacionados con la selección de los centros de interés pueden surgir? Paredes García avisa que los 16 CI clásicos, es decir, los utilizados en el tiempo del nacimiento de la disponibilidad léxica (y hasta hoy vigentes en los estudios contemporáneos del Proyecto Panhispánico), pretendían ser aplicables al estudio de cualquier individuo, sin que importen la situación o condición de este, y ser lo suficientemente coherentes como para estimular las asociaciones mentales de los encuestados. En cuanto a la universalidad de los CI clásicos, afirma que este aspecto ya ha sido criticado: el primero en hacerlo fue Mackey, quien llamó la atención sobre los factores con los que estaban relacionados los intereses humanos, como la edad, el nivel educativo, etc. (1971, como lo recoge Paredes García, 2014, p. 2). Aparte de los factores que pueden influir en la universalidad de los CI, Paredes García exalta la importancia de elegir apropiadamente la cantidad de los CI: Habla de la viabilidad de la encuesta, en otras palabras, la encuesta tiene que ser realizable sin ser demasiado extensa para que no se pierda el interés (2014, p. 2). Es adecuado esperar que la productividad de los encuestados no alcance la cumbre potencial una vez que ya estén demasiado aburridos, molestos o hasta frustrados.

En cuanto a la cualidad de los centros de interés, antes ya he dado a entender que el léxico disponible involucra principalmente los sustantivos. Borrego Nieto argumenta que la metodología de la encuesta estimula la producción de los sustantivos, sobre todo los concretos, y desfavorece la producción de otras categorías (2004, como lo recoge Paredes

⁶ Aunque esta última investigación no forma parte del Proyecto Panhispánico ni sigue la metodología usual del estudio de la disponibilidad léxica (principalmente por plantear objetivos distintos).

García, 2014, p. 2). Paredes García luego procede a la enumeración de los CI incorporados por varios investigadores para compensar esta falta y, así, asegurar la aparición de otro tipo de vocabulario, por ejemplo, “Defectos físicos y morales”, “Los colores” e incluso una categoría tan vaga como “Adjetivos” (2014, p. 3). Según Paredes García, otro indicador de la adecuación de los CI es la productividad: se ha demostrado que los distintos CI impulsan distintas cantidades de léxico producido. Los núcleos temáticos relacionados con una mayor productividad son “Los animales” y “Alimentos y bebidas” y entre los que presentan una menor productividad encontramos “Trabajos del campo y del jardín” e “Iluminación, calefacción y medios de airear un recinto” (una variante del centro de interés clásico), lo que se explica con un bajo grado de familiarización de los encuestados con estas áreas temáticas (2014, p. 3). Es posible esperar que, si un investigador opta por no introducir todos los CI, los menos productivos que acabo de nombrar sean buenos candidatos para ser eliminados, como hicieron en Díaz, Ibáñez y Oliva (2020). Carcedo González constata que el bajo grado de la productividad de algunos CI se debe a particularidades culturales (2000, p. 178). Samper y Hernández afirman la influencia del clima del territorio estudiado (Gran Canaria) y la falta de importancia en la vida cotidiana de los medios de calefacción en la región-sujeto de su investigación, Gran Canaria (1997, p. 233, como lo recoge Carcedo González, 2000, p. 178). Sin embargo, es indudable que este CI incluye el léxico de varias áreas climáticas, sea un lugar más familiarizado con la calefacción o con la ventilación, y eso sin mencionar siquiera la universalidad de la iluminación. Por otro lado, cabe preguntarse si los CI de tal pobreza léxica pueden ser descompuestos de manera que el léxico que abarquen se incluya dentro de otros CI (ya vigentes o no). Por ejemplo, “radiador”, que es el único vocablo que aparece en más del 80 % de la muestra en la comunidad de La Rioja (Pérez Jiménez, 2016, p. 179), está tan relacionado con el tema de la casa o los aparatos al menos tanto como las palabras “nevera” o “aspiradora”. Lo demuestra la misma investigación de La Rioja, dado que la palabra “radiador” no solamente aparece en el tema de la calefacción, sino también en “La cocina y sus utensilios”, aunque no se trata de un aparato exclusivo de la cocina (Pérez Jiménez, 2016, p. 110); así como en otros, como “Partes de la casa” (p. 387) y “Los muebles de la casa” (p. 394).

El último de los factores de cualidad planteados por Paredes García es el problema de la falta de homogeneidad: algunos núcleos temáticos funcionan como hiperónimos (es decir, que el centro de interés dado representa la denominación general de un conjunto y el léxico producido tras este estímulo designa los elementos que forman parte de conjunto),

como “Los animales”, y luego hay otros que facilitan la producción mucho más general y amplia, por ejemplo, “El campo” (2014, p. 3).

De hecho, está claro que no todos los CI son iguales ni igualmente válidos o importantes y que el conjunto de ellos es homogéneo externa o internamente. Efectivamente, la taxonomía de los CI también ha sido investigada por varios autores y hay varias propuestas, de las que presento algunas. Ya en el párrafo anterior he mencionado el hecho de que algunos centros de interés son hiperónimos. Hernández Muñoz estableció una dicotomía basada en el nivel de inclusividad y dividió los CI en inclusivos y relacionales, dependiendo de si la asociación entre el CI y su léxico es directa (es decir, la relación es hiponímica/hiperonímica) o indirecta (el léxico está meramente relacionado de alguna manera con el CI correspondiente), en orden respectivo (2006, como lo recogen Díaz, Ibáñez y Oliva, 2020). Un ejemplo claro del CI inclusivo es “Partes del cuerpo” (se supone que las unidades producidas por los informantes corresponden a las partes del cuerpo individuales – hipónimos del CI) y uno relacional es “La ciudad” (su planteamiento en una investigación no sugiere la enumeración de las ciudades, sino del léxico *relacionado* con la ciudad). En este sentido, es importante la cuidadosa selección del nombre del núcleo temático, porque, como señala Hernández Muñoz, el uso del CI “La escuela” supone resultados diferentes de los del CI “La escuela: muebles y materiales” (2006, como lo recogen Díaz, Ibáñez y Oliva, 2020). Aquí se ve claramente que la primera opción requiere léxico que está, más o menos, relacionado con la escuela y la segunda solo acepta palabras que designan los diferentes muebles y materiales. La primera opción es, por lo tanto, más abierta y menos estricta.

Aparte de las características internas, el aspecto externo de los CI tampoco es homogéneo (como ya he insinuado en el caso de la productividad). Alba aparentemente estuvo inspirado por Echeverría y su “índice de cohesión” y en su investigación del léxico disponible en República Dominicana ordena los CI según el criterio de la densidad, esto es, el grado de coincidencia en las respuestas. El valor de la densidad de cada CI lo determina dividiendo el total de las palabras producidas (la suma de respuestas) por el número de palabras diferentes (cada palabra está registrada una sola vez). Según los resultados de esta investigación, los CI de mayor densidad son “El cuerpo humano” (32), “Medios de transporte” (30) y “Animales” (26) (1995, p. 19-20).

En cuanto a la variación diacrónica del léxico disponible, su estudio se encuentra en desventaja en comparación con el estudio de la variación diatópica. Un proyecto que debo mencionar es *Observación del cambio lingüístico en tiempo real: El nuevo léxico disponible*

de los dominicanos, cuyo objetivo es investigar los cambios del léxico disponible en los dominicanos, comparando los datos obtenidos en 1990 y 2008. Es el primer trabajo de este carácter en el ámbito hispano (Alba, 2012, p. 23). Un aporte importante de este trabajo es la conclusión de que el léxico disponible de algunas áreas ha cambiado más que el de otras, por ejemplo, algunos de los núcleos más estables son “Animales” o “El cuerpo humano”; mientras que los temas “Comidas y bebidas”, “La ropa” y “Juegos y diversiones” han cambiado marcadamente. Esto se debe al carácter estático de algunas realidades (como el cuerpo humano o el reino animal) y al dinámico de otras (como el ámbito del vestido) (p. 129).

2.2 Las investigaciones del léxico disponible en el ámbito hispano

Después del lanzamiento de los estudios de la disponibilidad léxica en Francia, inmediatamente le siguieron estudios llevados a cabo en el Canadá francoparlante con una metodología similar para asegurar la comparabilidad de los datos (López González, 2014, p. 25). El concepto se difundió a otros países y, según Samper Hernández, aparte de las primeras aportaciones de Michéa y Gougenheim, las aportaciones del cambio de siglo presentan la mayor relevancia, especialmente las del mundo hispano (2002, p. 8).

Las investigaciones en el ámbito hispano empezaron su trayectoria en Hispanoamérica en los años 70 siguiendo cuidadosamente la metodología de los estudios francoparlantes (Samper Hernández, 2002, p. 9). Puerto Rico es la cuna de los estudios de la disponibilidad léxica en la lengua española y Humberto López Morales, con su libro *Disponibilidad léxica en escolares de San Juan* de 1973, es considerado el padre de estas investigaciones en español. Como avisa López González, aunque López Morales impulsó numerosas investigaciones en la América hispanohablante, estos trabajos carecen de metodología unificada. Las diferencias estriban en la cantidad y selección de los centros de interés, la edad de los informantes y el tamaño de la muestra, pero también en la consideración de las diferentes variables, tales como el sexo, nivel sociocultural y nivel escolar, tipo de centro, etc., dado que no todos los investigadores tienen en cuenta las mismas variables (2014, pp. 27-29). Así, no es sorprendente que surgiera un proyecto que facilitara la comparabilidad de los datos de todo el mundo hispano, dirigido y coordinado por el mismo Humberto López Morales.

2.2.1 Proyecto Panhispánico de Disponibilidad Léxica

Como he señalado, del interés por el estudio de la disponibilidad léxica en la lengua española y, a la vez, de la falta de conformidad metodológica, nació el Proyecto

Panhispánico de Disponibilidad Léxica (PPHDL) bajo la dirección de Humberto López Morales.

El objetivo del proyecto es este:

Elaborar diccionarios de Disponibilidad Léxica para las diversas zonas del mundo hispánico. La homogeneidad de criterios permitirá establecer comparaciones de tipo lingüístico, etnográfico y cultural, dibujar áreas de difusión y, en general, servirá de punto de partida para análisis posteriores. (Dispolex, s. f.)

Y, aparte de los mismos diccionarios (o justo para que estos pudieran finalizarse), el PPHDL supone un análisis tanto cuantitativo como cualitativo.

A continuación, presento brevemente las pautas metodológicas del PPHDL, recogidas por Samper Padilla, Bellón Fernández y Samper Hernández, establecidas en una reunión en Bilbao del año 1999 (2003, p. 46):

El primer punto es la selección de los centros de interés, que deben ser los mismos en todas las investigaciones que forman parte del proyecto: el PPHDL cuenta con los 16 CI clásicos, aunque ya han sufrido pequeñas modificaciones en sus nombres⁷ y siempre hay posibilidad de añadir otros, como ya se ha destacado. La mayor modificación, según Samper Padilla *et al.*, la ha sufrido el centro de interés “Iluminación, calefacción y medios de airear un recinto”, que originalmente carece del elemento “medios de airear un recinto”, cambio realizado para acercarse más a los hablantes de los países tropicales (2003, p. 49), pero aun así sigue siendo uno de los dos CI menos productivos del proyecto (p. 59). Gracias al requisito de análisis cuantitativo, cada investigación supone el resultado de la productividad del CI (según el promedio de las palabras, número total de palabras, etc.).

La siguiente pauta reside en la ejecución de la prueba: en vez de listas cerradas (que requieren un número concreto de respuestas sin límite temporal) se opta por el uso de las listas abiertas con límite temporal de dos minutos (sin límite de palabras), lo que estuvo posibilitado por la decisión de que los encuestados fueran estudiantes de 18 años (que no necesitan tanto tiempo como niños o adolescentes menores) (Samper Padilla *et al.*, 2003, pp.

⁷ A pesar de haberlos enumerado ya en el apartado 2.1.2, vuelvo a presentarlos para facilitar la lectura del trabajo: 01. ‘Partes del cuerpo’ 02. ‘La ropa’ 03. ‘Partes de la casa (sin los 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. ‘Iluminación, calefacción y medios de airear un recinto’ 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’ (Samper Padilla *et al.*, 2003, p. 47).

63-64). De hecho, la muestra debe estar compuesta de estudiantes preuniversitarios (en España, el último curso del bachillerato) para que los resultados no se distorsionen por los tecnicismos adquiridos en la universidad (DispoLex, s. f.). Además, según la revelación de Borrego Nieto, la media de las palabras producidas por los jubilados es mucho menor que en los jóvenes (2008, p. 230-231). Por eso, resulta menos problemático mantener la conformidad de la edad en las investigaciones.

Volviendo a Samper Padilla *et al.*, otro de los requisitos de la prueba es que se responda a las encuestas por escrito, lo que hace posible una recolección de gran cantidad de datos y de varios encuestados a la vez, en una sola sesión (2003, p. 65). En cuanto al tamaño de la muestra, se acordó que el número de encuestados fuera de 400 (p. 68). Además, al final del capítulo de Samper Padilla *et al.* se adjunta el modelo del formulario, incluyendo la primera hoja que recoge datos sobre el encuestado, necesario para el análisis según variables.

Efectivamente, las variables obligatorias del PPHDL o “condicionantes extralingüísticos”, son estas: sexo, nivel sociocultural, zona geográfica (urbana/rural), tipo de centro (público/privado) y también hay condicionantes particulares para el estudio de las comunidades bilingües (Samper Hernández, 2002, p. 10). López González afirma: “La sociolingüística se ha constituido en el pilar básico de los estudios de disponibilidad” (2014, p. 103). Es indudable que las variables que acabo de enumerar condicionan el lenguaje producido por cualquier hablante, entonces, es imprescindible incluirlas obligatoriamente en los estudios de la disponibilidad léxica y realizar una parte del análisis basándose específicamente en ellas.

En cuanto a la variable “sexo”, López González advierte de que, a pesar de la introducción obligatoria de esta variable, los informantes no se eligen en función de esta, sino que se trabaja con los estudiantes dados (sin que importe la ratio) y la variable simplemente se considera a la hora del procesamiento de los datos, lo que lleva a gran desproporción en favor de las mujeres (2014, p. 104). Afirma que esta desproporción no causa diferencias enormes en la productividad general del léxico. Sin embargo, señala que numerosas investigaciones muestran que las mujeres superan a los hombres en la productividad dentro de los CI “La ropa”, “Los muebles de la casa”, “La cocina” y “Los colores”, y que, por el contrario, varias investigaciones muestran mayor productividad de los hombres en “Calefacción e iluminación”, “Medios de transporte” y “Trabajos del campo y del jardín” (2014, p. 104-107).

La variable “nivel sociocultural”, según López González, muestra gran significación, ya que se repite con regularidad el modelo de menor disponibilidad en los informantes de nivel sociocultural bajo (2014, p. 110).

El PPHDL cuenta con varios criterios de edición esenciales para posibilitar el análisis comparativo (Samper Padilla *et al.*, 2003, p. 95). Por lo tanto, la elaboración de las investigaciones incluye un procesamiento profundo de las respuestas. Algunos de los ajustes son los siguientes: corrección ortográfica, unificación ortográfica si hay más formas correctas (este ajuste está principalmente explotado en el campo de los neologismos en favor de la forma adoptada por los diccionarios del español, por ejemplo *football* y “fútbol” se unifican bajo la segunda opción; excepto si las formas son muy variadas, como “fútbol” y “balompié”), fusión de las formas plenas con sus acortamientos (*boli[grafo]*) y otros; pero lo que causa complicaciones peculiares son los nombres de marcas comerciales, de los cuales solo se admiten aquellos totalmente lexicalizados (como *colacao*) (Samper Padilla *et al.*, 2003, pp. 95-99).

Ya he mencionado varias veces la importancia del análisis cuantitativo. La investigación del léxico disponible supone automáticamente un cálculo del índice de la disponibilidad para cada palabra individual y el PPHDL recurre a una fórmula introducida fuera del proyecto por López Chávez y Strassburger Frías en 1991 (Samper Padilla *et al.*, 2003, p. 101; López González, 2014, p. 44). El cálculo de este índice es una de las aportaciones más importantes del ámbito hispánico al estudio de la disponibilidad léxica (Samper Padilla *et al.*, 2003, p. 102) y la fórmula es la siguiente:

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

Donde:

n = máxima posición alcanzada en el centro de interés en la encuesta,

i = número de la posición en cuestión,

j = índice de la palabra tratada,

e = número real (2.718181818459045),

f_{ij} = frecuencia absoluta de la palabra j en la posición i,

I = número de informantes que participaron en la encuesta,

$D(P_j)$ = disponibilidad de la palabra j (Samper Padilla *et al.*, 2003, pp. 101-102; López González, 2014, p. 44).

Anteriormente he constatado que una palabra disponible no tiene que ser especialmente frecuente en el habla (ni aparecer como tal en los listados de frecuencias), pero a partir de la fórmula se evidencia que la disponibilidad de una palabra está fuertemente condicionada por su frecuencia. Sin embargo, hay que aclarar que aquí lo que nos interesa es la frecuencia aislada: la que se produce a lo largo de las encuestas realizadas en el entorno artificial de la prueba asociativa.

A pesar de la complejidad del cálculo, la tecnología permite un alto grado de procesamiento de datos extensivos. Desde el año 2000 funciona la página web DispoLex.com, creada por José Antonio Bartol Hernández y Natividad Hernández Muñoz desde la Universidad de Salamanca (López González, 2014, p. 10), hoy gestionada por los profesores Bartol Hernández y Julio Borrego Nieto. Por un lado, la página acumula y almacena los datos del PPHDL, además de la información adicional acerca del tema, y, por otro lado, permite a cualquier investigador-participante del proyecto introducir los datos de su investigación y aprovechar las herramientas para realizar los cálculos (aparte del índice de disponibilidad (ID), es posible calcular el porcentaje de aparición, promedios por informantes, comparaciones entre proyectos, etc.). No obstante, DispoLex no es la única plataforma para el procesamiento de los datos: ya desde el año 1995 está accesible el programa LexiDisp⁸ (López González, 2014, p. 45) y durante los últimos años ha sido posible acudir al programa LexPro⁹ (una extensión del proyecto DispoGram, que pretende estudiar la disponibilidad gramatical del español), resultado de la colaboración entre la Universidad de Salamanca y la Universidad Miguel Hernández de Elche. El programa brinda numerosas posibilidades, entre las que destaco el cálculo de los datos para las palabras individuales en cada CI por separado: el índice de la disponibilidad (según la fórmula anteriormente presentada), las frecuencias relativa (el porcentaje de la palabra entre el total de palabras de la muestra) y acumulada (la suma de las frecuencias relativas) y la aparición del vocablo (el porcentaje de los informantes que incluyen la palabra); pero también facilita la comparación de los datos entre grupos de informantes y la generación de gráficos y redes léxicas (Hernández Muñoz, Tomé Cornejo, López García y Bartol Hernández, 2023, pp. 7-9).

⁸ LexiDisp no proviene del ámbito hispanohablante.

⁹ <https://dispogram.usal.es/lexpro/>.

El surgimiento de las herramientas tecnológicas de este tipo en el ámbito hispanohablante demuestra el interés que hay por este tema. De hecho, en actualidad el PPHDL dispone con un repertorio extenso de investigaciones en el campo de la disponibilidad léxica en los países hispanoamericanos, y aún más más fecundo en España, donde cuenta con al menos 39 provincias ya estudiadas, lo que permite la progresiva elaboración del proyecto *Léxico disponible de España* (Bartol Hernández y Borrego Nieto, 2019, p. 58).

2.3 Aplicaciones

2.3.1 Español como lengua extranjera

Es imprescindible la relación entre el concepto lingüístico de la disponibilidad léxica y la enseñanza/el aprendizaje de lenguas extranjeras, dadas las circunstancias en las que surgió. Efectivamente, la DL no solamente es un recurso útil en este ámbito, sino que este fue el objetivo con el cual nació. No obstante, su utilidad en ELE también es variada. Samper Hernández sostiene:

[...] por un lado, pone de manifiesto carencias o deficiencias en la competencia léxica de los estudiantes, hecho que permite poner los medios para que esto no ocurra; por otro lado, nos proporciona el vocabulario que realmente se utiliza en una comunidad lingüística determinada, por lo que se evita caer en el error de incluir en los listados que se destinen a la enseñanza los ya citados arcaísmos, cultismos, etc. de los diccionarios *ad usum*. (2002, p. 13)

Dicho esto, acerquémonos a la DL en los estudiantes de ELE. En este tipo de estudios, se aplica una metodología parecida a la de los estudios en hablantes nativos. Naturalmente el carácter y el propósito de la investigación exigen ajustes, como variables adicionales: López Rivero cita “nivel de estudios”, “conocimiento de otras lenguas” y “lengua materna” (2008, pp. 37-38). La autora en su trabajo pretende acercarse a la disponibilidad léxica de los estudiantes de ELE, residentes en Madrid. Mencionemos un trabajo más reciente: el artículo que Rubio (2017) dedicó al léxico de los estudiantes italianos de ELE.

En cuanto al segundo tipo de investigación, y este es más significativo para los objetivos de este trabajo, las cosas se complican, ya que hay varios enfoques que pueden asumir los investigadores. Por un lado, existen investigaciones que exploran las diferentes maneras de mejorar la enseñanza de ELE (o lengua extranjera, en general) aprovechando las listas de léxico disponible de los nativos (véase 3.2) y, por el otro, hay trabajos que evalúan el reflejo este en los materiales para la enseñanza de lenguas extranjeras. Ya Benítez Pérez realizó un estudio de estas características en 1994 y llegó a la interesante conclusión de que,

además del gran número de palabras con un ID alto, los manuales también presentan muchas palabras poco disponibles y hasta palabras con el índice cero (1994^a, p. 332). Destaco también a Serfati (2019), que, de modo parecido, trata de encontrar la medida en la que dos manuales frecuentemente usados en Marruecos tienen en cuenta el léxico disponible de los hablantes nativos y compara ambos manuales. Igualmente, hay investigaciones que mezclan los dos enfoques metodológicos, como Hidalgo Gallardo (2017) en su tesis doctoral: estudia, por una parte, el léxico disponible de los estudiantes chinos de ELE, y, por otra parte, el léxico impuesto por los manuales usados.

En el apartado 2.1.2 he expuesto la oposición entre los CI inclusivos y relacionales. Es natural que percibamos algunos de los CI como inclusivos, al menos a la primera vista: – por ejemplo, “La ropa”; pero cada CI inclusivo es relacional al mismo tiempo, es decir, que hay léxico que es hipónimo del nombre de aquel CI, pero al mismo tiempo hay léxico que solo se relaciona con él. De hecho, si un informante está expuesto a este tipo de CI, podemos presumir que instintivamente lo primero que se le ocurrirá es el conjunto de hipónimos, y que de las palabras que solo se relacionan con el tema (sin ser hipónimos) se acordará más tarde o no se acordará de ellas en absoluto. Por lo tanto, me atrevo a decir que quedan dos opciones: o bien asumimos que las circunstancias de la realización de las encuestas ni se acercan a simular las situaciones comunicativas reales, o bien acordamos que el léxico disponible tampoco es la herramienta definitiva para la elaboración del vocabulario para los manuales. Para sostener esta atrevida declaración, recurro a un ejemplo que he obtenido del ya mencionado Serfati: una parte de su análisis cualitativo se centra en comparar la lista de palabras disponibles con la del vocabulario extraído de un manual y analizar las discrepancias entre estas. Como consecuencia, hay palabras que aparecen en el libro, pero no son lo suficiente disponibles, las que llama “superfluas” (2019, p. 320). Es crucial reflexionar sobre si las palabras como “talla” (una de las palabras que menciona Serfati) verdaderamente son “superfluas” en el tema de la ropa, cuando solamente nos basamos en el criterio de la disponibilidad (que en el caso de esta palabra puede ser relativamente baja o hasta 0). En resumen, desde mi punto de vista, todavía quedan huecos metodológicos acerca del análisis de los manuales y la presencia del léxico disponible en ellos.

2.3.2 Otras aplicaciones

Para bosquejar lo útil que puede llegar a ser el estudio de la disponibilidad léxica, además de la enseñanza de ELE, describo brevemente otros campos en los que se puede aplicar.

Dada la inherente relación entre las investigaciones y sus correspondientes zonas geográficas, es innegable la importancia de la disponibilidad léxica en la dialectología. Las comparaciones de distintas zonas geográficas fueron pilotadas por el anteriormente citado Mackey en 1971 y continúan hasta hoy, siendo el PPHDL un ejemplo de la enorme dedicación que recibe este ámbito lingüístico. Las aportaciones a la sociolingüística están aseguradas principalmente por el uso de las diferentes variables, que permiten la comparación de los grupos sociales dentro de la misma zona geográfica. En cuanto al segundo campo, la psicolingüística, esta se interesa en gran medida por el lexicón mental, lo que centra el objeto del estudio de la disponibilidad léxica. Mencionemos a Hernández Muñoz, Izura y Ellis (2006), que estudian el aspecto cognitivo del asunto.

3 La enseñanza y el aprendizaje del vocabulario de la lengua extranjera

En los siguientes apartados se expondrán varios aspectos de la enseñanza de la lengua extranjera: abro el gran tema de *vocabulario versus gramática*, la adquisición de la competencia léxica, pero también se establece una breve base teórica de la composición de los manuales.

3.1 Las peculiaridades del vocabulario en la lengua extranjera

Para empezar, conviene precisar qué se entiende por el término “vocabulario”. Parto de la definición propuesta por el *DLE*: “Conjunto de palabras de un idioma” (adjuntando el sinónimo “léxico”)¹⁰, lo que es una definición simple, pero ambigua. De hecho, a nosotros, los que ya hemos tenido experiencia con la enseñanza escolar de las lenguas extranjeras y que tenemos una imagen relativamente clara sobre qué es el vocabulario (simplemente dicho, son *palabras*), no nos aporta ninguna información nueva, ni tampoco nos ayuda a restringir lo que se entiende por “vocabulario” en el ámbito de la didáctica. Al afrontar la tarea capciosa de decir qué es el vocabulario o hasta dar ejemplos de lo que abarca, estamos expuestos a la tentación de enumerar palabras como “gato”, “coche” o “comer”. Según Loureiro, es una estrategia errónea, porque así se omiten unidades como “aunque” o “este” y eso nos lleva a concluir que el léxico solo abarca las palabras “que tienen carga semántica”, lo que es una conclusión equívoca. Efectivamente, el vocabulario incluye todo tipo de palabras, desde la gramática normativa hasta las muletillas y la jerga. Aún más curioso es que la autora incluya en el conjunto las expresiones formadas por más de un miembro, por ejemplo: “tomar el pelo” (2007, p. 457-8). No está claro si la autora las considera unidades léxicas enteras o si deben ser percibidas como tales. Thornbury menciona las *colocaciones*, unas agrupaciones de palabras que aparecen juntas con mayor frecuencia de lo habitual, que difieren de las *unidades formadas por varias palabras*¹¹ (cuyo nivel de fijación es variado, por ejemplo, en inglés, *out of the blue* o *look for*) en el nivel de fijación, ya que las colocaciones presentan asociaciones aún más relajadas, por ejemplo, *set a new world record* (2002, p. 6-7). Así, resulta evidente que ambos autores tienen claro que las combinaciones de palabras/colocaciones son unidades

¹⁰ *DLE*, <<https://dle.rae.es/vocabulario>>, fecha de consulta: 13/2/2024.

¹¹ *Multi-word units* o *lexical chunks*.

que forman parte del conjunto del vocabulario, pero partiendo del Thornbury, no las vamos a considerar *palabras* como tales. Por otra parte, a estas alturas del siglo XXI, los lingüistas suelen considerarlas parte de la fraseología y, por tanto, no se estudian ya generalmente dentro del vocabulario. Otro asunto en el cual los autores están de acuerdo es que incluso las palabras de menor carga semántica pertenecen al vocabulario de una lengua determinada. Thornbury destaca la oposición entre las *palabras gramaticales* o *funcionales* (preposiciones, conjunciones, determinantes y pronombres¹²) y las *palabras de contenido* (nombres, verbos, adjetivos y adverbios), y es este segundo tipo el cual sufre mayores cambios. A pesar de esta oposición y la tradición de introducir las palabras gramaticales en la enseñanza de la gramática y las de contenido en la enseñanza del vocabulario, la frontera de esta separación se va borrando (2002, p. 4).

Estrictamente hablando, una palabra podría entenderse como cualquier secuencia de letras o sonidos independiente, separada de las demás por pausas (en la forma escrita), lo que otorgaría autonomía total a cualquier tipo de palabra, sea de naturaleza gramatical o de contenido. Sin embargo, tal definición separaría hasta las formas como “trabajamos” y “trabajas”, que representarían dos palabras distintas. Es evidente que se trata del mismo verbo, “trabajar”, y por eso no hay razón para separar aquellas formas en dos unidades léxicas. Según Lliteras, hay un acuerdo general de que las formas flexivas pertenecen a un mismo paradigma y que el conjunto de ellas está constituido por meras variaciones flexivas de una palabra (2019, p. 248). En un sentido parecido, Thornbury menciona en el tema de la adquisición del vocabulario el término *familia léxica*: la base léxica con todas sus inflexiones y la mayoría de sus derivaciones más comunes (2002, p. 4). La familia léxica es un concepto importante en el ámbito del aprendizaje, porque Thornbury avisa que la mente suele agrupar los miembros de la familia léxica juntos (2002, p. 5). No está claro por qué Thornbury solo incluye los derivados más comunes dentro de una familia léxica (y no todos), pero supongamos que el factor decisivo en el aprendizaje consiste en si el hablante/estudiante reconoce la pertenencia del derivado a la familia léxica. Este criterio es muy vago, pero el reconocimiento de las derivaciones tampoco es objetivo, sino que es un asunto individual.

¹² Es necesario resaltar que las palabras están asignadas a una clase gramatical específica en un idioma determinado, pero que esta puede diferir de la asignada a su equivalente en otro idioma. Por ejemplo, la palabra *tento* en eslovaco es un pronombre, mientras que su equivalente en español, “este”, se puede clasificar tanto como un determinante como un pronombre.

Ya hemos mencionado que las formas flexivas son simples variantes y no palabras distintas: esto se debe a que la base léxica (o una derivación de ella) es el verdadero portador del significado, mientras que la flexión contribuye a la palabra especificando la categoría gramatical. La siguiente pregunta que surge es si las categorías gramaticales y, por lo tanto, las flexiones de la palabra no aportan ningún significado. La respuesta simple es que, más que aportar un significado nuevo a la palabra, modifican su significado. Ambas formas, “trabajas” y “trabajamos”, designan la acción de trabajar y sus diferentes desinencias precisan quién realiza la acción indicada. El significado que aporta la desinencia es gramatical, pero tampoco es posible que este se realice a menos que sea dentro de la palabra. Según avisa Loureiro, las estructuras de la lengua siempre se realizan en determinadas palabras (2007, p. 458), y no omitamos que el “conocimiento gramatical” se almacena en la mente por medio de las palabras individuales y las relaciones entre ellas, por lo que la corrección y fluidez gramaticales provienen de haber procesado las palabras individuales repetidamente y en varios contextos a lo largo de mucho tiempo (Bacroft, 2005, p. 569). Además de servir de base para la realización de la gramática, el conocimiento del léxico puede ser considerado la parte más indispensable del conocimiento de la lengua. La certeza de esta declaración depende principalmente de la jerarquía de los objetivos a la hora de usar la lengua. Por un lado, el conocimiento del léxico de una lengua, aunque sea extenso y profundo, no es suficiente para dominar la lengua, pero, por otro lado, si estamos de acuerdo en que la función básica de la lengua es la comunicativa y el objetivo principal es comunicarse, resulta indudable el papel que ocupa el vocabulario. Lightbown y Spada afirman:

As it has often been remarked, we can communicate by using words that are not placed in the proper order, pronounced perfectly, or marked with the proper grammatical morphemes, but communication often breaks down if we do not use the correct word. (2013, p. 60)

No es única esa declaración en este sentido: de hecho, Barcroft también apoya esta postura, demostrándola con el ejemplo del mensaje “me gustan los camarones” y sus posibles realizaciones erróneas “me gusta los camarones” y “me gustan los calamares”. Mientras que la primera frase es comprensible y transmite el mensaje a pesar del error gramatical, la segunda causa un ruido comunicativo, ya que el mensaje que recibe el receptor es muy distinto del que pretende expresar el emisor. Así, el autor demuestra la importancia del vocabulario en la comunicación, frente a la gramática (2005, p. 569).

Una vez comprendida la importancia del conocimiento del vocabulario de una lengua (extranjera), voy a tratar el tema de su adquisición y mantenimiento. Antes de nada,

definamos qué es el “conocimiento” de una palabra. Las definiciones usuales varían, pero lo que tienen en común es el rechazo a la simplificación de que conocer una palabra equivalga a conocer su equivalente en su lengua materna. Thornbury propone dos requisitos para conocer una palabra: conocer su forma y su significado. No obstante, advierte de lo insidiosa que puede ser esta simplificación, dado que en ella no se tienen en cuenta parámetros como el reconocimiento de la clase de palabra o el significado cultural. Según el autor, conocer verdaderamente una palabra significa conocer también las colocaciones comunes en que toma parte, el registro en el cual se usa (por ejemplo, coloquial), sus connotaciones y su uso correcto (2002, p. 15). En cuanto al aspecto externo, Oster especifica que conocer la forma de la palabra no solo es conocer su grafía, sino que incluye también la familiarización con el aspecto fonético (2009, p. 34). Para complicar el asunto aún más, Thornbury añade que el conocimiento de una palabra no es una cuestión de sí o no, sino que es un espectro, y que habitualmente se prioriza el tamaño del vocabulario en detrimento de su profundidad, es decir, de *cuánto* se conoce cada una de las palabras de forma individual. Por eso, al referirse al conocimiento del vocabulario, hay que tener en cuenta el conocimiento receptivo y productivo (2002, p. 21). Por una parte, el conocimiento receptivo se refiere al nivel de la capacidad de comprender, de manera que un nivel alto supone la facilidad de comprensión sin problema en cualquier forma o contexto; y por la otra, el conocimiento productivo indica la facilidad de producción de la palabra, esto es, su ortografía y pronunciación, la automaticidad de la actualización en la mente del hablante, etc. La adquisición del vocabulario es un cambio constante en la dinámica de estas dos capacidades, pero lo que está claro es que cierto nivel de conocimiento receptivo de una palabra es un requisito innegociable para poder aumentar la capacidad de producirla.

Esto se puede observar en los bebés, que (generalmente hablando y omitiendo los factores fisiológicos) primero solo absorben la lengua, a continuación, empiezan a mostrar que entienden y son capaces de comunicar en cierta medida, y solo después son capaces de producir, sobre todo las palabras frecuentes que escuchan diariamente. La impresión habitual de la adquisición de L1 es que los niños aprenden muy rápido con muy poco esfuerzo, pero lo que se omite es el nivel de inmersión y la cantidad de *input*, que son enormes en el caso de los niños pequeños. Lightbown y Spada sostienen que, aunque los niños aprenden miles de palabras en L1 sin esforzarse, no es el caso en la adquisición de L2. El factor más evidente es la cantidad de *input*, dado que los aprendientes de L2 suelen estar expuestos solo a pequeñas porciones del idioma meta y en contextos más complicados, pero también, en el caso de niños mayores o adultos, las palabras a las que están expuestos pueden ser más difíciles, con

significados que no resultan evidentes a partir del contexto. (2013, p. 61). Para explicar el último punto: cuando un niño aprende a hablar, primero comunica los conceptos más básicos y más o menos concretos, como su madre o la comida, avanzando paulatinamente a conceptos cada vez más abstractos y difíciles; mientras que los niños o adolescentes (hasta adultos) que empiezan a aprender L2 ya comunican sobre aquellos conceptos relativamente complejos, lo que se refleja en el currículo del idioma, que contiene temas más complejos. Sin embargo, ya he insinuado que la clave para la adquisición efectiva de la lengua es la repetición y la frecuencia del *input*, y esto vale tanto para la lengua materna como la extranjera. Efectivamente, Criado y Sánchez argumentan que la repetición y la práctica son los factores más importantes en el proceso del aprendizaje de cualquier actividad humana, dado que la repetición refuerza las conexiones neurales hasta lograr la automatización. Los autores están convencidos de la relación entre la frecuencia¹³ y la adquisición de la lengua, a pesar de la existencia de un desacuerdo sutil entre los académicos (2012, p. 79-80). Thornbury destaca que cualquier modo de repetir la palabra es útil: hasta la repetición inmediata después de la primera exposición a la palabra. No obstante, la repetición más efectiva es la que ocurre al menos 7 veces en intervalos espaciados (2002, p. 24). Lightbown y Spada afirman que el número de reencuentros con una palabra necesarios para poder adquirirla llega a ser hasta 16 veces (2013, p. 62). En estas últimas líneas me he estado refiriendo a la adquisición como a un proceso de reforzamiento de la posición de la palabra en el lexicón mental y el aumento de la posibilidad de recordarla más tarde, pero aquí se empieza a fundir la adquisición con el mantenimiento de la palabra. Como ya se ha destacado, la adquisición es un proceso continuo y es posible adquirir una palabra individual como para poderla producir en un contexto, pero no en otro. Por ejemplo, al aprender la palabra “mano” como una parte del cuerpo, el aprendiente es capaz de usarla como tal, pero eso no significa que sepa sobre la existencia de la expresión “a mano”, a pesar de que incluye esta palabra. Por lo tanto, no solo es importante la repetición para recordar el vocabulario, sino también para la profundización de su conocimiento. Esa es la razón por la que Thornbury recomienda “reciclar” las palabras en contextos variados (2002, p. 26-27). En consecuencia, la tarea del profesor y del manual es asegurar estas condiciones para efectivizar el aprendizaje de los alumnos.

¹³ Hay que avisar que, como se ha podido observar, varios autores reemplazan la palabra “repetición” por “frecuencia” y las usan como sinónimos.

3.2 Los manuales y la selección léxica

Lo más común en un ambiente escolar es utilizar manuales en las clases de idioma, aunque en las demás asignaturas no se haga. De hecho, la enseñanza de la lengua (extranjera) implica el uso de un manual que proporciona el material trabajado en clase (o en casa), además de determinar el currículo de la asignatura y servir de eje principal. Esta realidad causa que el manual decida sobre la materia, el contenido y el transcurso de la enseñanza. Por esa razón, las siguientes páginas se dedicarán a la aplicación de los conocimientos del subcapítulo anterior a la teoría de la elaboración del vocabulario en los manuales.

Según Benítez Pérez, la preocupación acerca de la enseñanza del vocabulario se centra poco en qué vocabulario enseñar y, en vez de eso, se enfoca en *cómo* hacerlo (1994^b, p. 9). Por este motivo, esta va a ser la primera pregunta que nos planteamos. Antes de nada, resulta imprescindible avisar que lo habitual es que una parte considerable de cualquier manual esté destinada principalmente a que el estudiante entienda el contenido (por ejemplo, de un texto que desarrolle el contenido semántico de la unidad), adquiera la gramática de manera implícita, mejore la comprensión y, así, la comunicación, etc., en resumen, se centra en perfeccionar el conocimiento general de la lengua. En consecuencia, el estudiante alcanza diferentes niveles de conocimiento de las palabras presentadas y el conocimiento de algunas llega a ser poco profundo o meramente receptivo. Sin embargo, si consideramos que cierto léxico debe constituir el núcleo del vocabulario en cualquier hablante, debemos suponer que, después de asistir a un curso completo de lengua extranjera, el estudiante debe tener conocimiento productivo y profundo de este léxico. Eso sugiere que debería establecerse una jerarquía, de manera que estas unidades léxicas estén priorizadas y se les preste más atención. De acuerdo con lo expuesto en el subcapítulo anterior, varios autores destacan la importancia de la exposición repetitiva del vocabulario (Barcroft, 2005, p. 569; Thornbury, 2002, p. 24), pero aun teniendo en cuenta esta premisa primordial, no es menos importante considerar la forma en la que se presenta y elabora el vocabulario. El camino más obvio y simple es presentar el vocabulario en listas extensas, apartadas del resto de la unidad, en el formato palabra-traducción. La didáctica moderna parece criticar apasionadamente la enseñanza mediante de las listas bilingües, asociándolas con el aprendizaje de memoria, que también es muy criticado. No obstante, Thornbury las considera útiles para el aprendizaje de gran cantidad de unidades léxicas y convenientes para la autocomprobación del alumno (2002, p. 32-33). Larrotta, por otro lado, considera esta herramienta aceptable para los principiantes que no poseen ningún conocimiento léxico, pero no la considera eficaz para los aprendientes

que ya tienen cierto conocimiento del L2 (2011, p. 8). Otro enfoque, uno mejor aceptado y aprovechado por los manuales modernos, es la integración del vocabulario nuevo dentro de las actividades basadas en texto. Thornbury señala que eso puede incluir la pre-enseñanza¹⁴ del vocabulario nuevo y esencial, para asegurar la comprensión del contenido, y las actividades subsiguientes, que refuerzan la adquisición del vocabulario (2002, p. 39-40). Además de las actividades anteriores y posteriores del texto, una estrategia efectiva es que el nuevo léxico imprescindible del tema se presente de manera llamativa, es decir, subrayado, en diferente tipo de letra o aumentando las apariciones de las estas palabras (Barcroft, 2005, p. 575). El esquema del aprendizaje del vocabulario propuesto por Loureiro es el siguiente: la presentación, la comprobación de la comprensión y la fijación de la forma y del significado de la palabra (2007, p. 461). Pero aparte del reforzamiento inmediato del vocabulario nuevo, ya antes se había mencionado la necesidad de su “reciclaje” regular. Thornbury sostiene que los manuales recientes¹⁵ toman en serio esta necesidad (2002, p. 39). En definitiva, lo óptimo es que el número de apariciones de una palabra en un manual, junto a la forma de su presentación y reforzamiento, corresponda a su nivel de importancia y que estas apariciones ocurran en ciertos intervalos.

Pasamos ahora a la siguiente pregunta: qué vocabulario enseñar y cuál es la fuente adecuada para una selección apropiada. Mencionemos primero el *Plan curricular del Instituto Cervantes*¹⁶ (*PCIC*), la opción que más fácilmente se ofrece, porque su apartado “Nociones específicas” contiene una base ya procesada de unidades léxicas para cada nivel y área temática, dispuesta a su uso en los materiales de ELE. Tomé Cornejo acentúa la similitud organizacional entre el *PCIC* y los trabajos de la disponibilidad léxica: los campos semánticos parecidos a los centros de interés; pero avisa de que el *PCIC* difiere de los estudios de DL en la manera de conseguir estas listas, ya que sus palabras han sido elegidas de manera subjetiva (2015, p. 360). Paredes García cita los casos concretos de discrepancia entre las unidades del léxico disponible y del propuesto por el *PCIC*: muchas palabras que alcanzan nivel de disponibilidad muy bajo y evidentemente no deben estar incluidas en un plan curricular de ELE o la ausencia de las palabras muy disponibles y tan básicas como “oreja” o “boca” (2015, p. 15). Es importante destacar que el *PCIC* incluye la unidad “boca de metro” en el nivel A2

¹⁴ Del inglés, *pre-teaching*, es decir, se refiere a la instrucción previa a la lectura.

¹⁵ Hay que tener en cuenta que la fecha de publicación de su trabajo es el 2002, pero se supone que la necesidad de reciclar el vocabulario a lo largo del manual no va desvaneciéndose en la didáctica moderna.

¹⁶ Instituto Cervantes. *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. <https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/> [fecha de consulta: 27/2/2024].

sin haber incluido previamente la palabra “boca” (*PCIC*)¹⁷. Benítez Pérez opina que el vocabulario usado en la clase debe reflejar fielmente el lenguaje del hablante nativo (1994^b, p. 12), pero esta intención puede ser estropeada por la subjetividad del *PCIC*.

Aparte de la clasificación según campos temáticos (centros de interés) y la autenticidad, es decir, la consecuente utilidad en la comunicación, el léxico disponible brinda al menos una ventaja más: según Bartol Hernández, el ID puede ser usado para determinar el tamaño de los conjuntos de palabras para cada uno de los niveles (2010, p. 93). A pesar del carácter matemático y, por lo tanto, objetivo, de la metodología, la medida del corte para el léxico relevante parece estar sometida a una decisión subjetiva. Ávila Muñoz sostiene que la medida de corte en los trabajos de DL tradicionalmente ha sido $> 0,1$ (2016) y ha sido empleada por diferentes investigadores, como Serfati (2019) o Hidalgo Gallardo (2017)¹⁸. De todas formas, Paredes García avisa que las primeras posiciones del ID están ocupadas por representantes prototípicos del CI en cuestión y, por eso, son los candidatos ideales para la selección léxica en ELE, pero también expone los diferentes datos estadísticos proporcionados por los diccionarios de la DL (2015, pp. 17-18). La duda que surge en este aspecto es la adecuación de cualquiera de los criterios de cálculo. El docente/editor tiene una variedad de posibilidades, pero elegir la más válida entre ellas es imposible. Por esta razón, uno de los procedimientos aceptables sería elegir un criterio y apoyarse, en caso de duda, en los demás. Además, la metodología va avanzando y es posible que en el futuro se invente alguno nuevo. De hecho, Paredes García cita el nuevo criterio de la prototipicidad, basada en el concepto matemático de los conjuntos difusos, introducido por Ávila Muñoz y Villena Ponsoda en 2010 (2015, p. 18).

Dicho esto, el consenso hoy en día es que la metodología actual tiene ventajas y limitaciones y que más investigación es necesaria, por ejemplo, una revisión de los CI, dado que varios autores citan esta falta en la metodología vigente en el estado actual del PPHDL (Bartol Hernández, 2010, p. 93; Paredes García, 2015, p. 11). Tomé Cornejo elabora una propuesta metodológica para adaptar el estudio de la DL a los objetivos del ELE, pero avisa de que se trata de un trabajo piloto y sus resultados no deben considerarse alternativos al

¹⁷ Instituto Cervantes. *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/niveles/09_nociones_especificas_inventario_a1-a2.htm, [fecha de consulta: 27/2/2024].

¹⁸ Este último autor usa la medida de corte $\geq 0,1$, pero se supone que se refiere al mismo valor, dado que la posibilidad de que el ID de una palabra sea exactamente igual a 0,1 es minúscula, a pesar de la discrepancia matemática entre los signos utilizados.

inventario del *PCIC*¹⁹ (2015, p. 485). Además de las carencias citadas que presenta la DL, no es recomendable recurrir a la selección completamente objetiva y automática del vocabulario utilizando solamente los diccionarios de DL. Tomé Cornejo advierte de la necesidad del empleo de otros criterios, como el apoyo en los cognados de L1, sinonimia y antonimia y hasta la frecuencia (Tomé Cornejo, 2015, p. 449). En definitiva, el surgimiento de un nuevo instrumento para la selección léxica no tiene que implicar la omisión total de los existentes.

Para resumir el tema de la composición de los manuales, se concluye que el punto de convergencia es sin duda la importancia de la repetición del vocabulario esencial y su elaboración variada, además de la eficacia de su presentación por campos léxicos. No obstante, el ámbito que sigue evolucionando y que presenta discrepancias es la pregunta de *qué* exactamente es ese vocabulario esencial. A pesar de las limitaciones que conlleva la introducción de la disponibilidad léxica en el proceso de la selección, es indisputable la serie de sus ventajas que aporta y, por eso, es una fuente fiable para la selección del vocabulario en ELE.

¹⁹ La autora se apoya en el *PCIC* en varios momentos de su investigación metodológica, cuya mayor innovación consiste en la adición de los CI para acercarse a los campos nocionales del *PCIC* (el número final es 47 CI) y, entre otros cambios, la consecuente división de los informantes entre varios grupos que recibieron cuestionarios distintos (2015, p. 371).

4 Metodología

4.1 Objetivo del análisis

El objetivo principal de esta investigación es saber si el léxico disponible de los hablantes nativos está reflejado en un manual originado y de uso frecuente en la República Checa. Como aún no existe ningún corpus unificado de los proyectos mencionados, PPHDL ni Español disponible de España, y, por lo tanto, no hay ningún diccionario universal que pueda servir de referencia para la selección léxica, no es el objetivo saber si se tuvo en cuenta el léxico disponible en el proceso de creación del manual, como ha sido afrontado el tema por Hidalgo Gallardo (2017, p. 571), sino que se procurará determinar si el léxico del manual *coincide* con aquel presentado en los estudios de la disponibilidad léxica, y en qué medida.

De este objetivo principal parten otros parciales:

- 1) Confirmar que una porción adecuada de las palabras más disponibles en los nativos está presente en el manual checo de español;
- 2) averiguar la porción del vocabulario recogido por el glosario del manual, en el área temática correspondiente, que forma parte del léxico más disponible;
- 3) comprobar si la presentación de las palabras más disponibles en el manual concuerda proporcionalmente con su disponibilidad, tomando en cuenta su presencia en el glosario, número de ocurrencias y su distribución (frecuencia) a lo largo del manual.

4.2 Selección de las fuentes

Para poder analizar la presencia del léxico disponible en un manual, es necesario seleccionar, primero, el manual y el diccionario de disponibilidad con los cuales se trabajará.

4.2.1 Selección del manual

Como manual-objeto de análisis se ha seleccionado la serie *Aventura*²⁰ de la editorial Klett, uno de los manuales de español frecuentemente usados en la República Checa, destinado a los estudiantes checos de secundaria (típicamente de 15 a 19 años). La serie, en su estado actual, consta de 3 tomos: *Nueva Aventura 1* (2018), *Nueva Aventura 2* (2019) y

²⁰ Para referirse a la serie entera, se utilizará este título universal, *Aventura*.

Aventura 3 (2011), correspondientes a los niveles A1-A2, A2-B1 y B1 del MCER, respectivamente. A pesar de que los tomos no pertenecen a la misma edición y de las consecuentes discrepancias en la fecha de publicación, para los fines de esta investigación, se omiten estas diferencias. Esta decisión se ha tomado por varias razones: por una parte, como ya se ha advertido, no es el objetivo demostrar la intención de los autores de consultar los estudios de DL a la hora de seleccionar el vocabulario, sino que se procura investigar la presencia de las palabras más disponibles, aunque no sea intencional. Por esta razón, el uso del manual *Aventura 3*, publicado en 2011, no excluye el uso de una investigación de DL con una fecha de publicación más reciente. Por otra parte, en el momento de la elaboración de este trabajo, la tercera parte del manual, *Aventura 3*, todavía se vende como parte de la serie *Nueva Aventura*²¹ y, siendo la única opción, todavía se usa, de modo que resulta relevante realizar un análisis de su contenido léxico.

Aunque los autores de las investigaciones citadas en el capítulo anterior tomaron por objetos de su análisis los manuales de nivel A1 o elemental (Hidalgo Gallardo, 2017; Serfati, 2019; Benítez Pérez, 1994^a), en este trabajo se analizará el curso completo del manual escogido con el propósito de poder evaluarlo de manera justa, presumiendo que, al organizar el vocabulario por áreas temáticas, es posible que el primer tomo carezca de términos pertenecientes a ciertos campos léxicos que podrían estar recogidos por los tomos posteriores.

Además de tomar en cuenta los tres tomos del manual, es necesario establecer cuáles de los materiales propuestos por el manual estarán incluidos: el sujeto del análisis será el libro del estudiante y el cuaderno de ejercicios correspondiente, que forma parte del manual. Para los objetivos de este análisis, no se distingue entre estos dos componentes, sino que ambos se consideran constituyentes plenos del material usado en clase.

4.2.2 Selección del diccionario de DL

Para crear una lista de palabras disponibles que sirva de referencia para el análisis, se consultará la investigación *El léxico disponible de Castilla y León* (Cruz Alonso, 2016). Se ha escogido este listado por tratarse de una de las investigaciones realizadas en una zona relativamente amplia de España y que no se ve influenciada por otros idiomas cooficiales. Además, es una publicación reciente cuyos resultados coinciden aproximadamente en el tiempo con la creación de los manuales analizados.

²¹ Klett. *Nueva Aventura*. <<https://klett.cz/aventura/nueva-aventura/>>, [fecha de consulta: 2/4/2024].

4.3 Procedimientos

4.3.1 Procedimientos para la obtención de los datos

En cuanto a la selección de las listas de palabras, se escogerán los tres centros de interés (CI) con el mayor índice de cohesión. Este se refiere al grado de coincidencia entre los informantes para cada CI, reflejando la proporción del total de palabras y las palabras distintas (Cruz Alonso, 2016, p. 156). En efecto, se supone que las palabras más disponibles dentro de los CI que presentan el mayor índice de cohesión ocupan unas posiciones muy consolidadas en el diccionario de disponibilidad léxica (DL), por ser asociaciones comunes con el CI dado, y es menos probable que estén influenciadas por decisiones individuales o minoritarias. Los tres CI que presentan el índice de cohesión mayor en la investigación escogida, *El léxico disponible de Castilla y León*, son los siguientes: “el cuerpo humano” (0,09), “la ropa” (0,07) y “los animales” (0,06).

Una vez seleccionados los CI, hay varias posibilidades de determinar las palabras más disponibles dentro de ellas, dando lugar, así, a la lista de palabras más disponibles: es muy común que los análisis de DL tengan en cuenta las primeras 20 palabras más disponibles para realizar el análisis cualitativo, por ejemplo, Alba (2013, p. 169) y Sánchez (2017, p. 148). A pesar de que los objetivos de esos análisis se alejan de los nuestros, su decisión de incluir las 20 palabras más disponibles implica que son las que consideran importantes y a las cuales atribuyen valor predictivo. En contraste, Tomé Cornejo, en su propuesta metodológica de selección léxica, opta por la delimitación del número de unidades léxicas basada en el número de unidades propuesto por el *PCIC* (2015, p. 405). En el contexto de este trabajo de fin de máster, sin embargo, se elige la delimitación tradicional: el índice de disponibilidad (ID) con el corte de medida de 0,1. Así, serán consideradas todas aquellas unidades que presenten el ID mayor o igual al valor 0,1. Esto causa, a diferencia de las opciones anteriormente citadas, que la selección del léxico de referencia no se base en la comparación de los valores del ID, sino que las palabras serán válidas o no dependiendo del valor absoluto del ID. Esta delimitación es, en mi opinión, la preferible para evitar excluir ciertas palabras por la mera extensión del léxico que presenta el ID mayor o igual a 0,1. En caso contrario, el CI que presenta una gran cantidad de palabras con un ID mayor de 0,1 podría sufrir pérdidas inadecuadas.

En cuanto a la terminología, hasta ahora, no se ha distinguido entre los términos “palabra” y “vocablo”, y este hecho se mantendrá igual. A pesar de la importancia de

distinguir entre ellos en los estudios de la disponibilidad léxica (Gallego Gallego, 2014, p. 99), no es el caso en este análisis, dado que no es necesario en ningún momento distinguir entre el total de las palabras y el número de las palabras *distintas* o vocablos.

En la lista final de las palabras más disponibles se apuntará solo la forma en singular, siempre que el diccionario de disponibilidad la admita, aunque esta no sea la que típicamente se usa (por ejemplo, *braga* en lugar de “bragas”, lo que es posible porque el diccionario de disponibilidad recoge la palabra en este formato: “braga(s)”). Con ello se busca posibilitar una orientación efectiva.

Una vez finalizada la lista de las palabras con el ID mayor o igual a 0,1 (en adelante, también LP0,1), que servirá de referencia para el análisis, se procederá a la comparación de aquella y del léxico incluido en el manual. Antes de presentar los procedimientos del propio análisis, se exponen los principales criterios empleados para la obtención de los resultados:

- La selección de los materiales –el libro del estudiante y el cuaderno de ejercicios– da lugar a la omisión de algunas entradas potenciales: respuestas correctas en los huecos, audiciones, fichas adicionales, etc. Por otro lado, la fuente de entradas válidas será todo tipo de texto (a menos que no esté de acuerdo con alguno de los demás criterios), es decir, los ejercicios de vocabulario o de gramática y los textos de comprensión lectora, pero también los sumarios gramaticales y de comunicación, ejemplos, explicaciones, instrucciones, etc.
- No son válidas las apariciones dentro del texto introductorio al principio del manual, porque no forma parte del material trabajado en clase. (Ejemplo: “*Aventura* ha pasado por unos nueve años de existencia y por las *manos* de decenas de miles de estudiantes y cientos de profesores.”)
- Tampoco son válidas las apariciones dentro de la tabla de contenidos, a menos que aparezcan, al mismo tiempo, en el título de una unidad didáctica, en cuyo caso se tomará por válida tan solo una vez, aunque aparezca más veces a lo largo de la unidad. Esto se debe a que los títulos, que encabezan cada página del capítulo dado, no forman parte de la materia estudiada en clase y, por eso, no se toma en cuenta cada una de las apariciones. (Ejemplo: “8. *Manos a la obra*”)

- Si la palabra en cuestión se repite dentro de un texto o un ejercicio, se tomará por válida cada una de sus apariciones. En este caso, se supone que el aprendiente le presta suficiente atención a cada una de ellas. (Ejemplo: “Jorge: ‘No me gusta nada tu *perro*.’ – A Jorge no le gusta nada mi *perro*.”.)
- No son válidas las apariciones dentro de los nombres propios, dado que no designan el significado real de la palabra. (Ejemplos: “Castilla y *León*” o “La Estatua del *Oso* y el *Madroño*”.)
- Son válidas las formas flexionadas (con las distintas desinencias) y derivadas mediante sufijos apreciativos, puesto que estas no modifican el significado original de la palabra. (Ejemplo: *osito*, *perrito*.)
- Se aceptan las apariciones dentro de los compuestos, tanto los univerbales como los sintagmáticos. (Ejemplos: *osito* de peluche, cepillo de *dientes*, *minifalda*, *pintalabios*)
- Son válidas las apariciones dentro de las unidades fraseológicas, metáforas y refranes, siempre que su relación con el significado original de la palabra sea transparente (ejemplos: “a *pie*”, “a *mano*”) o al menos semitransparente (ejemplos: “*corazón* roto”, “gritar a todo *pulmón*”, “¡*ojo!*”). No son válidos los casos como “buen *humor*” o “a *pie* juntillas”, donde la relación con el significado original de la palabra en cursiva o se ha perdido, como en el caso de la primera expresión, o se desconoce por completo, como en el último ejemplo.
- No son válidos los homónimos ni las palabras polisémicas que designan un significado muy alejado del original (o del deseado), aunque compartan el origen etimológico. (Ejemplos: *burrito* [plato/animal], *muñeca* [juguete/parte del cuerpo humano], *medias* [adjetivo/prenda de ropa]) Sin embargo, si la relación entre el significado deseado y el alternativo es lo suficientemente transparente, se consideran válidas ambas formas. (Ejemplos: *lengua* [idioma/parte del cuerpo], *ratón* [animal/aparato])
- También son válidas las apariciones en las que se supone que no es importante su significado, por ejemplo, en los ejercicios de pronunciación. Esto se debe a que el conocimiento de la forma, pronunciación y ortografía forma parte del conocimiento de la palabra y, por lo tanto, la aparición sirve de exposición a la palabra, aunque el componente semántico no esté involucrado. (Ejemplo: “Tres tristes *tigres* tragan trigo en un trival.”)

- Para averiguar el contexto o el significado de las palabras, se usan varios elementos de apoyo: las imágenes, el tema de la unidad, el resto del texto que rodea la palabra, el contexto de la página, la traducción en el glosario, etc. (Ejemplos: “Mi juguete preferido era un cochecito para *muñecas*...”, “a) una falda, b) unas *medias* c) unos calcetines”, “a *medias* – napùl, polovičatě”.)

4.3.2 Procedimientos del análisis

El análisis y sus partes constituyentes se basan en estas premisas principales, basadas esencialmente en la información expuesta en el capítulo anterior:

- 1) El requisito primordial para adquirir una palabra es estar expuesto a ella (tanto a su forma como a su significado).
- 2) El siguiente factor de la adquisición del vocabulario es la repetición o frecuencia de esta exposición.
 - En primer lugar, una posibilidad de repetición es la que ocurre en la enseñanza directa, durante la introducción de vocabulario nuevo y su práctica;
 - en segundo lugar, una gran parte del aprendizaje consta del autoaprendizaje, es decir, el ejecutado por el aprendiente fuera de la institución en la que estudia (en este caso, se trata principalmente de la escuela secundaria). Una de las herramientas que le sirven al aprendiente en este proceso es el glosario proporcionado generalmente por los manuales, como ocurre en la serie *Aventura*;
 - y, por último, la repetición posterior al tema tratado, que fortalece o consolida el mantenimiento del vocabulario ya adquirido en el cerebro. Esta repetición debe ocurrir a intervalos mayores que las anteriores.
- 3) La opción idónea consta de una combinación de los factores anteriormente citados: la presencia del vocabulario meta en el material de clase y en el glosario y su repetición frecuente y recurrente en ciertos intervalos.

4.3.2.1 Presencia de las palabras disponibles en el manual

La fase introductoria del análisis parte de la primera premisa, la que sugiere que el aprendiente no adquiere una palabra si no la percibe, y se dedicará a la determinación de la porción de palabras disponibles que están presentes en el manual. Aquí, lo único que nos interesará será la mera presencia de las palabras, esto es, sin tomar en cuenta el número de

sus apariciones (u ocurrencias). Es decir, solamente se distingue entre las palabras que aparecen una o más veces y las que no aparecen ni una vez en el manual.

Aparte de esto, se averiguará la proporción de las palabras disponibles presentes en el glosario, ya sea dentro de un tema relacionado con el CI (por ejemplo, el tema del glosario “ropa” está, sin duda, directamente relacionado con el CI “la ropa”) o no (el tema “historias extraordinarias” no está relacionada con el CI “los animales”, pero, de todas maneras, recoge el término *araña*).

4.3.2.2 *Contenido del glosario*

En esta sección del análisis se investigará el contenido del glosario temático adjuntado a la unidad didáctica correspondiente al tema del CI y este se comparará con la LP0,1. En primer lugar, se consultarán las partes del glosario explícitamente relacionadas con el CI respectivo (por ejemplo, el apartado del glosario titulado “partes del cuerpo”, dentro de la unidad “8. Más vale prevenir”, en el caso del CI “el cuerpo humano”) y, en segundo lugar, las unidades cuyos glosarios abarcan el vocabulario menos relacionado con los CI, pero aún suficientemente relevante para el tema (por ejemplo, el apartado “carne”, en la unidad “9. ¡Que aproveche!”, en relación con el CI “los animales”).

Se averiguará el número y el porcentaje de las palabras coincidentes entre el glosario y la LP0,1, así como la cantidad de las palabras propuestas por la lista que no están recogidas en dicho glosario. También se indicará la proporción de palabras presentes en el glosario temático, pero que presentan ID demasiado bajo para estar incluidas en la LP0,1. Así, se podrá comprobar si los resultados de la publicación de DL concuerdan con el léxico que los autores del manual consideran suficientemente importante como para ser incluido en el glosario del tema correspondiente.

4.3.2.3 *Presentación de las palabras disponibles en el manual*

En la última sección del análisis, se estudiará la presentación de las palabras de la LP0,1 en el manual. En primer lugar, se procurará averiguar si el número de ocurrencias de las palabras de la lista es proporcional al valor de su ID²². Entre otras investigaciones que consideran importante el número de ocurrencias de la palabra, mencionemos, por ejemplo, la

²² En realidad, no se tendrá en cuenta el valor absoluto del ID de la palabra, sino su posición en la LP0,1, donde las palabras se presentan en orden decreciente según el valor de su ID.

de Hidalgo Gallardo (2017, p. 574), quien estudia este aspecto de manera marginal: el autor separa las palabras que presentan 5 o más ocurrencias de las demás (2017, p. 582).

En segundo lugar, se fusionarán los resultados de varias de las secciones anteriores: la presencia y la ausencia de las palabras, su aparición en el glosario, el número de sus ocurrencias y se añadirá el criterio de la distribución de las ocurrencias a lo largo del curso. Es decir, no solo se prestará atención al aspecto cuantitativo, sino también al modo de presentar las palabras. Para evaluar el modo de la presentación de las palabras, de manera aproximada, se opta por una categorización simple: en 6 clases, con el fin de facilitar la representación gráfica de los resultados. Primero, se delimitan las palabras ausentes por completo (categoría 1) y el resto se divide entre las palabras que están presentes, pero no recogidas en el glosario (categorías 2 y 3) y las que, al contrario, aparecen en él (categorías 4-6). A su vez, las palabras recogidas o no en el glosario se dividen entre las que presentan de 1 a 4 ocurrencias (categorías 2 y 4) y las que aparecen 5 o más veces (categorías 3, 5 y 6). Para indicar, de modo aproximado, la medida de la distribución de las unidades a lo largo del manual, se delimitan, además, las unidades que no solo están en el glosario y se repiten abundantemente, sino que además se distribuyen en diferentes secciones (o unidades) del libro (categoría 6).

A continuación, se propone un resumen de las categorías:

Ausentes	No presentes en el glosario		Presentes en el glosario		
	1-4	5 y más	1-4	5 y más	
				No distribuidas	Distribuidas
1	2	3	4	5	6

Tabla 1. Diagrama de la clasificación de las palabras según su presentación

Los criterios de clasificación parten de varias fuentes. De hecho, el número límite de ocurrencias se basa parcialmente en el número propuesto por el anteriormente citado Hidalgo Gallardo, pero también proviene del número mencionado en el capítulo anterior, citado por Thornbury: 7 repeticiones. No obstante, se supone que los alumnos no solamente ven las palabras escritas, sino que también están expuestos a ellas cuando las leen en voz alta, las oyen y hasta cuando las usan en forma escrita u oral. Por eso, se considera suficiente establecer el número límite en 5. A continuación, el criterio de la presencia en el glosario se ve priorizado (en detrimento del criterio del número de ocurrencias) con base en la suposición de que el vocabulario recogido por el glosario es el que el estudiante estudia de forma más concisa.

5 Resultados

Antes de presentar los resultados del análisis, aquí se presenta la lista de palabras que obtuvimos aplicando el corte de medida 0,1 (LP0,1) para los CI escogidos:

El cuerpo humano	La ropa	Los animales
brazo	pantalón	perro
ojo	camisa	gato
cabeza	jersey	león
mano	camiseta	vaca
pierna	calcetín	caballo
pie	falda	tigre
dedo	chaqueta	elefante
nariz	calzoncillo	ratón
oreja	braga	cerdo
boca	zapato	oveja
rodilla	abrigo	gallina
cuello	cazadora	jirafa
uña	bufanda	conejo
codo	sujetador	serpiente
pelo	media	mono
hombro	guante	ballena
tobillo	gorro	pájaro
muñeca	sudadera	pez
pecho	chándal	burro
espalda	chaleco	oso
diente	blusa	toro
tronco	vestido	lobo
lengua	bota	tiburón
labio	zapatilla	cabra
cadera	(pantalón) vaquero	águila
corazón	corbata	canario
ceja	tanga	delfín

estómago	pijama	zorro
antebrazo	polo	loro
cara	sombrero	periquito
culo	cinturón	rata
pene	chubasquero	cebra
pulmón		tortuga
muslo		pantera
pestaña		leopardo
hígado		liebre
		gallo
		rinoceronte
		hipopótamo
		cocodrilo
		ciervo
		hormiga
		pato
		hámster
		jabalí
		araña
		mosca
		avestruz
		abeja
		cigüeña
		pollo

Tabla 2. Lista de palabras con el $ID \geq 0,1$ (LP0,1)

Más adelante se presentan los resultados de los análisis constituyentes de la investigación.

5.1 Presencia de las palabras disponibles en el manual

En la siguiente tabla se puede observar el resumen general de la presencia de las palabras disponibles en el manual para cada CI (sin que importe el número de sus apariciones o el tema tratado) además de su presencia en cualquiera de los glosarios adjuntados a cada unidad del libro:

CI	El cuerpo humano	La ropa	Los animales	Total
ID \geq 0,1	36	32	51	119
Presentes (% presentes)	28 (77,8 %)	26 (81,3 %)	41 (80,4 %)	95 (79,8 %)
Glosario % glosario	24 (66,7 %)	24 (75,0 %)	30 (58,8 %)	78 (65,5 %)

Tabla 3 Presencia del léxico disponible en el manual y su glosario

En cifras totales, de las 119 palabras de la LP0,1, 95 (79,8 %) aparecen al menos una vez en el manual y 78 (65,5 %) están presentes en su glosario.

Por su parte, el CI “el cuerpo humano” está representado por 28 palabras (77,8 %) de las 36 más disponibles, de las cuales 24 (66,7 %) están incluidas en el glosario. Un número parecido, 26 (81,3 %) de las 32 palabras de la LP0,1, es el perteneciente al CI “la ropa”, con 24 (75,0 %) unidades presentes en el glosario, con lo que este CI alcanza los porcentajes más altos de los tres en ambos aspectos, pero no supera los demás CIs en cifras absolutas. El CI con el mayor número de unidades en la LP0,1 es el de “los animales”, que también cuenta con el mayor número de unidades presentes en el manual, 41 (80,4 %) de las 51, y el glosario contiene 30 (58,8 %) de ellas. Además, le corresponden los valores porcentuales más bajos.

Para el CI “el cuerpo humano”, las palabras ausentes por completo son: *muñeca*, *tronco*, *ceja*, *antebrazo*, *culo*, *pene*, *muslo* y *pestaña*. Estas palabras designan, entre otros, diferentes partes de la cara (*ceja*, *pestaña*), de las extremidades (*muñeca*, *antebrazo*, *muslo*), una parte del cuerpo más amplia (*tronco*) o tienen significado principalmente sexual o vulgar (*pene*, *culo*). Las palabras presentes, pero excluidas del glosario, son: *corazón*, *pulmón* e *hígado*, las cuales designan órganos internos, y *cara*.

Las palabras pertenecientes a la lista del CI “la ropa” ausentes en el manual son: *braga*, *sujetador*, *chaleco*, *tanga*, *polo* y *chubasquero*. Se omiten las prendas de carácter misceláneo (*chaleco*, *chubasquero*, *polo*, *tanga*). Las palabras excluidas del glosario son: *pijama* y *cinturón*.

Las unidades del CI “los animales” que no figuran en el manual son las siguientes: *canario*, *periquito*, *pantera*, *leopardo*, *liebre*, *rinoceronte*, *hormiga*, *jabalí*, *mosca* y *cigüeña*, de las cuales algunas designan diferentes especies de pájaros (*canario*, *periquito*, *cigüeña*), de insectos (*hormiga*, *mosca*) y de félidos (*pantera*, *leopardo*). Las palabras no recogidas por el glosario son: *perro*, *león*, *tigre*, *serpiente*, *mono*, *burro*, *cebra*, *gallo*, *hipopótamo* y *hámster*.

Como ya se ha destacado, los resultados de este trabajo se han obtenido analizando el manual como conjunto, más que analizando individualmente cada uno de los niveles. Sin

embargo, la siguiente tabla señala la presencia (es decir, al menos una ocurrencia) de las palabras disponibles en cada uno de ellos:

CI	El cuerpo humano	La ropa	Los animales	Total
<i>Nueva Aventura 1</i>	10	6	16	32
<i>Nueva Aventura 2</i>	26	25	37	88
<i>Aventura 3</i>	22	14	15	51

Tabla 4. Palabras presentes en los tomos del manual

Aunque todos los niveles abarcan algunas de las palabras disponibles, son notables las diferencias tanto entre los valores totales de los tomos, como entre los CI dentro de cada nivel. El tomo con el mayor número de palabras disponibles presentes es, de lejos, *NA2* (nivel A2-B1), seguido por *NA3* (nivel B1). Esta supremacía de *NA2* se debe principalmente al enorme número de unidades en la LP0,1 del CI “los animales” presentes en ese tomo, pero también a que es el tomo con el mayor número de unidades de los dos CI restantes. Asimismo, advierto de la posible (y evidente) repetición de las palabras en los tres manuales, por lo cual la suma de los totales (171) no es igual a la suma de las unidades constituyentes de la LP0,1 (119).

En resumen, al observar el atributo binario de la mera presencia o ausencia de las palabras constituyentes de la LP0,1 en los tres tomos del manual, sin tomar en cuenta el número de ocurrencias, la tasa media de presencia de un 79,9 % puede ser considerada relativamente alta. Así, cuatro de cada cinco palabras disponibles tienen cierta representación en el manual. Las diferencias entre los valores porcentuales entre los diferentes CI son minúsculas. Respecto a la presencia de las unidades a lo largo del curso, sería poco probable que todos los tomos contengan todas las palabras propuestas, o que cada manual contenga el mismo número de palabras de la lista distintas. Todo lo contrario, los números relativos a la presencia a lo largo de los tres tomos de cada manual resultan muy inconsistentes. Es posible presumir que el manual con el mayor número de unidades presentes es el que introduce las áreas temáticas de los CI en cuestión, pero, para comprobarlo, es necesario llevar a cabo un análisis adicional.

Por otra parte, respecto al aspecto cualitativo, las palabras ausentes principalmente representan hipónimos (*cigüeña*) o componentes (*muslo*) de las palabras más disponibles. Sin embargo, eso no implica la ausencia total de los hipónimos ni componentes, dado que aparecen, por ejemplo, las palabras *avestruz*, *boca* y *vaquero*. Así, se puede presumir que no se hizo una eliminación total de subtipos ni hipónimos, sino que su selección estuvo basada en ciertas reglas o simplemente en la intuición de los autores.

A pesar de la cantidad considerable de las unidades en la LP0,1 presentes en el manual, los valores porcentuales de aquellas contenidas en el glosario no son tan favorables, sobre todo en el CI “los animales”; menos desfavorables son los datos obtenidos para el CI “el cuerpo humano” y el CI “la ropa”, que es el que mayor porcentaje recoge en su glosario. Como ya se ha señalado, se supone que las palabras contenidas en el glosario son las que forman el vocabulario y, por lo tanto, se debería revisar su composición.

5.2 Contenido del glosario

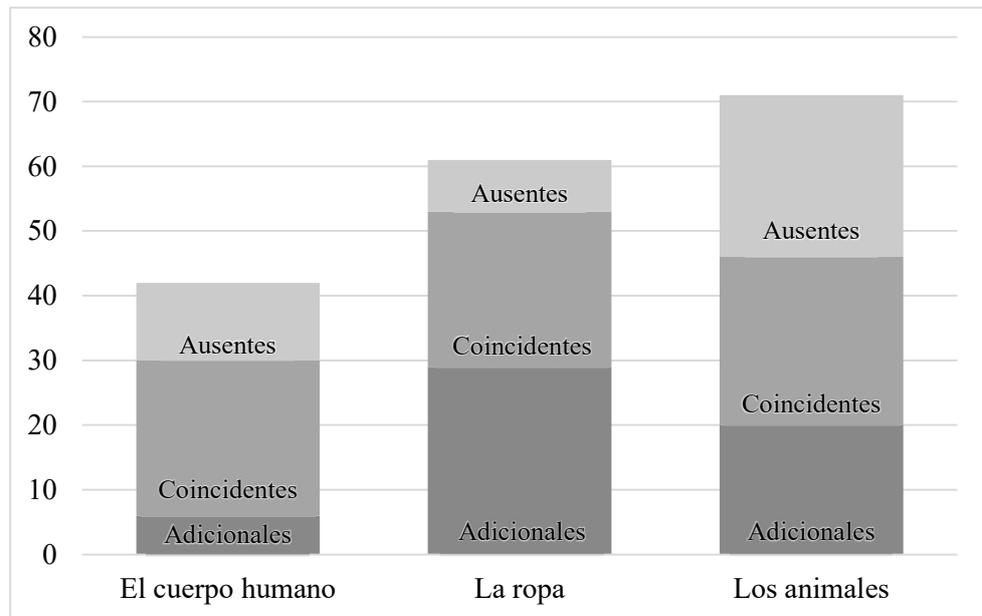
Para observar la presencia de las palabras propuestas por la LP0,1 en el glosario del manual, se han consultado las unidades cuyos glosarios tratan las áreas temáticas correspondientes a los CI escogidos: “8. Mi casa es tu casa” de *Nueva Aventura 1*, “8. Más vale prevenir”, “2. De tiendas” y “3. Vaya fauna” de *Nueva Aventura 2*; y las unidades cuyos glosarios abarcan el vocabulario estrechamente relacionado con los CI en cuestión: “5. Estos somos nosotros” y “9. ¡Que aproveche!” de *Nueva Aventura 1*. Los datos cuantitativos están recogidos en la tabla 5:

CI	El cuerpo humano	La ropa	Los animales	Total
Total ID $\geq 0,1$	36	32	51	119
Total glosario	30	53	46	129
Coincidentes	24	24	26	74
Ausentes	12	8	25	45
Adicionales	6	29	20	55
% de la LP0,1	66,7 %	75,0 %	51,0 %	62,2 %
% del glosario	80,0 %	45,3 %	56,5 %	57,4 %

Tabla 5. Presencia del léxico disponible en el glosario temático del manual

A primera vista, el total de las palabras abarcadas por el glosario (129) no difiere mucho del de la lista (119). Para los CI individuales, las cifras también son relativamente similares (en el caso de “el cuerpo humano” y “los animales”) o hasta las unidades del glosario superan el número de la LP0,1 (“la ropa”). Sin embargo, hay que recordar que estas cifras, a pesar de alcanzar niveles altos, no representan las mismas palabras. Por eso, lo que nos interesa más son las palabras coincidentes, cuyas sumas son iguales para los CI “el cuerpo humano” y “la ropa” (24) y es muy parecida en el caso de “los animales” (26).

La tabla 5 recoge, además del número de las palabras coincidentes, aquellas que presentan el ID $\geq 0,1$, pero están ausentes del glosario, y las recogidas en el glosario, pero que no están presentes en la LP0,1 (adicionales). Para contextualizar las relaciones entre las palabras coincidentes, ausentes y adicionales, consultemos la siguiente gráfica:



Gráfica 1. Proporciones de las palabras coincidentes, ausentes y adicionales

La gráfica 1 permite observar visualmente tres situaciones distintas: en el caso del CI “el cuerpo humano”, las palabras ausentes del glosario superan las adicionales, pero el conjunto de las coincidentes forma la parte más amplia, mientras que una situación contraria ocurre en el CI “la ropa”, dado que el número de las palabras ausentes corresponde a una fracción de las adicionales, que superan las coincidentes. Por fin, el CI “los animales” presenta mayor cantidad de palabras ausentes que adicionales, pero la diferencia entre estos dos conjuntos es relativamente pequeña.

Combinando, ahora, los datos de la tabla 5 y la gráfica 1, el total de las palabras en el glosario (129) sobrepasa el número de las palabras de la lista (119) por 10 unidades, pero solo 74 de ellas coinciden, número que corresponde al 62,2 % de la lista de las palabras disponibles y solo al 57,4 % de las unidades recogidas por el glosario. Como ya se ha dicho, a nivel de los CI individuales, ocurren situaciones distintas: En el caso del CI “el cuerpo humano”, hasta el 80 % del glosario está formado por las palabras disponibles, es decir, con el $ID \geq 0,1$, pero también hay bastantes palabras ausentes, dado que solo el 66,7 % de las palabras de la LP_{0,1} está presente. Por otro lado, los valores porcentuales del CI “la ropa” están invertidos. De hecho, el glosario cubre el 75 % de la lista, pero estas palabras solo constituyen el 45,3 % del glosario. En otras palabras, más de la mitad de las palabras propuestas por el glosario está formada por palabras con el ID menor a 0,1. En cuanto al CI “los animales”, los valores porcentuales presentan una gran proximidad entre sí. Es decir, las 26 palabras coincidentes forman el 56,5 % del glosario y, a la vez, corresponden al 51 % de la lista.

Respecto al carácter cualitativo, en la siguiente tabla se presenta el contenido del glosario, junto con el ID respectivo para cada palabra²³:

El cuerpo humano		La ropa		Los animales	
NA1, U5: "Estos somos nosotros"		NA1, U5: "Estos somos nosotros"		NA1, U9: "¡Que aproveche!"	
"Aspecto"		"Aspecto"		"Carne"	
barba	0	gafas	0,002	(carne de) cerdo	0,31
bigote	0,001	llevar	0	(carne de) ternera	0,03
ojo	0,82	pendiente	0,004	(carne de) vaca	0,55
pelo	0,34	sombrero	0,12	pollo	0,1
NA2, U8: "Más vale prevenir"		NA1, U8: "Mi casa es tu casa"		"Pescado"	
"Partes del cuerpo"		"Ropa"		atún	0,01
barriga	0,05	bañador	0,009	bacalao	0,01
boca	0,58	camiseta	0,75	gamba	0,01
brazo	0,85	pantalones	0,94	marisco	0,002
cabeza	0,82	zapatillas	0,2	merluza	0,01
cadera	0,18	"Muebles"		salmón	0,04
cintura	0,098	armario	0,002	"Otros alimentos"	
codo	0,41	cajón	0	huevo	0
cuello	0,45	NA2, U2: "De tiendas"		NA2, U3: "¡Vaya fauna!"	
dedo	0,74	"Tienda de deportes"		"Animales"	
diente	0,27	gorro de baño	0	abeja	0,12
espalda	0,28	"Joyería"		águila	0,2
estómago	0,17	anillo	0	avestruz	0,12
frente	0,07	collar	0,0004	ballena	0,26
garganta	0,02	pendiente	0,004	búho	0,09
hombro	0,33	"De compras"		caballo	0,52
labio	0,18	marca	0,005	cabra	0,21
lengua	0,21	probador	0,003	cerdo	0,31
mano	0,82	probar	0	ciervo	0,13
nariz	0,72	talla	0	cocodrilo	0,13
oreja	0,67	"Prendas de vestir"		conejo	
pecho	0,28	abrigo	0,42	elefante	0,41
pie	0,76	blusa	0,21	gallina	0,29
pierna	0,79	botas	0,2	gato	0,99
rodilla	0,5	bufanda	0,38	iguana	0,05
tobillo	0,32	calcetines	0,71	insecto	0,04
uña	0,42	calzoncillos	0,5	jirafa	0,29
		camisa	0,85	lobo	0,23
		cazadora	0,38	lobo marino	0

²³ Los valores se han redondeado a dos decimales, salvo en los casos donde el número alcanzaría, tras el redondeo, el umbral de 0,1 o, al contrario, se reduciría al valor vago de 0,0. En esos casos, se opta por el redondeo al siguiente decimal posible. Esta decisión se ha tomado para facilitar la orientación, dado que en este momento no nos interesa el valor exacto, sino que los valores mostrados sirven solamente para contextualizar las unidades determinadas y compararlas entre sí. A las palabras que no están en absoluto presentes en el diccionario de disponibilidad para el CI determinado, se les ha asignado el valor 0.

corbata	0,16	oso	0,24
chándal	0,25	oveja	0,31
chaqueta	0,51	pájaro	0,25
falda	0,54	pato	0,13
gorro	0,28	pingüino	0,06
guante	0,33	rana	0,096
jersey	0,83	ratón	0,31
medias	0,34	tortuga	0,15
pantalones cortos	0,08	vaca	0,55
slips	0,04	zorro	0,17
sudadera	0,27	"Mundo animal"	
traje	0,06	cuernos	0
vaqueros	0,19	hembra	0
vestido	0,2	macho	0
zapatos	0,47	mamífero	0,03
zapatos de tacón	0	nido	0
"Hablar de ropa"		pata	0
de colores	0,002		
de lunares	0		
elegante	0,008		
cómodo	0		
formal	0		
informal	0,0009		
llevar	0		
manga	0,005		
ponerse	0,001		
quedar bien	0		
tacón	0,002		

Tabla 6. Palabras presentes en el glosario del manual

Es evidente que los subgrupos que tienen la menor cantidad de palabras adicionales son los estrechamente enlazados con determinadas áreas temáticas, es decir, “partes del cuerpo”, “prendas de vestir” y “ropa” (pertenecientes a dos unidades distintas) y “animales”. Dada la extensión de estos subgrupos, se supone que son los que más palabras de la LP0,1 abarcan. Veamos, ahora, el contenido léxico de cada uno de los temas.

Dentro del CI “el cuerpo humano” y la parte del glosario correspondiente, “partes del cuerpo”, las palabras adicionales incluyen un tracto interno, *garganta*, cuya presencia se puede considerar funcional en el tema de la salud (“más vale prevenir”) y otras palabras que designan las partes del cuerpo (*barriga*, *frente* y *cintura*). El tema “aspecto” contiene estas palabras poco disponibles: *barba* y *bigote*, ambas referentes al vello facial, elementos que no son prototípicos ni universales en el cuerpo humano. Sin embargo, se trata de dos características que pueden ser importantes en la descripción física de los hombres adultos. Respecto a las palabras ausentes del glosario, aparte de las que no se presentan en ningún

momento del curso y, por lo tanto, citadas ya en el subcapítulo anterior, se documentan solamente órganos internos (*corazón, pulmón e hígado*) y *cara*.

En el caso del CI “la ropa”, el subgrupo de palabras correspondiente, “prendas de vestir”, contiene las siguientes palabras adicionales: un tipo de pantalones (*pantalones cortos*), de ropa interior (*slips*), de zapatos (*zapatos de tacón*) y un atuendo formal (*traje*). En los demás subgrupos, las palabras adicionales hacen alusión, sobre todo, a accesorios (*gafas, pendiente, anillo y collar*) y algunas prendas de uso ocasional (*bañador y gorro de bañar*), pero también a palabras cuya relación con el tema de la ropa tiene un menor valor denotativo o son más específicos (*llevar, marca, probador, probar, talla, de colores, de lunares, elegante, cómodo, formal, informal, manga, ponerse, quedar bien, tacón*). Sin embargo, también se clasifican como adicionales palabras más alejadas de este campo semántico y más próximas al tema de la casa (*armario y cajón*). Las palabras ausentes del glosario, aparte de las no recogidas por el manual en absoluto, incluyen *cinturón*, un accesorio, y *pijama*, una prenda de uso diario.

En relación con “los animales”, el CI con la mayor discrepancia de los tres, las palabras adicionales en el subgrupo dado, “animales”, son estas: *búho, iguana, insecto, lobo marino, pingüino y rana*. Algunos de ellos son hiperónimos (*insecto*) o un tipo de pájaro (*búho, pingüino*), pero también está presente la palabra *lobo marino*, que ni siquiera está presente en el diccionario de disponibilidad en cuestión. Las demás palabras adicionales, igual que en el caso del CI “la ropa”, están relacionadas con los animales sin denotarlos (*huevo, cuernos, hembra, macho, mamífero, nido y pata*) o se usan más en el contexto de la comida (*ternera, atún, bacalao, gamba, marisco, merluza, salmón* y, otra vez, *huevo*). Se registra una cantidad mayor de palabras ausentes del glosario que, aun así, están presentes en alguna parte del manual que en los demás CI: *perro, león, tigre, serpiente, mono, ballena, pez, burro, toro, tiburón, delfín, loro, cebra, gallo, hipopótamo, hámster, araña y pollo*.

En definitiva, hay que recordar que no solo se han analizado las partes del glosario correspondientes temáticamente a los centros de interés escogidos, sino que también se escogieron varios subgrupos de palabras en las unidades relacionadas, en cierta medida, con ellos. En efecto, en su conjunto el número de palabras recogidas en el glosario supera el número de las unidades en la lista de palabras cuyo ID es mayor o igual a 0,1. Mientras que en el caso del CI “cuerpo humano” el número de palabras coincidentes es más alto, este número es más bajo en comparación con los números de las palabras adicionales y ausentes para los CI “la ropa” y “los animales”. A pesar de analizar un número muy limitado de centros

de interés, se observa que el grado de acuerdo entre el glosario y la lista de disponibilidad para los CI aumenta proporcionalmente con el grado de cohesión.

La mayoría de las palabras adicionales, a pesar de pertenecer temáticamente al mismo campo semántico que el CI dado, no son los hipónimos de su título (por ejemplo, *talla* o *pata* no pertenecen a los conjuntos “la ropa” ni “los animales”, respectivamente). Sin embargo, está claro que estas palabras son de gran uso en las situaciones del tema respectivo. Otras unidades, aunque cumplan la característica que acabamos de citar, están más relacionadas con un tema correspondiente a otro CI (como *armario* o los diferentes tipos de pescado). Eso es, su disponibilidad posiblemente alcance niveles más altos en las listas de otros CI (como “comidas y bebidas” para los pescados), lo que justificaría su presencia en el glosario del manual, pero la verificación de esta hipótesis requeriría un análisis más exhaustivo. No obstante, en algunos casos es cuestionable la decisión de introducir las palabras en el glosario (como *cintura*, *slips* o *lobo marino*), mientras que otros parecen estar escogidos de manera aleatoria, sobre todo por la presencia de algunos animales (*pingüino*, *iguana*).

En este sentido, la ausencia de algunas de las palabras con el ID mayor a 0,1 en el glosario también es un aparente resultado de la selección aleatoria: como las diferentes especies de animales, teniendo en cuenta la inconsistencia en el ecotipo (por ejemplo, entran *abeja* e *insecto*, pero no así otros insectos probablemente más prototípicos como *araña* u *hormiga*, o la ausencia de los animales exóticos *tigre* y *león* a pesar de la presencia de *elefante*, etc.). Aparte de “los animales”, en los casos de los demás CI no hay un número tan alto de ausencias.

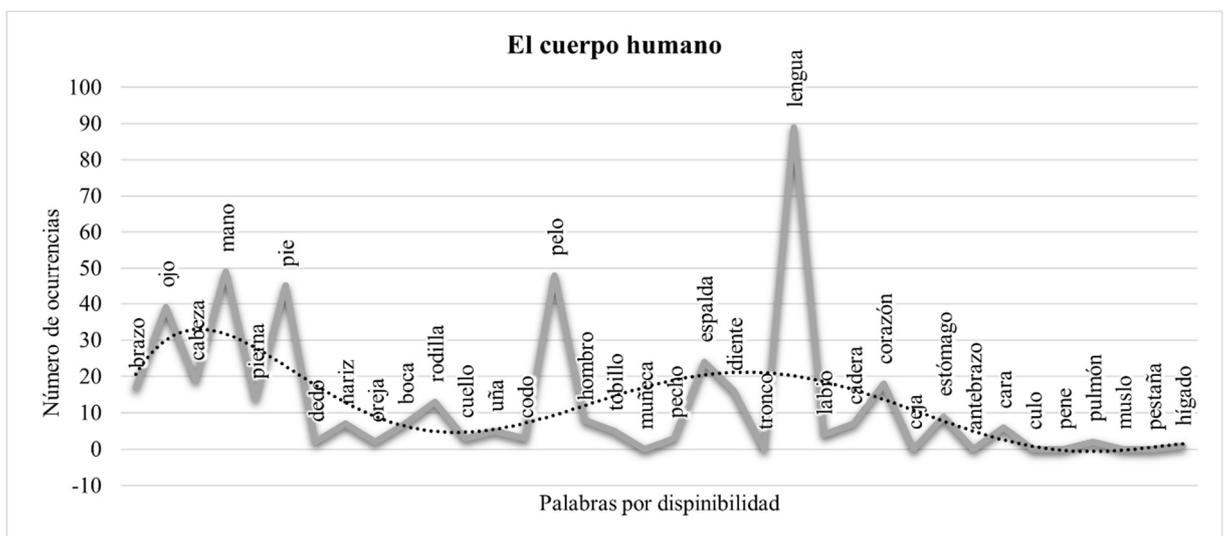
Es imprescindible recordar que esta parte del análisis no cuenta con las palabras sueltas que están presentes en una parte del glosario que no tiene relación con el tema de los CI tratados. Sin embargo, si la adquisición del vocabulario está facilitada por la exposición a las palabras de manera que se introduzcan dentro de sus respectivos campos semánticos, es cuestionable la omisión de algunas (poca) palabras altamente disponibles en los nativos, como *perro*, *bragas* o *corazón*.

5.3 La presentación de las palabras disponibles en el manual

En la siguiente parte del análisis, nos centramos en el modo de representación de las unidades pertenecientes a la lista de palabras con el ID mayor o igual a 0,1 (LP0,1). De un modo parecido a ambas revisiones anteriores, nos fijaremos tanto en la presencia de las

palabras en cualquier lugar del manual, como en su aparición en cualquiera de los glosarios adjuntados a las unidades didácticas del libro. En vez de un enfoque binario, nos fijamos en la medida en la que aparecen las palabras individuales en el manual, es decir, su número de ocurrencias. Para una comprensión más profunda y una orientación más fácil, se recomienda consultar los anexos 9.1, 9.2 y 9.3. Sin embargo, en las siguientes páginas se presentan los resultados del análisis.

Empezando con el CI “el cuerpo humano”, en la siguiente gráfica se presentan los resultados del análisis, siendo las palabras ordenadas de acuerdo con su posición en el listado de disponibilidad (de izquierda a derecha):



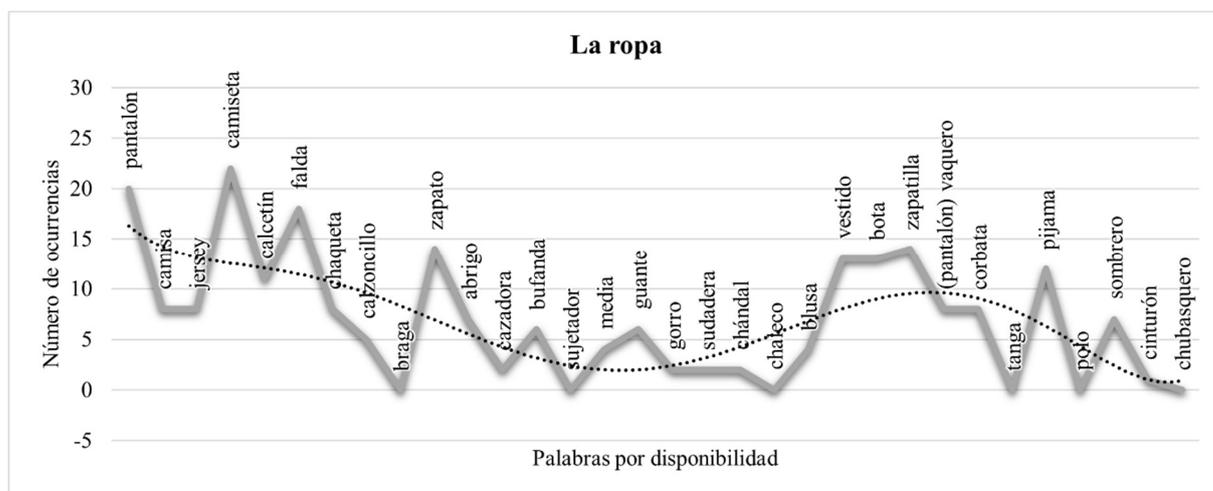
Gráfica 2. Número de ocurrencias de las palabras del CI “el cuerpo humano”

Aparte de la representación gráfica, son imprescindibles los datos numéricos. El mayor número de ocurrencias lo alcanza la palabra *lengua*, con un total de 89. Este número tan alto es posible por ser usada en diferentes contextos. Efectivamente, además de aparecer en el contexto de las partes del cuerpo, está presente en el glosario dentro de la expresión “sacar la lengua” y, además, la mayoría de sus apariciones se refieren al significado “idioma” (en la sección 4.3.1 ya indicamos que este tipo de palabras plurisemánticas se contarían siempre y cuando la relación entre el significado deseado y el alternativo fuera lo suficiente transparente). A esta palabra la siguen, con una diferencia notable en el número de ocurrencias, las palabras *mano* (49), *pelo* (48), *pie* (45) y *ojo* (39). Todas estas palabras, aparte del uso literal, es decir, el que se refiere a partes del cuerpo, forman parte de unidades fraseológicas (“¡pero ojo!”, “a mano”, “a pie”, etc.) o se emplean, además, en un contexto diferente pero relacionado con el cuerpo humano, como el aspecto físico (*ojo* y *pelo*). Las cuatro palabras presentan unos ID relativamente altos, sobre todo *ojo*, *mano* y *pie*. No

obstante, es significativo el impacto que ejercen las palabras productivas en la formación de las unidades fraseológicas, dado que otras palabras de ID muy parecido o hasta más alto, no llegan a un número de ocurrencias tan alto, por ejemplo, *brazo* (17), *cabeza* (19) o *pierna* (14). Por otro lado, aun las palabras con menor número de ocurrencias aparecen dentro de unidades fraseológicas, por ejemplo, *cara* (6) (“de muchas caras”) o *pulmón* (2) (“[gritar] a todo pulmón”).

La gráfica dispone, además de la curva de valores absolutos, de una línea de tendencias, que resume la propensión de las ocurrencias en relación con la disponibilidad decreciente de las palabras. Esta línea tiene dos picos: uno causado por unas cifras extraordinariamente elevadas, en el caso de *pelo* y *lengua*, y el otro producido por los números de ocurrencias relativamente altos (*ojo*, *mano* y *pie*) junto con la bajada causada por el número menor de ocurrencias en el caso de la palabra más disponible: *brazo*. En definitiva, la línea de tendencia disminuye, aunque no lo hace de manera uniforme. Por último, debemos señalar que la media del número de ocurrencias es 12,9.

Pasando al próximo CI, “la ropa”, observemos la siguiente gráfica:



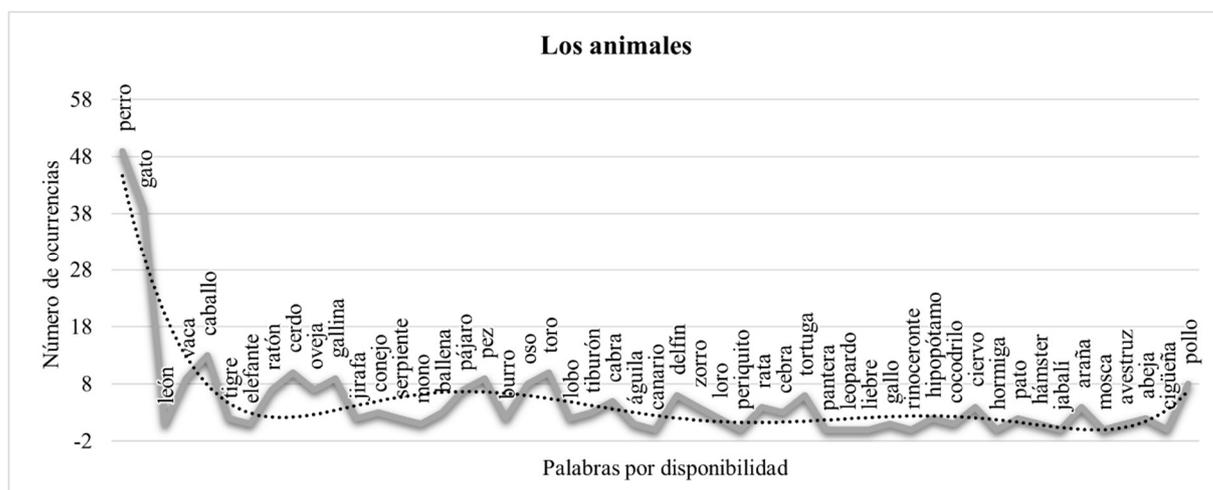
Gráfica 3. El número de ocurrencias de las palabras del CI “la ropa”

En comparación con el CI anterior, los números absolutos de ocurrencias no alcanzan niveles tan altos. De hecho, el valor máximo del número de ocurrencias está representado por la palabra *camiseta* (22), seguida con una mínima diferencia por *pantalón* (20) y *falda* (18). Sin embargo, la gráfica evidentemente no presenta ningún valor especialmente destacado en comparación con los valores vecinos. Las diferencias mostradas por dos prendas básicas y muy prototípicas, *pantalón* y *camiseta*, tampoco es sorprendente. Aunque las unidades de este CI no han sido tan afectadas por su uso en unidades fraseológicas, también hay casos de contextos especiales que aumentan el número de ocurrencias de la

palabra dada. Uno de los pocos casos es la palabra *falda*, que es usada, por ejemplo, en el compuesto “minifalda”, mientras que *pijama* (12) aparece varias veces y repetidamente en la letra de una canción (“No me [gustaba] el pijama, el pijama”).

A su vez, la línea de tendencias sigue el modelo del CI “el cuerpo humano” y se desarrolla de forma parecida: se forma un valle, causado por valores más bajos de ocurrencias alcanzados por las palabras de ID relativamente alto (especialmente las no recogidas en absoluto por el manual, *braga*, *sujetador* y *chaleco*), entre dos picos de la curva. El segundo pico demuestra, al contrario, el aumento de ocurrencias de varias unidades de ID relativamente bajo, sobre todo *vestido* (13), *bota* (13), *zapatilla* (14) y *pijama* (12). Resumiendo, la media del número de ocurrencias es 7²⁴ y, de manera similar al CI anterior, la gráfica muestra una tendencia descendente, aunque esta tampoco es constante.

Para el último CI, “los animales”, la siguiente gráfica presenta varios desafíos:



Gráfica 4. Número de ocurrencias de las palabras del CI “los animales”

De manera parecida a los valores de “el cuerpo humano”, hay un pico muy pronunciado, dado que las primeras dos palabras con el mayor ID presentan un número extraordinariamente alto de ocurrencias: *perro* (49) y *gato* (39). Es interesante la diferencia abismal entre estas y la tercera palabra más disponible: *león* (0). Sin embargo, esa palabra no es la única “culpable” del valle formado en la línea de tendencias. De hecho, la tercera palabra con mayor número de ocurrencias es *caballo* (13), número muy alejado del máximo. Si bien este CI no es tan productivo en la formación de los fraseologismos (aunque se documentan algunos, como “cada *oveja* con su pareja”), varias de las palabras se emplean refiriéndose a

²⁴ Redondeado a los decimales.

los animales como a su carne, es decir, en el tema de la comida, entre ellos *cerdo* (10), *vaca* (9), *pollo* (8) y hasta *caballo* (13). Otras veces, las unidades no se refieren al animal, sino a un objeto que se asemeja a él, como *oso* (8) en la expresión “osito de peluche”, pero también *ratón* (7) en el tema de la tecnología y *cebra* (3) en el compuesto “paso de cebra”.

Ya se ha destacado el número notable de unidades de la LP0,1 en este CI documentadas en el manual, pero aquí se puede observar que la medida de ocurrencias es generalmente muy baja para las palabras individuales. En efecto, la media del número de ocurrencias es solo 5; sin embargo, es necesario considerar que esa cifra está afectada tanto por las palabras con el valor especialmente alto como por las que no tienen ninguna ocurrencia en absoluto.

Aparte del pico inicial seguido por un leve valle, la tendencia es claramente descendente, propensión esta que queda perturbada por la “anomalía” en la última palabra, *pollo*, y su empleo frecuente en el contexto de la comida. Esta relativa uniformidad del decrecimiento en la tendencia se debe, sobre todo, a la cantidad de palabras que no tienen ninguna ocurrencia en el manual, cuyo número crece hacia el final del eje horizontal, esto es, con la disminución del ID.

Por último, para sintetizar los resultados previamente expuestos, se recurre al estudio aproximado de los factores que facilitan la adquisición del vocabulario: la presencia en el glosario, el número de ocurrencias y la eventual distribución de estas entre varias secciones del manual. Es decir, se observa la presentación de cada palabra individual y, también, se evalúa la presentación de su conjunto.

Primero se aplica el criterio del glosario, después, la palabra se clasifica según el número de ocurrencias y, al final, se separan las que están distribuidas en 3 o más unidades didácticas. Avisamos nuevamente de que esta parte del análisis es una aproximación y su propósito es dar una visión general del tema.

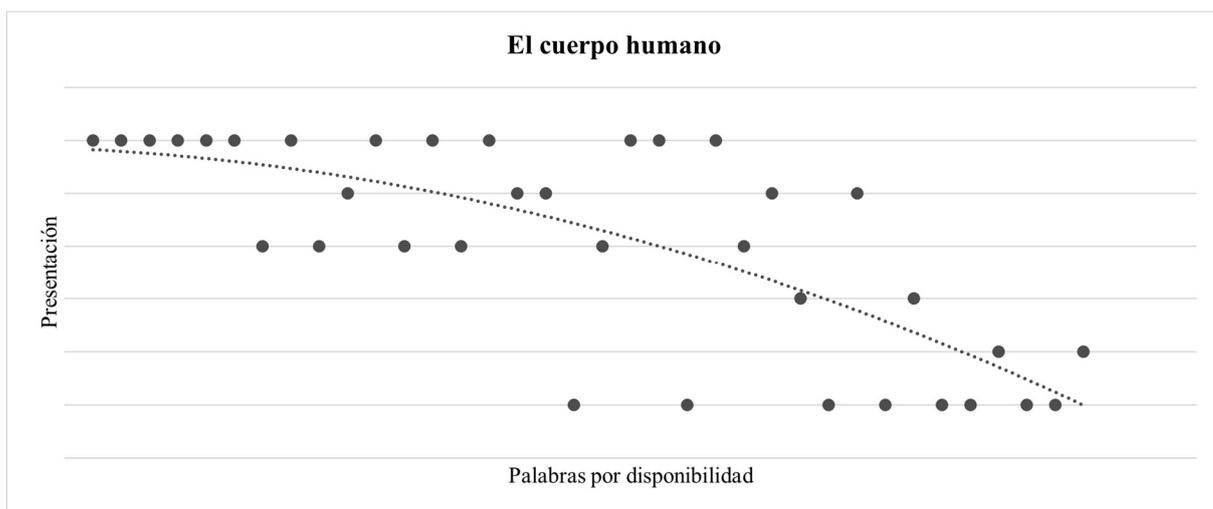
La clasificación de las palabras y su consecuente evaluación se ejercen según las categorías apuntadas en la tabla 1, ya presentada anteriormente en la sección 4.3.2.3²⁵:

²⁵ Para más detalles, se recomienda consultar los anexos 9.1, 9.2 y 9.3.

Ausentes	No presentes en el glosario		Presentes en el glosario		
	1-4	5 y más	1-4	5 y más	
				No distribuidas	Distribuidas
1	2	3	4	5	6

Tabla 1. Diagrama de la clasificación de las palabras según su presentación

En la siguiente gráfica, se puede observar la dispersión²⁶ de las palabras disponibles pertenecientes al centro de interés “el cuerpo humano”:

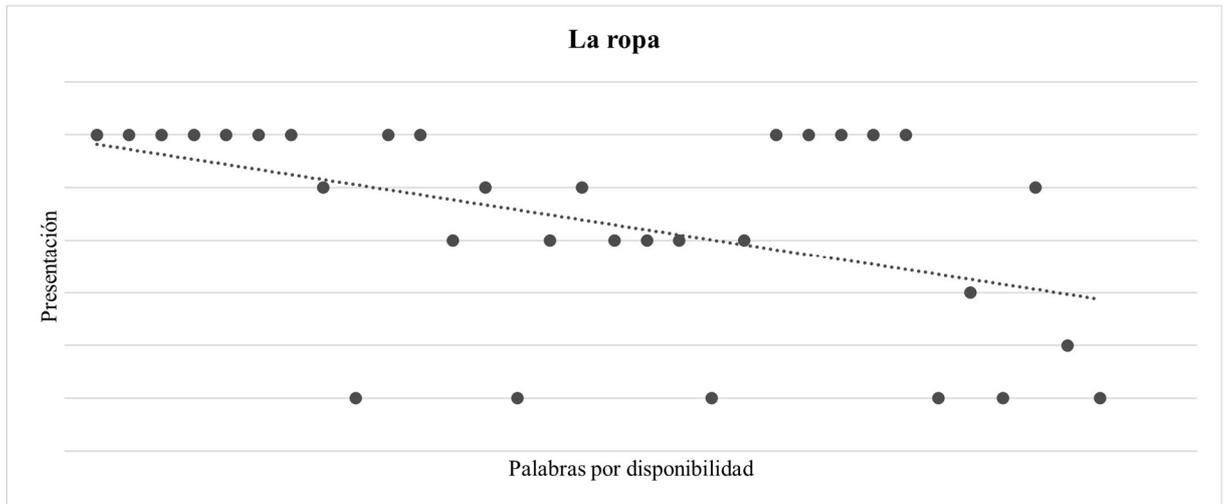


Gráfica 5. Representación gráfica de la clasificación del CI “el cuerpo humano”

A primera vista, resulta evidente que, con la disminución de la disponibilidad, cada vez más palabras pertenecen a las categorías más bajas, es decir, que cumplen menos factores de adquisición efectiva. Aproximadamente la primera mitad de las palabras encajan en una de las categorías más altas (4-6), dado que están presentes en el glosario, con menos o más ocurrencias, distribuidas o no. Además, se nota que las primeras 6 palabras, las más disponibles, están distribuidas por varias unidades del manual, además de presentar 5 o más ocurrencias y estar recogidas por el glosario. La abundancia de palabras ausentes por completo aumenta con el decrecimiento de la disponibilidad, pero omitiendo estas, la gráfica se vuelve aún más equilibrada, dado que, hasta el final de la lista de disponibilidad, las palabras se clasifican en los primeros grupos (1-3), esto es, no están presentes en el glosario. Así, junto con las palabras ausentes, producen que la curva de tendencia descienda con cierta intensidad.

²⁶ El eje horizontal indica, como en las gráficas anteriores, las palabras de la LP0,1 ordenadas por el índice de disponibilidad, mientras que el eje vertical indica la pertenencia a la categoría de 1 a 6 de acuerdo con la Tabla 1.

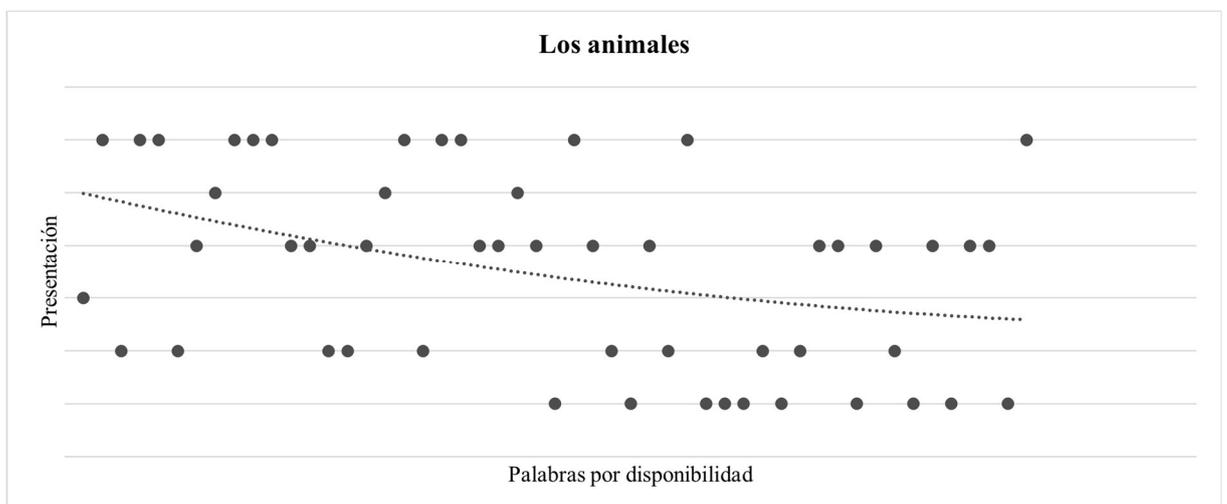
A continuación, podemos observar la gráfica del CI “la ropa”, que presenta resultados distintos:



Gráfica 6. Representación gráfica de la clasificación del CI “la ropa”

La peculiaridad de este CI es el número bajo de palabras ausentes en combinación con el número elevado de palabras presentes en el glosario. Además de ser estas numerosas, una gran parte de ellas tiene 5 o más ocurrencias: algunas están acumuladas en una unidad del manual o dos, pero hay un número considerable que están distribuidas por varias unidades. Estas palabras, las encasilladas en la última categoría, se presentan en la gráfica en dos cúmulos mayores, correlativos a los picos en la gráfica 6. Es gracias a estas, y a pesar de las ausentes, que el decrecimiento de la línea de tendencia es moderado. Por eso, hay pocas unidades situadas en la porción inferior de la gráfica.

Para cerrar el análisis, se presenta la gráfica que representa los datos obtenidos para el CI “los animales”:



Gráfica 7. Representación gráfica de la clasificación del CI “los animales”

A pesar de la estabilidad relativa de la curva en el análisis del número de ocurrencias, se nota que, en la categorización de las palabras, no hay tanta discrepancia entre las primeras posiciones y el resto de la lista. Por otro lado, hay una enorme variación a medida que avanzamos en la lista de palabras. Efectivamente, aparte de las ausencias (unidades pertenecientes a la categoría 1) que se hacen abundantes aproximadamente en la mitad de la lista, la gráfica muestra que la correlación entre la disponibilidad y la presentación de las palabras es minúscula. Es positivo que la mayoría de las palabras pertenecen a las categorías 4-6, es decir, están abarcadas por el glosario, y muchas de ellas presentan 5 o más ocurrencias distribuidas en el manual. De todos modos, a pesar de la dispersión heterogénea, la curva tiende a decrecer. No obstante, esa realidad se debe, sobre todo, a la disminución de las palabras en la categoría 6 y el crecimiento repentino en la 1, es decir, de las que se hallan ausentes por completo.

En resumen, la tendencia general presenta un descenso en la frecuencia de ocurrencias a lo largo de la lista de palabras, aunque hay variaciones notables en cada centro de interés. Se ha observado que, para algunos CI, a pesar de mostrar un gran número de palabras disponibles presentes en el manual, el número de ocurrencias es más bajo en las palabras individuales, como se puede observar en el CI “los animales”. Este CI presenta una característica peculiar, debido a la discrepancia entre las primeras dos palabras y las demás, mientras que el CI “la ropa” contiene los números de ocurrencias más homogéneos. En cuanto al estudio de los factores como la presencia de la palabra en el glosario, su número de ocurrencias y la posible distribución de ellas entre varias unidades didácticas, se ha observado una tendencia decreciente en los tres CI, pero el caso de “los animales” se ha revelado como el más problemático, dado que las formas de la presentación de las palabras poseen una gran cantidad de discrepancias entre las unidades vecinas, es decir, con el ID parecido. La tendencia del CI “la ropa”, al contrario, decrece de manera menos repentina, esto es, sus unidades se tienden a presentar de manera similar y, así, se presta mucha atención a una gran cantidad de palabras disponibles.

Es importante tener en cuenta que los resultados propuestos, sobre todo en la segunda parte de este último análisis, son aproximados y, para poder sacar conclusiones aún más precisas, resultaría necesario analizar las frecuencias de las ocurrencias a lo largo del curso del español de forma más detallada.

6 Conclusiones

Este trabajo se plantea la pregunta de si la selección del vocabulario en un manual checo de ELE, *Aventura*, es adecuada y refleja el lenguaje usado por los hablantes nativos de español. Como punto de partida y referencia se ha usado una investigación de disponibilidad léxica, que forma parte del Proyecto Panhispánico de la Disponibilidad Léxica (PPHDL), *El léxico disponible de Castilla y León* (Cruz Alonso, 2016), y se ha aplicado el corte de medida tradicional, 0,1, a su diccionario de disponibilidad para los centros de interés “el cuerpo humano”, “la ropa” y “los animales”. Se ha comparado la lista obtenida con el contenido léxico del manual, tomando en cuenta las palabras de la lista presentes en el manual y las ausentes, su número de ocurrencias y su presencia en el glosario del manual. Además de eso, se ha explorado, de modo aproximado, la distribución de las ocurrencias a lo largo del manual. Se han registrado todas las apariciones: desde los ejercicios de vocabulario hasta las instrucciones, incluyendo la mayoría de las unidades fraseológicas. Sin embargo, la selección de las apariciones válidas ha presentado varios desafíos, sobre todo en relación con el descarte de algunas palabras constituyentes de unidades fraseológicas y de algunas que denotan significados muy alejados del significado propio del CI. Se han seguido las pautas establecidas, pero, de todos modos, este procedimiento ha sido objeto de una evaluación subjetiva, ya que se ha evaluado la validez de cada ocurrencia de manera intuitiva.

Por un lado, se ha demostrado la presencia en el manual de un porcentaje alto de las palabras propuestas por la lista, con variación minúscula entre los centros de interés, de manera que el conjunto de las palabras ausentes consta principalmente de algunos de los hipónimos o componentes de las palabras presentes. A pesar de la cantidad de palabras disponibles presentes en el manual, los valores porcentuales de las unidades recogidas por el glosario del manual son poco favorables, y eso tomando en cuenta tanto las partes del glosario correspondientes a las áreas temáticas escogidas como las que abarcan palabras sueltas sin relacionarlas con el campo semántico en cuestión. Este hecho y la presencia de las unidades con el índice de disponibilidad más bajo implican la necesidad de una revisión del glosario proporcionado por el manual. Sin embargo, es innegable que algunos de los términos con el ID bajo son útiles en las situaciones relacionadas con el centro de interés en cuestión, lo que demuestra la necesidad de un componente subjetivo en la selección léxica para los materiales didácticos. Asimismo, no solo la abundancia de las coincidencias entre el glosario y la lista, sino también las ratios de las unidades ausentes y adicionales varían según el centro de interés. Por eso, la medida de desacuerdo entre las listas comparadas no es constante para todas las

áreas temáticas. No obstante, lo que todos los centros de interés tienen en común es el decrecimiento del número de ocurrencias en el manual en relación con la disminución de disponibilidad, aunque esta tendencia decreciente no se desarrolla de manera constante ni igual para cada uno de los centros de interés. En este sentido, tanto la posible presencia en el glosario como el número de ocurrencias y su posible distribución por el manual tienden a decrecer en relación con el índice de disponibilidad.

Los resultados citados implican varias posibles explicaciones: dado el elevado porcentaje de palabras de la lista presentes en el manual, una de las explicaciones sería el uso de la publicación en cuestión, *El léxico disponible de Castilla y León* (Cruz Alonso, 2016), o de otro listado de disponibilidad léxica a la hora de componer el manual. Sin embargo, esta opción es poco probable. Más aun, se ofrece la posibilidad de que el alto grado de coincidencia se deba a una casualidad o, mejor dicho, a la coincidencia entre lo que los autores, o los recursos de los que partieron, consideran imprescindible y lo que consideran importante los informantes del estudio, si bien de una manera subconsciente. Esta coincidencia, aunque sea accidental, es positiva y favorece la posible adquisición de la lengua realmente utilizada por los hablantes nativos. La discrepancia entre los glosarios temáticos y la lista, a su vez, posiblemente radique en gran medida en que los autores incluyen el vocabulario relacionado con las situaciones presentadas en el manual (por ejemplo, *probar* o *talla* en el tema de la ropa) y no de forma *abstracta* como sí hacen los informantes. Este hecho, junto con la omisión de algunas unidades muy disponibles, pero de uso cuestionable, demuestra el papel que desempeña la evaluación subjetiva en el proceso de la selección del vocabulario y la falta de fiabilidad de un enfoque puramente objetivo, aunque se use para ello una herramienta basada en la disponibilidad léxica. Por otro lado, resulta evidente la deficiencia del glosario, o del manual mismo, causada por la falta de algunos términos muy disponibles cuya utilidad podría considerarse significativa incluso basándose en un juicio subjetivo, lo que reafirma, al contrario, la indudable necesidad de una selección del vocabulario (al menos parcialmente) objetiva.

Habiendo declarado lo anterior, es importante destacar que los resultados de este trabajo pueden presentar distorsiones causadas por estos factores: primero, para realizar el análisis, solo se han escogido tres centros de interés de la publicación de disponibilidad léxica sobre la que se basa la investigación. En este sentido, la siguiente variable es la selección del listado de referencia: el uso de otra publicación o de una fusión de varias, dando lugar a una lista unificada, posiblemente produciría resultados distintos. Resulta complicado saber a qué

publicación recurrir a la hora de realizar tal análisis, pero esta incertidumbre puede solucionarse por la finalización y la eventual unificación de los resultados obtenidos por varios estudios del Proyecto Panhispánico de la disponibilidad léxica, que sirva de referencia tanto para los análisis de los materiales didácticos como para su creación.

Por otra parte, es común que el manual no sea el único recurso utilizado en el aula, sino que, por el contrario, los profesores suelen recurrir a diversas fuentes. Por lo tanto, es posible que haya oportunidades para abordar las carencias en el contenido del manual. En este sentido, el presente trabajo de fin de máster puede servir como una guía de análisis, de modo que, con base en sus resultados, los profesores podrían compensar las deficiencias de los manuales utilizados en sus cursos.

7 Resumé

Pre skvalitnenie vyučovania cudzích jazykov je veľmi dôležitý výskum v tejto oblasti, pričom optimálna výuka cudzieho jazyka pracuje s potrebou používať skutočný jazyk rodených hovoriacich. Táto práca rozoberá relatívne nový nástroj v oblasti lingvistiky, ktorého cieľom je vytvárať a zhromažďovať zoznamy slovnej zásoby používanej rodenými hovoriacimi v rôznych komunikačných situáciách, a jeho použitie pri vývoji učebníc španielčiny.

Teoretická časť práce poskytuje súhrn poznatkov o tomto nástroji a jeho možnostiach vo vyučovaní španielčiny ako cudzieho jazyka, ako aj základné teoretické poznatky o problematike osvojovania slovnej zásoby a tiež jej výbere v učebniciach.

Hlavným predmetom analýzy v tejto diplomovej práci je slovná zásoba v českej učebnici *Nueva Aventura* a jej príslušnosť k jazyku, ktorý používajú rodení hovoriaci. Hlavným cieľom práce je teda zistiť, či je v tejto učebnici zastúpená slovná zásoba používaná rodenými hovoriacimi. V praktickej časti sa porovnáva zoznam slovnej zásoby získanej od rodených hovoriacich s lexikálnym obsahom učebnice, pričom sa berú do úvahy prítomné a neprítomné slová, ako aj ich frekvencia a ich prítomnosť v slovníku učebnice. Okrem toho sa skúma rozloženie týchto slov v celej učebnici. Vychádza sa pritom zo zoznamov slov získaných z výsledkov výskumu, ktorý je súčasťou projektu *Proyecto Panhispánico de la disponibilidad léxica* (PPHDL).

Výsledkom práce je kvantitatívna aj kvalitatívna analýza údajov získaných porovnaním slovnej zásoby učebnice a zoznamu slovíčok zo zvoleného zoznamu.

Výsledky tejto práce podnecujú k niekoľkým možnostiam pre budúci výskum. Jednou z možností je preskúmanie širšieho výberu sémantických polí, ako aj porovnanie výsledkov medzi rôznymi publikáciami z projektu PPHDL či vytvorenie jednotného zoznamu. Okrem toho by výskum mohol zahŕňať otázku, či a ako učitelia cudzích jazykov kompenzujú nedostatky v učebniciach využívaním rôznych zdrojov.

8 Referencias bibliográficas

- Alba, O. (1995). *Léxico disponible de la República Dominicana*. Pontificia Universidad Católica Madre y Maestra.
- Alba, O. (2013). Variación diacrónica del léxico disponible dominicano. *LEA: Lingüística Española Actual*, 35(1), pp. 149-180.
- Alba, O. (2014). *Observación del cambio lingüístico en tiempo real: el nuevo léxico disponible de los dominicanos*.
- Ávila Muñoz, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos. *Journal of Spanish Language Teaching*, 3(1), pp. 31-43.
- Barcroft, J. (2005). La enseñanza del vocabulario en español como segunda lengua. *Hispania*, 88(3), pp. 568-583.
- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En R. M. Castañer Martín & V. Lagüéns Gracia (Coords.), *De moneda nunca usada: Estudios dedicados a José M^a Enguita Utrilla*, pp. 85-107.
- Bartol Hernández, J. A. y Borrego Nieto, J. (2019). El léxico disponible de la ropa en Cádiz, Valladolid y el conjunto de España: norma y variación. *Ogigia. Revista electrónica de estudios hispánicos*, 25, pp. 57-82.
- Benítez Pérez, Pedro. (1994^a). Léxico real/irreal en los manuales de español para extranjeros. En Actas del Segundo Congreso Nacional de ASELE. En Montesa, S. y Garrido, A. (eds.) *Español para extranjeros: didáctica e investigación*, pp. 325-333. Málaga: ASELE.
- Benítez Pérez, P. (1994^b). Qué vocabulario hay que enseñar en las clases de español como lengua extranjera. En L. Miquel y N. Sans (eds.), *Didáctica del español como lengua extranjera II*, pp. 9-12. Madrid: Fundación Actilibre.

- Borrego Nieto, J. (2004). Algunas preguntas en relación con el concepto de 'léxico disponible'. En Parado Aragonés, J. y Galloso Camacho, M. V. (eds.), *Diccionario, léxico y cultura*, pp. 59-69. Universidad de Huelva.
- Borrego Nieto, J. (2008). Edad y culturas léxicas. En M. L. Arnal Purroy (Ed.), *Estudios sobre disponibilidad léxica en los jóvenes aragoneses*, pp. 227-244. Zaragoza: Institución Fernando el Católico.
- Brožová, K., Ferrer Peñaranda, C., & Caba, M. (2011). *Aventura 3: Španělština pro střední a jazykové školy*. Praga: Klett nakladatelství s. r. o.
- Carcedo González, A. (2000). Índices léxicos-estadísticos y graduación del vocabulario en la enseñanza de ELE: (aspectos culturales). *Nuevas perspectivas en la enseñanza del español como lengua extranjera: actas del X Congreso Internacional de ASELE* (Cádiz, 22-25 de septiembre de 1999), pp. 175-184. Servicio de Publicaciones.
- Criado, R., Sánchez, A. (2012). Lexical frequency, textbooks and learning from a cognitive perspective. A corpus-based sample analysis of ELT materials. *RESLA*, 2012, pp. 77-94.
- Cruz Alonso, R. (2016). *El léxico disponible de Castilla y León*. Salamanca: Ediciones Universidad de Salamanca.
- Davies, M. y Hayward Davies, K. (2018). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners* (2ª ed.). Nueva York: Routledge.
- Díaz, I. C. S., Ibáñez, E. T. y Oliva, M. F. R. (2020). Propuesta de una taxonomía de los centros de interés en los estudios de disponibilidad léxica. *DELTA: Documentação De Estudos Em Lingüística Teórica E Aplicada*, 36(4).
- DispoLex. (s. f.). *¿Qué es el Proyecto Panhispánico?* <<http://www.dispoxex.com/info/el-proyecto-panhispanico>> Accedido el 3/2/2024.
- Eurostat. (2024). *Foreign languages in upper secondary education in 2022*. <<https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20240410-1>>. Accedido el 16/4/2024.

- Gallego Gallego, D. J. (2014). Léxico disponible de 82 inmigrantes estudiantes de español en la ciudad de Alcalá de Henares. *Lengua y migración/Language and Migration*, 6(2), pp. 95-123.
- González Fernández, J. (2014). Idoneidad de los centros de interés clásicos en los estudios de disponibilidad léxica aplicados al español como lengua extranjera. *Revista Nebrija de Lingüística Aplicada* (2014) 16.
- Hernández Muñoz, N. (2006). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: El léxico disponible de los estudiantes castellano-manchegos* [Tesis doctoral]. Salamanca: Universidad de Salamanca.
- Hernández Muñoz, N., Izura, C. y Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology*, 18(5), pp. 730–755.
- Hernández Muñoz, N., Tomé Cornejo, C. y López García, M., Bartol Hernández, J. A. (2023). *Manual de LexPro*. <<https://dispogram.usal.es/>> Accedido el 5/2/2024.
- Hidalgo Gallardo, M. (2017). *La disponibilidad léxica como método de detección del vocabulario y de su selección en manuales. Aplicación en una muestra de estudiantes sinohablantes de ELE* [Tesis doctoral]. Universidad de Jaén.
- Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el español* [en línea]. <https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/>. Accedido el 28/2/2024.
- Klett. (s. f.). *Nueva Aventura*. <<https://klett.cz/aventura/nueva-aventura/>>. Accedido el 2/4/2024.
- Larrotta, C. (2011). Second Language Vocabulary Learning and Teaching: Still a Hot Topic. *Journal of Adult Education*, 40(1), pp. 1-11.
- Lightbown, P. M., Spada, N. (2013). *How Languages are Learned*. Oxford University Press.
- López González, A. M. (2014). *Disponibilidad léxica. Teoría, método y análisis*.

- López Morales, H. (1995). Los estudios de disponibilidad léxica: pasado y presente. *Boletín De Filología*, 35(1), pp. 245–259.
- López Morales, H. (2005). Un nuevo corpus para el estudio del español: la disponibilidad léxica. *Oralia*, 8, pp. 141–159.
- López Rivero, E. L. (2008). *Estudio de disponibilidad léxica en 43 estudiantes de ELE* [Tesis de fin de máster]. Universidad Antonio de Nebrija.
- Loureiro, V. J. (2007). ¿Por qué y para qué enseñar vocabulario? En *IV simposio internacional José Carlos Lisboa de didáctica del español como lengua extranjera del Instituto Cervantes*, pp. 457-465. Rio de Janeiro.
- Mackey, William Francis. (1971). *Le vocabulaire disponible du Français*, 2 vols. Paris: Didier.
- Oster, U. (2009). La adquisición de vocabulario en una lengua extranjera: De la teoría a la aplicación didáctica. *Porta Linguarum*, 11, pp. 33-50.
- Paredes García, F. (2014). A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE. *Revista Nebrija de Lingüística Aplicada*, 16.
- 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), pp. 1-32.
- Paredes García, F., Guerra Salas, L. y Gómez Sánchez, M. E. (2022). *Léxico disponible de los jóvenes preuniversitarios de la Comunidad de Madrid*. Editorial Universidad de Alcalá.
- Pérez Jiménez, E. (2016). *El léxico disponible de los estudiantes preuniversitarios en La Rioja* [Tesis doctoral]. La Rioja: Universidad de La Rioja.
- Real Academia Española. 1000 lemas más frecuentes (CREA Anotado) [en línea]. En *Corpus de referencia del español actual (CREA)*. <<http://www.rae.es>>. Accedido el 29/1/2024.

- Real Academia Española. (2014). *Diccionario de la lengua española* [en línea]. <<https://www.rae.es>>. Accedido el 13/2/2024.
- Rubio Sánchez, R. (2017). Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de Español como Lengua Extranjera. En Barrios de la Rosa, F. (coord.), *Palabras Vocabulario Léxico: La Lexicología Aplicada a La Didáctica Y a La Diacronía*, pp. 143–163.
- Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.
- Samper Padilla, J. A. y Hernández Cabrera, C. E. (1997). El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística. En Almeida, M. y Dorta, J. (eds.) *Contribuciones al estudio de la lingüística hispánica: homenaje al profesor Ramón Trujillo* (Vol. 2), pp. 229-239. Santa Cruz de Tenerife: Montesinos.
- Samper Padilla, J. A., Bellón Fernández, J. J. y Samper Hernández, M. (2003). El proyecto de estudio de la disponibilidad léxica en español. En G. Wotjak (coor.), *Pautas y pistas en el análisis del léxico hispano(americano)*, pp. 27–140. Lingüística Iberoamericana.
- Sánchez, R. R. (2017). Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de Español como Lengua Extranjera. En del Barrio de la Rosa, F. (coord.), *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica ya la diacronía*, pp. 143-161. Edizioni Ca'Foscari.
- Serfati, M. (2019). Disponibilidad léxica y la selección léxica en los manuales de ELE en Marruecos. *El Español Por El Mundo*, 2, pp. 311-324.
- Thornbury, S. (2002). *How to teach vocabulary*. Essex: Pearson Education Limited.
- Tomé Cornejo, C. (2019). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE* [Tesis doctoral]. Universidad de Salamanca.
- Zlesáková, K., & Ferrer Peñaranda, C. (2018). *Nueva aventura 1: Španělština pro střední a jazykové školy*. Praga: Klett nakladatelství s. r. o.

Zlesáková, K., Ferrer Peñaranda, C., & Caba, M. (2019). *Nueva aventura 2: Španělština pro střední a jazykové školy*. Praga: Klett nakladatelství s. r. o.

9.3 Lista de palabras con el ID $\geq 0,1$: “los animales”

Los animales	G	Total	NA														NA														A														G	
			1	0	1	2	3	4	5	1	6	7	8	9	10	2	2	0	1	2	3	4	5	1	6	7	8	9	10	2	3	0	1	2	3	4	5	1	6	7	8	9	10	2		
1	perro	49	17	2	1	0	5	0	3	0	3	0	2	0	0	1	16	0	2	0	3	0	4	0	3	1	2	0	1	0	16	0	1	0	0	1	1	0	3	0	1	2	0	0	7	0
2	gato	39	2	0	0	1	0	0	0	0	1	0	0	0	0	25	1	0	0	9	0	1	0	3	5	0	1	1	4	12	0	6	0	1	2	0	0	0	0	0	1	1	1	0		
3	león	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
4	vaca	9	2	0	0	0	0	0	0	0	0	0	0	2	0	4	0	0	0	3	0	0	0	0	0	0	0	1	3	0	0	3	0	0	0	0	0	0	0	0	0	0				
5	caballo	13	2	0	0	0	0	0	0	0	0	0	2	0	9	0	0	0	3	0	0	0	0	3	0	3	0	2	0	1	0	0	0	0	0	0	1	0	0	0	0					
6	tigre	2	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
7	elefante	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
8	ratón	7	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	2	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
9	cerdo	10	5	0	0	0	0	0	0	0	0	0	5	0	4	0	0	0	4	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0						
10	oveja	7	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	5	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
11	gallina	9	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	3	0	0	1	0	0	0	0	0	5	0	0	5	0	0	0	0	0	0	0	0	0						
12	jirafa	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
13	conejo	3	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0						
14	serpiente	2	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
15	mono	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
16	ballena	3	2	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
17	pájaro	7	1	0	0	0	0	0	0	0	1	0	0	0	6	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
18	pez	9	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	3	3	0	0	0	0	0	0	0	3	0	0	0	0	2	0	0	0	0	1	0	0						
19	burro	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
20	oso	8	2	0	0	0	0	0	0	0	0	0	2	0	4	0	0	0	1	2	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	2	0						
21	toro	10	10	2	0	0	0	0	2	0	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
22	lobo	2	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
23	tiburón	3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0						
24	cabra	5	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						

Anotación

Nombre y apellido de la autora: Bc. Monika Husárová

Departamento y facultad: Departamento de Lenguas Románicas, Facultad de Filosofía, Universidad Palacký de Olomouc

El título de la tesis: Disponibilidad léxica y selección léxica en un manual checo de ELE

El supervisor de la tesis: doc. Mgr. Enrique Gutiérrez Rubio, Ph.D.

Número de páginas: 73

Número de caracteres del propio trabajo: 123 451

Número total de caracteres: 143 888

Número de apéndices: 3

Número de recursos utilizados: 52

Palabras claves: disponibilidad léxica, vocabulario, manual, ELE

Anotación: Este trabajo de fin de máster estudia la abundancia de las unidades del léxico disponible de los hablantes nativos españoles en un manual de ELE checo, *Aventura*. La base teórica aporta un sumario de los conocimientos sobre el concepto de la disponibilidad léxica y su potencial en la enseñanza de ELE, así como la teoría básica de la adquisición de vocabulario y su selección en los manuales. El análisis pretende analizar tanto las coincidencias entre ambos sujetos del estudio como las deficiencias léxicas del manual y determinar, así, el grado de representación del léxico disponible en dicho manual.

Annotation

The author's name and surname: Bc. Monika Husárová

Department and faculty: Department of Romance Languages, Faculty of Arts, Palacký University Olomouc

Title of the thesis: Lexical availability and lexical selection in a Czech SFL textbook

Thesis supervisor: doc. Mgr. Enrique Gutiérrez Rubio, Ph.D.

Number of pages: 73

Number of characters of the actual work: 123 451

Total number of characters: 143 888

Number of annexes: 3

Number of sources: 52

Keywords: lexical availability, vocabulary, manual, SFL

Annotation: This master's thesis studies the abundance of lexical units available to Spanish native speakers in a Czech Spanish as foreign language (SFL) textbook, *Aventura*. The theoretical background provides a summary of knowledge about the concept of lexical availability and its potential in SFL teaching, as well as the basic theory of vocabulary acquisition and its selection in textbooks. The analysis strives to analyze both the commonalities between the two subjects of the study and the lexical deficiencies of the textbook and thus determine the extent of representation of the available lexicon in the textbook.