

Selección del vocabulario meta en español: propuesta metodológica basada en la triangulación de fuentes léxicas

Target vocabulary selection: a methodological proposal based on the triangulation of lexical sources

Matías Hidalgo Gallardo

EOI de Madrid-Embajadores
España

Anna Rufat Sánchez*

Universidad de Extremadura
España

ONOMÁZEIN 58 (diciembre de 2022): 121-147

DOI: 10.7764/onomazein.58.07

ISSN: 0718-5758



Matías Hidalgo Gallardo: EOI de Madrid-Embajadores, España. Orcid: 0000-0001-6129-4979.

| E-mail: mhidalgogallardo@educa.madrid.org

Anna Rufat Sánchez: Departamento de Filología Hispánica y Lingüística General, Facultad de Filosofía y Letras, Universidad de Extremadura, España. Orcid: 0000-0002-7016-0814. | E-mail: arufat@unex.es

Fecha de recepción: noviembre de 2019

Fecha de aceptación: abril de 2020

Resumen

Este trabajo presenta una propuesta de selección del vocabulario meta para estudiantes de español como lengua extranjera (ELE) basada en la triangulación o complementación de tres fuentes de información léxica: (1) el léxico disponible en nativos hispanohablantes, (2) el léxico más frecuente del español y (3) las nociones específicas del *Plan curricular del Instituto Cervantes*. Previamente, se ha realizado un análisis de necesidades léxicas de una muestra de estudiantes de ELE, cuyos resultados han sido cotejados con las mencionadas fuentes de vocabulario. Las carencias encontradas han sido evaluadas con el fin de observar cómo dichas fuentes las cubren. Se ha comprobado así que cada fuente satisface necesidades léxicas diferentes (con una similitud del 12,5 % en las necesidades cubiertas y una compatibilidad léxica de las fuentes del 24 %), por lo que la viabilidad de la triangulación queda justificada. La propuesta resultante ofrece una recopilación léxica nivelada que aúna las bondades de varios criterios de selección ya asentados y supera las limitaciones derivadas de su empleo por separado.

Palabras clave: vocabulario meta; selección del vocabulario; disponibilidad léxica; léxico frecuente; español como lengua extranjera (ELE).

Abstract

This paper presents a proposal on how to select suitable vocabulary items to teach Spanish as a foreign language (SFL) based on the triangulation or complementation of three sources of lexical information: (1) the available lexicon in native Spanish speakers, (2) the most frequent vocabulary in Spanish and (3) the specific notions of the *Plan curricular del Instituto Cervantes*. Previously, an analysis of the lexical needs of a sample of SFL students has been carried out, whose results have been collated with the abovementioned sources of vocabulary. The found shortcomings have been evaluated in order to observe how they are covered by each of these three sources. It has been proven that each source

* Agradecimientos de la autora Anna Rufat a la Consejería de Economía e Infraestructuras de la Junta de Extremadura (Proyecto IB180809) y FEDER.

covers different lexical needs (with a similarity of 12.5% in the covered needs and a lexical compatibility of the sources of 24%), so the viability of the triangulation is justified. The resulting proposal offers a lexical compilation in different levels that combines the benefits of several well-established selection criteria and overcomes the limitations arising from using them individually.

Keywords: target vocabulary; vocabulary selection; lexical availability; frequent vocabulary; Spanish as a foreign language (SFL).

1. Introducción

El vocabulario determinado como objeto prioritario de aprendizaje varía de forma llamativa en los distintos cursos y materiales correspondientes al ámbito de la enseñanza del español como lengua extranjera (ELE). Esto se debe a la coexistencia de varias fuentes útiles, pero dispares e incompletas, para llevar a cabo esa labor, como el *Plan curricular del Instituto Cervantes (PCIC, 2006)* o las listas de vocabulario frecuente. A dichas fuentes se han sumado desde hace algún tiempo los estudios de disponibilidad léxica (DL), que se aplican a la enseñanza de lenguas con dos propósitos: (1) contribuir a establecer un léxico de referencia fundamentado en datos procedentes de hablantes nativos y (2) medir el dominio léxico de los aprendientes. El presente trabajo combina ambas perspectivas, pues tiene como objetivo determinar las necesidades léxicas de una muestra italoфона de aprendientes universitarios de español con respecto, en primer lugar, a un vocabulario meta elaborado a partir de datos recabados de estudiantes universitarios nativos. Después, en un afán por superar la disparidad antes señalada, los datos de los aprendientes italofonos se comparan también con las otras dos fuentes mencionadas, el *PCIC* y la lista de frecuencia que constituye el diccionario de Davies y Davies (2018), de modo que se obtienen tres conjuntos de necesidades léxicas más o menos diversos. Como un primer paso hacia un vocabulario meta homogéneo, se triangulan, finalmente, los tres listados de referencia —el léxico disponible de los nativos, el *PCIC* y el diccionario de Davies y Davies— para establecer sus coincidencias y discrepancias.

Dado el carácter inicial de esta investigación, cuyas bases habrán de servir para abordar nuevos campos temáticos con idéntica finalidad, el estudio realizado se limita al centro de interés “Partes del cuerpo”, que suele copar gran atención en los estudios de DL y es, asimismo, el primer campo nocional que se muestra en el *PCIC*. Se trata, además, de un campo léxico sin demasiadas dificultades de delimitación, lo cual facilita el manejo de los datos y lo hace especialmente adecuado para una investigación fundacional como esta. Se trata, en definitiva, de establecer un procedimiento riguroso y abarcador para determinar un vocabulario de referencia en la enseñanza del español que pueda extenderse a nuevos dominios léxicos.

2. Vocabulario meta. Criterios de selección del vocabulario

El conocimiento léxico es esencial para el desarrollo de la competencia lingüística de los aprendientes de lenguas extranjeras en todas las destrezas (Chao y otros, 2000; Alderson, 2005). Diversos estudios defienden que la comprensión de un texto se logra satisfactoriamente solo cuando el aprendiente conoce entre el 95 y el 98 % de las palabras que lo conforman (Nation, 2006; Laufer y Ravenhorst-Kalovski, 2010). Según eso, la enseñanza del vocabulario debiera considerarse un objetivo primordial en el aula, para cuya consecución

es necesario un ejercicio de planificación docente que implica, antes que nada, determinar cuál será el vocabulario que prioritariamente debe aprenderse. Para ello conviene tener en cuenta el conocimiento léxico previo de los aprendientes. En el ámbito de la enseñanza del inglés, existen test de diagnóstico estandarizados que informan de las necesidades precisas de los estudiantes (Schmitt, Schmitt y Clapham, 2001), y que además pueden complementarse con el Productive Vocabulary Levels Test (Laufer y Nation, 1999), que proporciona indicaciones sobre el vocabulario productivo del estudiante¹. En el caso del español, sin embargo, no disponemos de test de este tipo, con lo que la selección de vocabulario meta debe realizarse forzosamente *a priori*. Sí que es posible, de todos modos, atender a los intereses particulares de los estudiantes y trabajar con vocabulario propio de los ámbitos en los que vayan a desenvolverse, aunque, como se verá después, esta parcela constituye solo una parte de los contenidos léxicos que deben considerarse en un curso.

Por otro lado, las pruebas de diagnóstico léxico solo son posibles si se parte de un vocabulario que se considere de referencia, para comprobar cuánto se acerca o se aleja el aprendiente de él. En tal sentido, desde la primera mitad del siglo XX muchos lingüistas han tratado de establecer ese vocabulario central o nuclear de las lenguas, fundamentalmente del inglés (Ogden, 1934; West, 1953)². Y es que, como señala Gardner (2013: 35), si está demostrado que un puñado de palabras son las que hacen la mayor parte del trabajo, tiene sentido conceder a ese grupo de palabras un papel preponderante en el proceso de enseñanza-aprendizaje. Hablamos, en definitiva, del vocabulario más frecuente, en oposición al léxico especializado o específico de un ámbito discursivo concreto. Así, aunque, como se ha dicho, no disponemos en español de test de diagnóstico léxico —lo cual explica la inexistencia de datos sobre el número de palabras de cada una de las bandas de frecuencia que conocen los estudiantes de español en diferentes niveles de dominio—, sí existen listas de vocabulario frecuente elaboradas a partir de documentos de lengua natural que pueden contribuir a diseñar dichos test. La más destacada corresponde sin duda al diccionario de Davies y Davies (2018), que proporciona las 5000 palabras más frecuentes del español a partir del Corpus del Español, formado por más de dos mil millones de palabras de textos recientes (en su mayoría posteriores a 1990) tanto de España como de Latinoamérica de diferente tipo (páginas web, ficción, no ficción, conversaciones, etc.). De acuerdo con Davies (2005), conocer las 1000 palabras más frecuentes permite a los aprendientes reconocer

-
- 1 De hecho, se está trabajando en la aplicación de pruebas de disponibilidad léxica o tareas de categorización para la estimación del tamaño del léxico productivo (Roghani y Milton, 2017), donde se ha apreciado que los resultados son comparables a los de otras pruebas como, precisamente, el Productive Vocabulary Levels Test y que corresponden a un tercio del léxico receptivo medido por el X_Lex (Meara y Milton, 2003).
 - 2 No en vano, para el inglés ya existen cientos de listados léxicos de referencia. A modo de ejemplo, Burkett (2017) recoge en su tesis hasta 104 listas.

entre el 75-80 % de todos los lexemas en el español escrito, y hasta el 88 % en el oral; con las 4000 palabras más frecuentes el aprendiente puede reconocer más del 90 % de las palabras de una conversación, pero necesitaría alrededor de 6000 para reconocer ese mismo porcentaje en un texto escrito. Los primeros puestos los ocupan palabras gramaticales básicas como *el* o *de*, desde las que se progresa hacia palabras léxicas intermedias, y se tiene también en cuenta la dispersión o alcance de las palabras, es decir, el hecho de que estas se extiendan de manera equilibrada a lo largo de todo el corpus y no deban una presencia relevante a un solo tipo de texto o de registro.

Lo anteriormente dicho podría hacer pensar que la frecuencia es un criterio muy sólido para determinar el vocabulario meta de un grupo de aprendientes de español. Sin embargo, las listas de vocabulario frecuente presentan dos grandes inconvenientes. En primer lugar, no dan cabida a los bloques léxicos, con lo que unidades léxicas tan frecuentes como, por ejemplo, *sistema nervioso* no pueden ser rastreadas. Y, por otro lado, obvian el problema de la homonimia, por lo que una palabra situada en una posición elevada de la lista podría deber su preeminencia a la suma de varios significados; no se contempla, por lo tanto, la posibilidad de que una palabra frecuente tenga sentidos infrecuentes. Así, lo mismo que ya señaló Lorge (en West, 1953: XI), a propósito de su publicación *General Service List*, podemos decir hoy:

The selection of words for the teaching of English to foreigners, to a large degree, has been based on the frequency of the occurrence of words in printed English. Such facts about word frequency have been useful in the choice of basic vocabularies of specified sizes. The teaching of these words, however, left much to the judgment of the teacher, for he had to choose which meaning or meanings of these words to impart. In general, the more frequent the occurrence of a word, the greater is the variety of meanings in which it is used.

Además, como sugeríamos más arriba, las listas de frecuencia contienen una gran abundancia de vocablos sin significado propiamente léxico, sino simplemente gramatical, y no recogen otros de contenido léxico que, sin ser frecuentes, son muy comunes en determinados ámbitos temáticos. Son las llamadas *palabras temáticas* —propias de un tema determinado—, que designan seres y objetos (sustantivos, en su mayoría) y que suelen aparecer en agrupaciones, de modo que el empleo de alguna favorece el de otras, y que se actualizan repetidamente cuando se aborda el tema que las engloba.

Con el objetivo de paliar algunas de estas carencias, han cobrado fuerza en los últimos tiempos los estudios de DL aplicados a la enseñanza del español, cuyas virtudes se han señalado en oposición a las debilidades de las listas de frecuencia: “Pero esta identificación frecuencia = útil se ha demostrado inexacta; pues aunque admitiéramos que una palabra frecuente es útil, no podríamos aceptar que las palabras poco frecuentes no sean útiles” (Bartol Hernández, 2010: 88). Se trata, por tanto, de determinar palabras que, no siendo necesariamente frecuentes, sí se encuentran “disponibles” para los hablantes nativos en

su comunicación cotidiana. Ello se hace a partir de una metodología fundamentada en “técnicas de la psicología asociativa para acceder a la competencia léxica de los hablantes” (Paredes García, 2015: 4) que aspira, por tanto, a completar los vacíos que las listas de frecuencia puedan presentar (Ávila Muñoz, 2016). No obstante, debe reconocerse que esta metodología presenta algún que otro inconveniente. Para empezar, las listas de frecuencia se extraen de corpus confeccionados con muestras de lengua natural, es decir, procedentes de situaciones reales de comunicación, lo cual no es posible en los estudios de DL. Esta circunstancia implica discutir las afirmaciones que equiparan el léxico obtenido en estos estudios con el que se utiliza en la comunicación habitual nativa. A ello hay que añadir el componente subjetivo presente en la recogida de datos, que desemboca en la variabilidad de los resultados obtenidos al margen de las diferencias geográficas, que, obviamente, también los condicionan (Ávila Muñoz, 2016).

Por otro lado, en la aplicación de los resultados de estos estudios a la determinación del vocabulario meta del español tampoco se ha resuelto definitivamente el asunto de los centros de interés, pues los que se utilizan habitualmente no parecen los más adecuados para fijar un vocabulario de referencia orientado a aprendientes. Al respecto, varios especialistas (Ávila Muñoz, 2016: 8; Santos Díaz, 2017: 137; Tomé Cornejo, 2015: 440) han señalado la conveniencia de acercar en lo posible los centros de interés a los campos nocionales propuestos por el *PCIC*, que, al fin y al cabo, es el documento de referencia en la programación y creación de materiales en la enseñanza del español. Como explicamos más abajo, esta es la vía que sigue nuestra investigación.

Aparte de las listas de frecuencia y de los datos que puedan sugerir los estudios de DL, lo cierto es que los criterios más habituales para la selección del vocabulario meta en un curso de español suelen ser el propio manual de aula o el repertorio de “Nociones específicas” del *PCIC*. Con respecto al manual, suele proporcionar vocabulario relacionado con el tema de cada unidad (comida, transporte, enfermedad, ropa, etc.), lo que conlleva la sobreaparición de palabras que se refieren a conceptos concretos y que al mismo tiempo estén infrarrepresentadas las palabras cuyos referentes son conceptos abstractos (Davies y Face, 2006). Asimismo, el manual normalmente no incorpora ninguna indicación sobre el uso habitual de ese vocabulario en la conversación cotidiana o en textos reales. Como señalan Davies y Davies (2018: 1), algunas de las palabras que incluyen los manuales son tan infrecuentes en los textos cotidianos que es posible que los estudiantes no se las encuentren fuera del aula. A esto se une el hecho de que, precisamente por la disparidad de criterios que existen para su selección, en cada manual —y en cada curso— se trabaje con un vocabulario diferente para los mismos niveles, según los temas que se aborden. En cuanto al inventario del *PCIC*, se trata únicamente de una guía orientativa que no pretende determinar aspectos como el número de unidades léxicas por nivel o el orden de aprendizaje de esas unidades, sino tan solo “dar una idea del tipo de unidades léxicas que el hablante debe conocer en cada uno de los niveles” (cf. “Nociones específicas. Introduc-

ción”). Esto contribuye a la mencionada variación presente en los manuales, pues, aunque algunos partan del inventario del *PCIC*, al incluir este listas breves y abiertas, mucho del vocabulario finalmente seleccionado será forzosamente distinto.

Ante esta situación, tal vez lo más conveniente sea combinar varios criterios (Hidalgo-Gallardo, 2017a, 2017b y 2019). Desde la perspectiva del profesor, probablemente lo más aconsejable es partir del vocabulario que presente el manual utilizado y trabajarlo en profundidad en caso de que el manual no lo haga o no lo haga suficientemente. Conviene, asimismo, dejar espacio para el vocabulario personal, aquel “que necesitará un usuario de la lengua para representar su propia identidad social y psicológica y para desenvolverse en los ámbitos en los que se verá involucrado” (Troitiño, 2017: 147). Aun así, como adelantábamos más arriba, esta vía será por sí sola insuficiente, pues resulta imprescindible contar con un vocabulario de referencia a partir del cual poder programar un curso. Y ahí toma importancia el vocabulario central o nuclear de una lengua, para cuya determinación, según hemos ido argumentando, parece necesario encontrar maneras para conjugar los datos proporcionados por las listas de frecuencia con los hallazgos que aporten los estudios de DL. A esto último pretende contribuir esta investigación, cuyos detalles pasamos a exponer.

3. Método

3.1. Muestra

Para este estudio preliminar nos hemos basado en una muestra mixta conformada por 239 universitarios nativos españoles y 38 universitarios italianos con un nivel de competencia equivalente al B2 según los parámetros del *Marco común europeo de referencia para las lenguas* (MCER; Consejo de Europa, 2002). La demarcación universitaria se debe a que nuestro foco didáctico es la enseñanza de ELE en contextos universitarios, pues constituyen un ámbito consolidado en el que los repertorios léxicos que se produzcan hallarán una aplicación inmediata, pero además, tal y como afirma Tomé Cornejo (2015: 359), la especialización académica de este tipo de informantes parece no afectar a sus respuestas y, a su vez, parecen mostrarse más maduros a la hora de responder. Asimismo, en aras de determinar un vocabulario meta más representativo, siguiendo las recomendaciones de Ávila Muñoz (2016: 9) —y en la línea de proyectos en marcha en el seno del Instituto Cervantes—, hemos procurado que la muestra sea geográficamente dispersa, ya que los “materiales generales de español como L2 deben seleccionar el léxico con criterios geolectales variados”. No obstante, al tratarse del primer acercamiento, la dispersión geográfica se limita, por el momento, a España, aunque el propósito final es el de incorporar también las demás variedades del español con objeto de confeccionar un vocabulario meta panhispánico.

Por su parte, en relación con los informantes italo parlantes, que responden a un perfil similar a los nativos, hemos privilegiado un nivel intermedio avanzado de español para

potenciar una actualización léxica que vaya más allá de los ítems más básicos y prototípicos del centro de interés en cuestión, lo cual nos permitirá localizar de forma más eficaz y realista las posibles necesidades léxicas que presenten.

3.2. Centros de interés

Una de las grandes novedades metodológicas que distinguen este trabajo de otros que se basan en la metodología léxico-estadística de la disponibilidad léxica es la nómina de estímulos nocionales que se han empleado. En nuestro caso, no nos encorsetamos en los dieciséis centros de interés (CI) clásicos del Proyecto Panhispánico de Disponibilidad Léxica (PPHDL)³, sino que, adaptándolos a nuestro objetivo final de corte didáctico, partimos del inventario de nociones específicas del *PCIC*. Dado el estadio inicial de este proyecto piloto, por el momento solo abordamos once de estas nociones, y en este trabajo en concreto presentamos los datos para una de ellas: *1.1. Partes del cuerpo*. Nuestra intención futura, como ya indicamos en un trabajo precedente (Jiménez Calderón y Rufat, 2019), es la de analizar todas las categorías nocionales presentes en este documento, realizando modificaciones o aclaraciones mínimas para facilitar la comprensión por parte de los informantes y distribuyéndolas de forma asequible en diversas pruebas que no superen los 20 minutos de duración. Una de las bondades de esta adaptación didáctica de los centros de interés es la correspondencia total de los mismos a la hora de realizar cotejos entre los resultados de las encuestas de léxico disponible, tanto en nativos como en estudiantes, y los contenidos léxicos de los documentos que usamos como referencia. Se trata, por tanto, de otra innovación que nos distingue de trabajos previos que, además de no perseguir —o al menos no de forma exclusiva— la determinación de un vocabulario meta, difieren entre los CI empleados con nativos y los usados con extranjeros y entre los que sirven como estímulo en las encuestas y las categorías recogidas por el *PCIC*, con las consiguientes divergencias en los contrastes.

3.3. Herramienta

Para llevar a cabo la investigación hemos utilizado una aplicación informática en línea, LexicalEx, que permite responder a un test de DL a través de cualquier dispositivo con acceso a Internet. Su empleo, sin embargo, va más allá del mero carácter instrumental, pues implica determinadas transformaciones en relación con los estudios habituales de este tipo. Para empezar, como ya indicábamos, esta herramienta nos ha permitido adaptar los CI a las categorías nocionales proporcionadas por el *PCIC*, lo cual facilita la comparación de los datos con este documento de referencia y también sus potenciales aplicaciones didác-

3 Puede consultarse más información sobre el PPHDL, incluidos los CI tradicionales, en www.disporex.com.

ticas. Por otro lado, al estar alojada en una página web, permite la recogida de un número ilimitado de muestras; los resultados se actualizan de manera automática tras cada nueva respuesta; y estos pueden recuperarse de acuerdo a distintos criterios como el índice de disponibilidad o la frecuencia de aparición, lo cual favorece varias posibilidades de aproximación y facilita el análisis.

Dado que todo el procedimiento es automático, el empleo de LexicalEx evita al investigador tener que pasar por la transcripción o digitalización de las respuestas, lo que reduce significativamente la carga de trabajo y acelera el proceso. Además, es posible trabajar con informantes de perfiles muy diversos, pues LexicalEx es capaz de filtrar los datos atendiendo a cualquiera de las variables que se introduzcan al recuperar los resultados, variables, a diferencia de en otros programas de DL, ilimitadas. Asimismo, la herramienta posibilita también la recogida de bloques léxicos, por lo que tanto en las instrucciones iniciales que aparecen en el test como en las instrucciones parciales que se muestran para cada centro de interés se especifica y enfatiza que pueden aportarse tanto palabras como expresiones. Por último, la aplicación informática incorpora una herramienta de edición de datos llamada “Normalizar vocablos”, que nos permite unificar distintas formas para un mismo ítem, lo cual asegura el rigor de los resultados obtenidos; así, si los datos incluyen, por ejemplo, las formas *ojo* y *ojos*, o *pie* y *pié*, la herramienta mencionada nos permite agrupar las formas en un mismo lema con total fiabilidad. El usuario simplemente ha de entrar en la página web, clicar en “Realizar test”, introducir una serie de datos que permitirán a la aplicación filtrar posteriormente los resultados, leer las instrucciones e iniciar el procedimiento: los centros de interés se van sucediendo cada dos minutos, y en ese tiempo el informante debe escribir cuantas unidades léxicas le sugiera el centro en cuestión, con la posibilidad de eliminar formas si repara en algún error ortográfico, por ejemplo. Puesto que este primer test consta de 11 centros, resulta una prueba de 22 minutos, una duración razonable. Para fases posteriores, como ya indicamos más arriba, se agruparán nuevos centros en test de similar duración hasta que se obtengan datos para todos los campos nocionales del *PCIC*.

3.4. Procedimiento

Nuestro proceso de trabajo se ha desarrollado fundamentalmente en tres fases: (1) la recopilación de la información léxica referida al estímulo “partes del cuerpo” en las distintas fuentes que nos servirán de referencia, (2) el cotejo léxico de los listados obtenidos y (3) la triangulación de las fuentes de referencia: léxico disponible, léxico frecuente y el contenido léxico del *PCIC*.

En la primera de las fases, la de recopilación, en primera instancia hemos sometido a los informantes nativos a pruebas de disponibilidad léxica para obtener su conocimiento léxico con respecto al estímulo en cuestión (LD-N). Seguidamente, ha sido la muestra de

estudiantes italianos la que ha cumplimentado las encuestas de léxico disponible (LD-ELE). En ambos casos nos hemos servido de la herramienta web LexicalEx. En cuanto al léxico frecuente (LF), hemos analizado el diccionario de frecuencias de Davies y Davies (2018), para lo cual nos hemos basado en el cuadro temático sobre el cuerpo humano que la propia obra incorpora y lo hemos ampliado con otras unidades referidas a este centro de interés contenidas también en el diccionario. Por su parte, para el caso del *PCIC*, el compendio léxico ha consistido en agrupar los ítems que ya este documento ofrece bajo la noción “partes del cuerpo” (L-PCIC) en los distintos niveles de competencia, aunque hemos agregado algunos ítems presentes en otros campos pero íntimamente ligados al centro de interés aquí analizado. Debemos indicar que hemos aplicado medidas de corrección en función de las características de la submuestra italófona (tamaño muestral y nivel) que han supuesto una reducción de los ítems léxicos en la fase de cotejo y que explicaremos con más detalle en el epígrafe dedicado al análisis de los datos.

Para la fase de cotejo léxico, dado que el principal objetivo es el de determinar las posibles necesidades léxicas de los informantes extranjeros, hemos llevado a cabo contrastes por pares tomando siempre como referencia el léxico disponible de los universitarios italianos. Las diferencias encontradas con respecto a la submuestra nativa, al léxico frecuente y al *PCIC*, respectivamente, supondrán una serie de necesidades léxicas de los discentes foráneos. Ante esta realidad, habrá que analizar la naturaleza de las mismas y observar cómo estas fuentes las cubrirían, es decir, comprobar si cada una de ellas cubriría necesidades similares, por lo que sería indiferente a qué fuente acudir para la selección de vocabulario meta, o si, por el contrario, cada una haría un aporte diferente, lo que justificaría su complementación para determinar el vocabulario objeto de aprendizaje por parte de los estudiantes de español. A tal propósito, se ha añadido el cálculo del grado de correspondencia de dichas necesidades y, también, la compatibilidad del componente léxico aportado por las tres fuentes de contraste, hechos que nos permitirán trazar un perfil más detallado de las diferencias existentes entre ellas. Estos cálculos se han hecho a través de la *cardinalidad del conjunto* (López Chávez, 1992): a partir de los vocablos de cada submuestra (*cardinalidad*), hemos considerado la suma de (1) las unidades convergentes en todos los subconjuntos (*intersección*), (2) de las léxicas que no están presentes en todos ellos —aunque sí en varios— (*unión*) y (3) de las exclusivas de cada subconjunto (*complemento*) —la suma de la unión y de los complementos se denomina *suma disyuntiva*—. La compatibilidad será, pues, el peso de la intersección con respecto al total de vocablos (*intersección + suma disyuntiva*), es decir, el porcentaje de unidades léxicas compartidas.

En última instancia, con el soporte de los datos obtenidos en la fase de cotejo, hemos procedido a una triangulación de los listados que barajamos que nos permite determinar una propuesta de vocabulario meta basada en la complementación de fuentes. El proceso de triangulación lo explicaremos, de nuevo, con más detalle en el siguiente apartado.

4. Análisis de los datos

El rendimiento léxico de los informantes nativos tras las encuestas de disponibilidad ha producido un total de 207 vocablos o unidades léxicas diferentes, cantidad que se reduce hasta los 64 ítems en el caso de los universitarios italianos. En el diccionario de frecuencias de Davies y Davies (2018) se recogen 78 unidades entre cuyos significados alguno se refiere a las partes del cuerpo, mientras que los inventarios del *PCIC* aportan 82 ítems. No obstante, como ya anticipábamos, estas cantidades se han corregido y reducido al tener en cuenta las diferencias de tamaño de las submuestras y el nivel de los informantes extranjeros. De este modo, en pos de realizar cotejos equilibrados y atenuar la diferencia de tamaño muestral, la nómina de 207 unidades de léxico disponible nativo se ha igualado, siguiendo un criterio arbitrario pero relevante por sí mismo, con la nómina resultante de los italianos, es decir, hemos considerado solo las 64 unidades más disponibles. Igualmente, pero atendiendo a criterios de nivel, solo hemos conservado 54 ítems de los 78 más frecuentes para las partes del cuerpo. Esta cantidad responde a que hemos contemplado exclusivamente aquellas unidades que se encuentran dentro de las 4000 palabras más frecuentes, franja que, según diversos estudios (West, 2012; Meara y Milton, 2003), se aproximaría al conocimiento de vocabulario necesario en un nivel B2. Algo similar hemos aplicado al listado de 82 ítems contenidos en el *PCIC*, ya que hemos mantenido solo aquellos catalogados como B2 por el propio documento, y la cifra definitiva para los cotejos es de 60 unidades.

Una vez que disponemos de los datos corregidos, hemos llevado a cabo los distintos cotejos. En el contraste entre el léxico más disponible de los italianos y el de los nativos, observamos 19 unidades ausentes en el listado de los primeros: *abdomen, ano, antebrazo, axila, cabello, columna vertebral, extremidad, gemelo, glúteo, jugos gástricos, moflete, músculo, nuca, pene, pulmón, sistema nervioso, talón, tronco y vagina*. Esta cifra supone que casi un 30 % del LD-N (64 ítems) no está entre el más disponible de los italo parlantes. Sin embargo, si ahondamos más en la correspondencia o similitud cualitativa de ambos listados, tenemos que la compatibilidad (véase *cardinalidad del conjunto* en § 3.4.) entre ambos subconjuntos es del 54 %, lo que acentúa las discrepancias entre los listados.

En la comparación entre el LD-ELE y el léxico frecuente, las necesidades léxicas que apreciamos alcanzan los 15 ítems, es decir, cerca de un 28 % del LF no está presente en el léxico más disponible de los estudiantes de ELE: *barba, cabello, columna (vertebral), músculo, nervio, palma, pata, pechos, pulmón, puño, sangre, seno, tronco, vena e índice*. Los cálculos de la compatibilidad, en cambio, arrojan que el grado de correspondencia es del 49 %.

El cotejo del LD-ELE con respecto al contenido léxico del *PCIC* revela un déficit de otras 19 unidades (casi un 32 %) por parte de los primeros: *arteria, articulación, cabello, cana, columna vertebral, costilla, cráneo, esqueleto, extremidad, músculo, nervio, pulmón, páncreas,*

sistema digestivo, sistema nervioso, tendón, tibia, tronco y vello. La compatibilidad, por su parte, es también de un 49 %.

A simple vista, ya podemos apreciar que las necesidades léxicas que se desprenden de cada cotejo no son las mismas, sino que cada fuente cubre necesidades diferentes. Es más, si calculamos la compatibilidad entre estas necesidades, es decir, en qué casos nos daría igual acudir a una fuente u otra para seleccionar el mismo ítem léxico, tenemos que esto solo se produce para cinco unidades: *cabello, columna vertebral, músculo, pulmón y tronco*, es decir, solo en un 12,5 % del total de unidades aportadas por las tres fuentes. Si a esto añadimos que el grado de similitud de los compendios léxicos de las tres fuentes presenta solo un 37 % de compatibilidad⁴ —compatibilidad que se reduce hasta el 24 % si consideramos los listados sin las correcciones previamente realizadas en función de las características de la muestra, con una intersección de 52 ítems y una mayor diferencia por parte del LD-N— y, además, tenemos en cuenta el distinto tratamiento que recibe el léxico en cuestiones como, por ejemplo, la homonimia, la polisemia o los bloques léxicos —con claras limitaciones o deficiencias para el léxico frecuente—, podemos afirmar que cada una de ellas cubre necesidades léxicas distintas, lo que justifica la viabilidad de la triangulación perseguida en este trabajo.

De este modo, la propuesta de triangulación que hacemos en este estudio —proceso para el que hemos tenido en cuenta todas las unidades actualizadas y localizadas en las tres fuentes que usamos como referencia, sin reducciones— consiste en la división de cada listado (léxico disponible de nativos, léxico frecuente y PCIC) en seis grupos y en la asignación a cada uno de ellos, de mayor a menor disponibilidad y frecuencia y de nivel inicial (A1) a nivel superior (C2), de una puntuación de 6 a 1. Con esto, el listado de léxico disponible, a falta de establecer un criterio definitivo más objetivo, se ha agrupado en seis intervalos en función del índice de disponibilidad (ID) tomando como referencia el máximo y el mínimo obtenidos en las encuestas. De este modo, en orden descendente de disponibilidad y, por ende, de puntuación, tenemos: grupo 1, ID = 0,5966-0,7147; grupo 2, ID = 0,4765-0,5955; grupo 3, ID = 0,3574-0,4764; grupo 4, ID = 0,2383-0,3573; grupo 5, ID = 0,1192-0,2382; y grupo 6, ID = 0,0027 a 0,1191. El listado de léxico frecuente está compuesto por seis grupos seleccionados según franjas de 1000, es decir, el primer grupo está formado por las primeras 1000 palabras más frecuentes; el segundo grupo va de la 1001 a la 2000; el grupo 3, de la 2001 a la 3000; el cuarto grupo comprende el intervalo 3001-4000 de las palabras más frecuentes; el quinto, el intervalo 4001-5000; y el grupo 6 se extiende a partir de la palabra 5001. Por

4 Solo 34 ítems están presentes en los tres listados, por orden alfabético: *brazo, cabello, cabeza, cara, ceja, cerebro, cintura, columna, corazón, cuello, dedo, diente, espalda, estómago, frente, garganta, hombro, hueso, mano, mejilla, muñeca, músculo, muslo, nariz, oído, ojo, pecho, pelo, pie, pierna, pulmón, rodilla, tronco y uña*.

último, el léxico del *PCIC* se agrupa en seis bandas en función del propio nivel asignado por el documento, de A1 a C2. La suma de los distintos puntos asignados a cada listado nos indicará la relevancia de cada ítem léxico. Así, los mayores puntajes corresponderán a las unidades léxicas de mayor relevancia en nuestra propuesta de vocabulario meta y serán objeto de estudio, por consiguiente, en los niveles iniciales, mientras que los más bajos se asignarán a las unidades reservadas a niveles superiores.

En la tabla 1 podemos observar una representación de la baremación realizada para el total de unidades presentes en los tres listados y ordenada por la puntuación resultante (véase el anexo para consultar todas las unidades). El baremo total consta de 222 ítems léxicos⁵ con indicaciones sobre el ID en nuestro estudio, y su posición entre paréntesis; el rango de frecuencia general (RF) en Davies y Davies (2018), y entre paréntesis el rango dentro del campo de “partes del cuerpo”; el nivel indicado en el *PCIC* y el campo al que pertenece en este documento o, en su defecto, una serie de observaciones pertinentes que nos pueden orientar sobre la consideración semántica de los mismos; las puntuaciones individuales asignadas; y el puntaje total. Para facilitar la lectura de la tabla, lo ilustramos a continuación con el ejemplo *labio*: según nuestro estudio de disponibilidad léxica en nativos, con un ID = 0,2860 (posición 26.^ª), a esta unidad le corresponden 2 puntos; en cambio, en el listado de frecuencia aparece en la posición 1060 —la 21.^ª dentro del vocabulario relativo a las partes del cuerpo—, lo que le otorga 5 puntos; y en el *PCIC* dicho ítem se encuentra en el inventario del nivel B1 (en 1.1. Partes del cuerpo), de manera que le pertenecen 4 puntos. La suma de estas puntuaciones arroja un total de 11 puntos.

TABLA 1

Ejemplo de la triangulación del LD, LF y *PCIC*

	ÍTEM LÉXICO	ID	RF	PCIC	NOCIÓN PCIC	PUNTAJE			
1	ojo	0.6250 (6)	169 (2)	A1	1.1. Partes del cuerpo	6	6	6	18
2	cabeza	0.7147 (1)	265 (5)	A2	1.1. Partes del cuerpo	6	6	5	17
7	dedo	0.5951 (7)	716 (15)	A2	1.1. Partes del cuerpo	5	6	5	16
9	pelo	0.2995 (17)	873 (18)	A1	1.1. Partes del cuerpo	3	6	6	15
10	espalda	0.3196 (15)	942 (19)	A2	1.1. Partes del cuerpo	3	6	5	14

5 Esta cifra es resultado de sumar los 207 extraídos en las encuestas de LD a nativos, los que se encuentran exclusivamente en el LF y en el *PCIC* —1 y 12 unidades respectivamente— y los presentes en estos dos últimos documentos, es decir, 2 ítems.

12	cara	0.2178 (21)	337 (6)	A2	1.1. Partes del cuerpo	2	6	5	13
16	pecho	0.2009 (24)	971 (20)	B1	1.1. Partes del cuerpo	2	6	4	12
20	labio	0.1703 (26)	1060 (21)	B1	Aparece como parte de una unidad pluriverbal (<i>barra de labios</i>) en 13.7. Estética	2	5	4	11
29	cerebro	0.1486 (30)	1593 (27)	B2	1.1. Partes del cuerpo	2	5	3	10
35	músculo	0.1017 (40)	2715 (45)	B1	1.1. Partes del cuerpo	1	4	4	9
39	ceja	0.2109 (22)	3700 (56)	B2	1.1. Partes del cuerpo	2	3	3	8
48	pestaña	0.2090 (23)	4703 (68)	B2	1.1. Partes del cuerpo	2	2	3	7
55	barriga	0.1874 (30)	—	B1	1.1. Partes del cuerpo	2	0	4	6
61	cadera	0.147 (107)	3595 (55)	C2	1.2. Características físicas	1	3	1	5
69	culo	0.1306 (31)	4527 (65)	—	—	2	2	0	4
81	ombligo	0.1079 (35)	—	C1	1.1. Partes del cuerpo	1	0	2	3
99	antebrazo	0.1549 (29)	—	—	—	2	0	0	2
114	nuca	0.1130 (34)	—	—	—	1	0	0	1

5. Discusión y conclusiones

Pese a la dificultad que puede entrañar la selección del vocabulario y la disparidad de criterios existente para emprender esta tarea —cada uno con sus pros y sus contras—, consideramos, desde nuestra concepción del léxico como elemento nuclear en el aprendizaje/enseñanza de lenguas, que la demarcación de un vocabulario meta es la base para la programación y el desarrollo de la competencia lingüística. Con nuestra propuesta de triangulación (anexo) basada en los resultados de este estudio preliminar, la cual ha quedado justificada por las diferencias reveladas tras el cotejo entre el conocimiento léxico de los informantes italo-fonos y el contenido léxico de las distintas fuentes consultadas, contribuimos con una metodología viable que aúna con equilibrio los criterios de mayor difusión en la delimitación del caudal del vocabulario objeto de enseñanza. Con esta aportación damos respuesta a problemáticas relativas al tratamiento de la homonimia y la polisemia y a la inclusión de bloques léxicos, elementos descuidados especialmente en los listados de léxico frecuente, pero sí incorporados en nuestra propuesta. Del mismo modo, dotamos a nuestro listado de una jerarquía de relevancia del vocabulario —basada en el puntaje obtenido tras la triangulación— que permitirá llevar a cabo eventuales nivelaciones del léxico, aunque debemos mencionar que en ningún caso esta ordenación de las lexías responde a principios prescriptivos; es más, cada docente podrá y deberá ajustar esa nivelación a las necesidades y contextos en los que desempeñe su labor educativa.

No obstante, somos conscientes de la arbitrariedad en algunas de las decisiones tomadas y de la necesidad de buscar alternativas más objetivas con respecto, por ejemplo, a la distribución de los grupos para la asignación de las puntuaciones o a las medidas de corte para la nivelación final del vocabulario de acuerdo al *MCER*. A este propósito, una de las posibles soluciones es la aplicación de un índice de compatibilidad basado en la teoría de los *conjuntos difusos* —algo que ya proponen Ávila Muñoz (2016 y 2017) y Santos Díaz (2017)—, índice basado en pruebas que también arrojan resultados agrupados en seis niveles. En esta línea de investigación y mejora futuras, también querríamos hacer mención de la necesidad de analizar con profundidad y detenimiento la consideración (por inclusión o exclusión) que las fuentes de referencia aquí usadas hacen de la homonimia, de la polisemia y de los bloques léxicos de cara a una baremación del vocabulario más cuidada en el proceso de triangulación. En el último de los casos, el de los bloques léxicos, y en consonancia con esta baremación más precisa, habría que añadir un estudio pormenorizado del *PCIC* en relación con la ausencia o existencia (a veces parcial) de unidades equivalentes. Por ejemplo, en el documento del Instituto Cervantes encontramos *aparato digestivo* en el nivel C1, por lo que le correspondería una puntuación de dos puntos; pero no hallamos *sistema digestivo*, lo que confiere a esta unidad cero puntos. Ante esta realidad, y fundamentado en la equivalencia de los términos, deberíamos, cuando menos, replantearnos la adjudicación de las puntuaciones. Asimismo, sería pertinente también abordar qué tratamiento hay que hacer de las unidades con una fuerte carga terminológica (p. ej., *esternocleidomastoideo*, *uretra*, *cóclea*) de cara a calibrar su puntuación.

De todo esto se desprende la pertinencia de acudir a criterios objetivos corregidos de selección léxica (Grève y Van Passel, 1971), es decir, a partir de una delimitación inicial objetiva —en nuestro caso basada en la triangulación—, someter los resultados iniciales a una fase de revisión en función de criterios subjetivos y didácticos.

Por último, no podemos olvidar que la lengua española es el punto de encuentro de numerosos países y culturas, y, por ende, el propósito final es reflejar esta variedad en la confección de un vocabulario meta panhispánico. Para ello, la intención es extender y diversificar las muestras de informantes de las pruebas de disponibilidad léxica. Este hecho implica, a su vez, la asunción de más variables y la consecuente evolución y mejora de la herramienta de recogida de datos: *LexicalEx*.

6. Bibliografía citada

ALDERSON, J. Charles, 2005: *Diagnosing Foreign Language Proficiency*, Londres: Continuum.

ÁVILA MUÑOZ, Antonio, 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), 31- 43.

ÁVILA MUÑOZ, Antonio, 2017: "The available lexicon: A tool for selecting appropriate vocabulary to teach a foreign language", *Iranian Journal of Language Teaching Research* 5 (1), 71-91.

BARTOL HERNÁNDEZ, José Antonio, 2010: "Disponibilidad léxica y selección del vocabulario" en Rosa CASTAÑER MARTÍN y Vicente LAGÜENS GRACIA (eds.): *De moneda nunca usada. Estudios filológicos dedicados a José M. Enguita Utrilla*, Zaragoza: Institución Fernando el Católico, 85-107.

BURKETT, Theodore, 2017: *An investigation into the Use of Word List in University Foundation Programs in the United Arab Emirates*. Tesis doctoral, University of Exeter.

CONSEJO DE EUROPA, 2002: *Marco común europeo de referencia para las lenguas: Aprendizaje, Enseñanza, Evaluación*, Madrid: Ministerio de Educación, Cultura y Deporte.

DAVIES, Mark, 2005: "Vocabulary Range and Text Coverage: Insights from the Forthcoming Routledge Frequency Dictionary of Spanish" en David EDDINGTON (ed.): *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, Somerville, MA: Cascadilla Proceedings Project, 106-115.

DAVIES, Mark, y Kathy Hayward DAVIES, 2018: *A Frequency Dictionary of Spanish. Core Vocabulary for Learners*, Nueva York y Abingdon: Routledge.

DAVIES, Mark, y Timothy L. FACE, 2006: "Vocabulary Coverage in Spanish Textbooks: How Representative is It?" en Nuria SAGARRA y Almeida Jacqueline TORIBIO (eds.): *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, Somerville, MA: Cascadilla Proceedings Project, 132-143.

GARDNER, Dee, 2013: *Exploring Vocabulary. Language in Action*, Londres: Routledge.

GRÈVE, Marcel, y Frans VAN PASSEL, 1971: *Lingüística y enseñanza de lenguas extranjeras*, Madrid: Fragua.

HIDALGO-GALLARDO, Matías, 2017a: "La disponibilidad léxica como herramienta de evaluación interventiva del vocabulario en ELE: aplicación en un contexto universitario sinohablante", *SinoELE* 16 (septiembre), 43-63.

HIDALGO-GALLARDO, Matías, 2017b: *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.

HIDALGO-GALLARDO, Matías, 2019: *Fuentes y factores que inciden en el léxico disponible de estudiantes sinohablantes de ELE. Monografías ASELE 23*, ASELE.

CHAO, Hsueh, Marcella HU y Paul NATION, 2000: "Vocabulary density and reading comprehension", *Reading in a Foreign Language* 13 (1), 403-430.

INSTITUTO CERVANTES, 2006: *Plan curricular del Instituto Cervantes*, Madrid: Instituto Cervantes y Biblioteca Nueva.

JIMÉNEZ CALDERÓN, Francisco, y Anna RUFAT, 2019: "Estudio de léxico disponible para la selección de vocabulario meta del español en enseñanza universitaria" en Rosabel ROIG-VILA (ed.): *Investigación e innovación en la Enseñanza Superior. Nuevos contextos, nuevas ideas*, Barcelona: Octaedro, 234-251.

LAUFER, Batia, y Paul NATION, 1999: "A vocabulary size test of controlled productive ability", *Language Testing* 16 (1), 33-51.

LAUFER, Batia, y Geke C. RAVENHORST-KALOVSKI, 2010: "Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension", *Reading in a Foreign Language* 22 (1), 15-30.

LÓPEZ CHÁVEZ, Juan, 1992: "Alcances panhispánicos del léxico disponible", *Lingüística* 4, 26-124.

MEARA, Paul, y James MILTON, 2003: *X_Lex, The Swansea Levels Test*, Newbury: Expres.

NATION, Paul, 2006: "How Large a Vocabulary Is Needed for Reading and Listening?", *Canadian Modern Language Review* 63 (1), 59-82.

OGDEN, Charles, 1934: *The System of Basic English*, New York: Harcourt, Brace.

PAREDES GARCÍA, Florentino, 2015: "Disponibilidad Léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula", *Linred* XIII, 1-32.

ROGHANI, Shadan, y James MILTON, 2017: "Using Category Generation Tasks to Estimate Productive Vocabulary Size in a Foreign Language", *TESOL International Journal* 12 (1), 143-159.

SANTOS DÍAZ, Inmaculada, 2017: "Selección del léxico disponible: propuesta metodológica con fines didácticos", *Porta Linguarum* 27, 122-139.

SCHMITT, Norbert, Diane SCHMITT y Caroline CLAPHAM, 2001: "Developing and exploring the behavior of two new versions of the Vocabulary Levels Test", *Language Testing* 18, 55-88.

TOMÉ CORNEJO, Carmela, 2015: *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. Tesis doctoral, Universidad de Salamanca.

TROITIÑO, Sergio, 2017: "Implicaciones de crear materiales desde una perspectiva léxica", en Francisco HERRERA (ed.): *Enseñar léxico en el aula de español. El poder de las palabras*, Difusión: Barcelona.

WEST, Michael, 1953: *A General Service List of English Words*, Londres: Longman, Green.

WEST, Richard, 2012: *Word Family Framework*, TeachingEnglish, British Council.

7. Anexo. Resultados de la triangulación por orden de puntuación

ÍTEM LÉXICO	ID	RF	PCIC	NOCIÓN PCIC	PLD	PLF	PIC	TOTAL	
1	ojo	0.6250 (6)	169 (2)	A1	1.1. Partes del cuerpo	6	6	6	18
2	cabeza	0.7147 (1)	265 (5)	A2	1.1. Partes del cuerpo	6	6	5	17
3	pierna	0.6735 (2)	776 (16)	A2	1.1. Partes del cuerpo	6	6	5	17
4	brazo	0.6706 (3)	470 (9)	A2	1.1. Partes del cuerpo	6	6	5	17
5	mano	0.6647 (4)	135 (1)	A2	1.1. Partes del cuerpo	6	6	5	17
6	pie	0.6379 (5)	365 (7)	A2	1.1. Partes del cuerpo	6	6	5	17
7	dedo	0.5951 (7)	716 (15)	A2	1.1. Partes del cuerpo	5	6	5	16
8	nariz	0.5523 (8)	1570 (26)	A1	1.1. Partes del cuerpo	5	5	6	16
9	pelo	0.2995 (17)	873 (18)	A1	1.1. Partes del cuerpo	3	6	6	15
10	espalda	0.3196 (15)	942 (19)	A2	1.1. Partes del cuerpo	3	6	5	14
11	diente	0.2398 (20)	1365 (24)	A1	Aparece como parte de una unidad pluriverbal (<i>lavarse los dientes</i>) en 13.6. Higiene	3	5	6	14
12	cara	0.2178 (21)	337 (6)	A2	1.1. Partes del cuerpo	2	6	5	13
13	rodilla	0.4258 (11)	1839 (30)	B1	1.1. Partes del cuerpo	4	5	4	13
14	cuello	0.3592 (14)	1298 (23)	B1	1.1. Partes del cuerpo	4	5	4	13
15	cuerpo	0.0118 (117)	232 (3)	A1	Aparece como parte del título de la noción 1.1. Partes del cuerpo	1	6	6	13
16	pecho	0.2009 (24)	971 (20)	B1	1.1. Partes del cuerpo	2	6	4	12
17	corazón	0.1651 (27)	475 (10)	B1	1.1. Partes del cuerpo	2	6	4	12
18	hombro	0.2770 (19)	1146 (22)	B1	1.1. Partes del cuerpo	3	5	4	12
19	dedo del pie	0.0240 (76)	716 (15)	A2	No aparece el bloque como tal, pero sí sus elementos por separado en 1.1. Partes del cuerpo	1	6	5	12
20	labio	0.1703 (26)	1060 (21)	B1	Aparece como parte de una unidad pluriverbal (<i>barra de labios</i>) en 13.7. Estética	2	5	4	11
21	estómago	0.1224 (32)	2285 (36)	A2	1.1. Partes del cuerpo	2	4	5	11
22	oído	0.0714 (46)	1659 (29)	A2	1.1. Partes del cuerpo	1	5	5	11

23	muñeca	0.3015 (16)	2562 (43)	B1	1.1. Partes del cuerpo	3	4	4	11
24	piel	0.0313 (71)	670 (14)	B1	1.1. Partes del cuerpo	1	6	4	11
25	uña	0.3937 (12)	2578 (44)	B2	1.1. Partes del cuerpo	4	4	3	11
26	sangre	0.0111 (120)	555 (11)	B1	13.4. Centros de asistencia sanitaria	1	6	4	11
27	boca	0.5040 (9)	465 (8)	—	Aparece con un significado metafórico y como parte de una unidad pluriverbal (<i>boca de metro</i>) en 14.3. Sistema de transportes (A2)	5	6	0	11
28	barba	0.0029 (179)	2759 (47)	A1	1.2. Características físicas	1	4	6	11
29	cerebro	0.1486 (30)	1593 (27)	B2	1.1. Partes del cuerpo	2	5	3	10
30	hueso	0.0663 (47)	1649 (28)	B1	1.1. Partes del cuerpo	1	5	4	10
31	frente	0.0568 (51)	238 (4)	B2	1.1. Partes del cuerpo	1	6	3	10
32	garganta	0.0505 (56)	2359 (39)	A2	1.1. Partes del cuerpo	1	4	5	10
33	codo	0.3753 (13)	4384 (63)	B1	1.1. Partes del cuerpo	4	2	4	10
34	bigote	0.0027 (196)	3716 (57)	A1	1.2. Características físicas	1	3	6	10
35	músculo	0.1017 (40)	2715 (45)	B1	1.1. Partes del cuerpo	1	4	4	9
36	cintura	0.0794 (45)	2737 (46)	B1	1.1. Partes del cuerpo	1	4	4	9
37	columna (vertebral)	0.0519 (55)	1869 (31)	B2	1.1. Partes del cuerpo	1	5	3	9
38	oreja	0.4977 (10)	2095 (34)	—	—	5	4	0	9
39	ceja	0.2109 (22)	3700 (56)	B2	1.1. Partes del cuerpo	2	3	3	8
40	lengua	0.1647 (28)	586 (12)	—	—	2	6	0	8
41	muslo	0.1214 (33)	3382 (52)	B2	5.3. Alimentos	2	3	3	8
42	mejilla	0.1038 (38)	2365 (40)	B2	1.1. Partes del cuerpo	1	4	3	8
43	pulmón	0.0578 (50)	3158 (51)	B1	1.1. Partes del cuerpo	1	3	4	8
44	cabello	0.0419 (63)	1510 (25)	C1	13.7. Estética	1	5	2	8
45	tobillo	0.2860 (18)	5123 (72)	B1	1.1. Partes del cuerpo	3	1	4	8
46	rostro	0.0218 (82)	596 (13)	C2	1.2. Características físicas	1	6	1	8
47	nervio	0.0098 (124)	2856 (49)	B2	1.1. Partes del cuerpo	1	4	3	8
48	pestaña	0.2090 (23)	4703 (68)	B2	1.1. Partes del cuerpo	2	2	3	7
49	vientre	0.0313 (72)	2294 (37)	C1	1.1. Partes del cuerpo	1	4	2	7
50	pechos	0.0240 (77)	971 (20)	—	Aparece en su forma singular, con la consiguiente diferencia semántica, en 1.1. Partes del cuerpo (B2)	1	6	0	7

51	planta (del pie)	0.0136 (112)	768 (17)	—	—	1	6	0	7
52	palma	0.0123 (115)	2127 (35)	C1	1.1. Partes del cuerpo	1	4	2	7
53	índice	0.0034 (168)	2540 (42)	C1	1.1. Partes del cuerpo	1	4	2	7
54	pata	0.0027 (197)	2091 (33)	C1	1.1. Partes del cuerpo	1	4	2	7
55	barriga	0.1874 (30)	—	B1	1.1. Partes del cuerpo	2	0	4	6
56	tronco	0.0854 (43)	2777 (48)	C2	1.1. Partes del cuerpo	1	4	1	6
57	hígado	0.0539 (53)	4387 (64)	B2	1.1. Partes del cuerpo	1	2	3	6
58	articulación	0.0127 (113)	4636 (66)	B2	1.1. Partes del cuerpo	1	2	3	6
59	huevo	0.0079 (130)	1994 (32)	—	Aparece con un significado diferente en 5.3. Alimentos (A1)	1	5	0	6
60	puño	0.0038 (158)	3060 (50)	C1	1.1. Partes del cuerpo	1	3	2	6
61	cadera	0.147 (107)	3595 (55)	C2	1.2. Características físicas	1	3	1	5
62	intestino	0.0382 (67)	5998 (76)	B2	1.1. Partes del cuerpo	1	1	3	5
63	seno	0.0163 (95)	2301 (38)	—	—	1	4	0	5
64	cráneo	0.0159 (98)	4321 (62)	C1	1.1. Partes del cuerpo	1	2	2	5
65	riñón	0.0139 (111)	5010 (71)	B2	1.1. Partes del cuerpo	1	1	3	5
66	arteria	0.0036 (166)	5604 (75)	B2	1.1. Partes del cuerpo	1	1	3	5
67	costado	0.0029 (180)	2440 (41)	—	—	1	4	0	5
68	muela	—	—	A2	1.1. Partes del cuerpo	0	0	5	5
69	culo	0.1306 (31)	4527 (65)	—	—	2	2	0	4
70	párpado	0.1070 (36)	3581 (54)	—	—	1	3	0	4
71	barbilla	0.0619 (48)	—	B2	1.1. Partes del cuerpo	1	0	3	4
72	recto	0.0317 (70)	3990 (58)	—	Aparece con otro significado en 16.3. Matemáticas (B2)	1	3	0	4
73	costilla	0.0218 (83)	—	B2	1.1. Partes del cuerpo	1	0	3	4
74	vena	0.0160 (97)	3405 (53)	—	—	1	3	0	4
75	tendón	0.0145 (108)	—	B2	1.1. Partes del cuerpo	1	0	3	4
76	faringe	0.0070 (131)	—	—	Aparece la palabra de familia <i>faringitis</i> en 13.1. Salud y enfermedades	1	0	3	4
77	pulgar	0.0059 (140)	6742 (77)	C1	1.1. Partes del cuerpo	1	1	2	4
78	átomo	0.0042 (153)	—	B2	16.5. Física y química	1	0	3	4
79	esqueleto	—	5186 (74)	B2	1.1. Partes del cuerpo	0	1	3	4
80	coco	—	4649 (69)	C1	1.1. Partes del cuerpo	0	2	2	4

81	ombligo	0.1079 (35)	—	C1	1.1. Partes del cuerpo	1	0	2	3
82	trasero	0.0395 (65)	4247 (60)	—	—	1	2	0	3
83	páncreas	0.0377 (68)	—	C1	1.1. Partes del cuerpo	1	0	2	3
84	órgano re-productor	0.0238 (78)	—	—	—	1	0	2	3
85	órgano	0.0205 (86)	—	C1	Aparece como parte del bloque <i>donar órganos</i> en 13.4. Centros de asistencia sanitaria	1	0	2	3
86	nalga	0.0181 (90)	4645 (67)	—	—	1	2	0	3
87	vello	0.0173 (91)	—	C1	13.7. Estética	1	0	2	3
88	pupila	0.0155 (102)	4180 (59)	—	—	1	2	0	3
89	tímpano	0.0127 (114)	—	C1	1.1. Partes del cuerpo	1	0	2	3
90	espinilla	0.0115 (118)	—	C1	13.7. Estética	1	0	2	3
91	lomo	0.0059 (139)	4655 (70)	—	Aparece con un significado diferente en 5.3. Alimentos (B1)	1	2	0	3
92	vértebra	0.0052 (150)	—	C1	1.1. Partes del cuerpo	1	0	2	3
93	cartílago	0.0033 (169)	—	C1	16.2. Biología	1	0	2	3
94	colon	0.0033 (170)	—	C1	1.1. Partes del cuerpo	1	0	2	3
95	peca	0.0029 (181)	—	C1	1.2. Características físicas	1	0	2	3
96	lumbares	0.0029 (182)	—	C1	13.3. Síntomas	1	0	2	3
97	aparato digestivo	0.0029 (183)	—	C1	1.1. Partes del cuerpo	1	0	2	3
98	cana	—	—	B2	1.1. Partes del cuerpo	0	0	3	3
99	antebrazo	0.1549 (29)	—	—	—	2	0	0	2
100	extremidad	0.0429 (61)	—	C2	1.1. Partes del cuerpo	1	0	1	2
101	vesícula	0.0212 (84)	—	C2	Aparece como parte del bloque <i>tumor en la vesícula</i> en 13.1. Salud y enfermedades	1	0	1	2
102	mandíbula	0.0159 (99)	5127 (73)	—	—	1	1	0	2
103	tibia	0.0148 (105)	—	C2	1.1. Partes del cuerpo	1	0	1	2
104	clavícula	0.0112 (119)	—	C2	1.1. Partes del cuerpo	1	0	1	2
105	fosa nasal	0.0052 (151)	—	C2	1.1. Partes del cuerpo	1	0	1	2
106	coronilla	0.0042 (154)	—	C2	1.1. Partes del cuerpo	1	0	1	2
107	tronco inferior	0.0036 (167)	—	—	—	1	0	1	2
108	retina	0.0027 (198)	8356 (78)	—	—	1	1	0	2
109	diafragma	0.0027 (199)	—	C2	18.5 Fotografía	1	0	1	2

110	entraña	—	4289 (61)	—	—	0	2	0	2
111	napia	—	—	C1	11. Partes del cuerpo	0	0	2	2
112	córnea	—	—	C1	11. Partes del cuerpo	0	0	2	2
113	sistema circulatorio	—	—	C1	11. Partes del cuerpo	0	0	2	2
114	nuca	0.1130 (34)	—	—	—	1	0	0	1
115	pene	0.1039 (37)	—	—	—	1	0	0	1
116	sistema nervioso	0.1035 (39)	—	—	Aparece <i>nervioso</i> como parte del bloque <i>trastorno nervioso</i> en 13.3. Síntomas	1	0	0	1
117	vagina	0.0983 (41)	—	—	—	1	0	0	1
118	talón	0.0967 (42)	—	—	Aparece con otro significado en 15.1. Finanzas y bolsa (C2)	1	0	0	1
119	gemelo	0.0818 (44)	—	—	—	1	0	0	1
120	moflete	0.0586 (49)	—	—	—	1	0	0	1
121	jugo gástrico	0.0564 (52)	—	—	—	1	0	0	1
122	pantorrilla	0.0539 (54)	—	—	—	1	0	0	1
123	abdomen	0.0491 (57)	—	—	—	1	0	0	1
124	glúteo	0.0472 (58)	—	—	—	1	0	0	1
125	ano	0.0469 (59)	—	—	—	1	0	0	1
126	axila	0.0463 (60)	—	—	—	1	0	0	1
127	torso	0.0424 (62)	—	—	—	1	0	0	1
128	radio	0.0417 (64)	—	—	—	1	0	0	1
129	húmero	0.0394 (66)	—	—	—	1	0	0	1
130	empeine	0.0347 (69)	—	—	—	1	0	0	1
131	pectorales	0.0300 (73)	—	—	—	1	0	0	1
132	ingle	0.0280 (74)	—	—	—	1	0	0	1
133	teta	0.0272 (75)	—	—	—	1	0	0	1
134	esternocleido-mastoideo	0.0234 (79)	—	—	—	1	0	0	1
135	pómulo	0.0227 (80)	—	—	—	1	0	0	1
136	tríceps	0.0221 (81)	—	—	—	1	0	0	1
137	testículo	0.0208 (85)	—	—	—	1	0	0	1
138	cuádriceps	0.0200 (87)	—	—	—	1	0	0	1

139	bíceps	0.0187 (88)	—	—	—	1	0	0	1
140	cerebelo	0.0184 (89)	—	—	—	1	0	0	1
141	falange	0.0169 (92)	—	—	—	1	0	0	1
142	pelvis	0.0169 (93)	—	—	—	1	0	0	1
143	fémur	0.0166 (94)	—	—	—	1	0	0	1
144	esófago	0.0161 (96)	—	—	—	1	0	0	1
145	peroné	0.0158 (100)	—	—	—	1	0	0	1
146	nudillo	0.0156 (101)	—	—	—	1	0	0	1
147	iris	0.0151 (103)	—	—	—	1	0	0	1
148	pezón	0.0149 (104)	—	—	—	1	0	0	1
149	bazo	0.0148 (106)	—	—	—	1	0	0	1
150	apéndice	0.0142 (109)	—	—	—	1	0	0	1
151	peras	0.0142 (110)	—	—	—	1	0	0	1
152	cúbito	0.0122 (116)	—	—	—	1	0	0	1
153	vulva	0.0108 (121)	—	—	—	1	0	0	1
154	tráquea	0.0104 (122)	—	—	—	1	0	0	1
155	cogote	0.0100 (123)	—	—	—	1	0	0	1
156	tórax	0.0098 (125)	—	—	Aparece el adjetivo <i>torácico</i> en 13.2. Heridas y traumatismos (C2)	1	0	0	1
157	pubis	0.0091 (126)	—	—	—	1	0	0	1
158	colmillo	0.0085 (127)	—	—	—	1	0	0	1
159	rabadilla	0.0085 (128)	—	—	—	1	0	0	1
160	laringe	0.0085 (129)	—	—	Aparece la palabra de la familia <i>laringólogo</i> en 13.4. Centros de asistencia sanitaria (B2)	1	0	0	1
161	mentón	0.0070 (132)	—	—	—	1	0	0	1
162	globo ocular	0.0070 (133)	—	—	—	1	0	0	1
163	lagrimal	0.0070 (134)	—	—	—	1	0	0	1
164	ovario	0.0064 (135)	—	—	—	1	0	0	1
165	abdominales	0.0063 (136)	—	—	Aparece como adjetivo en 13.2. Heridas y traumatismos (C2)	1	0	0	1
166	sesos	0.0063 (137)	—	—	—	1	0	0	1
167	genitales	0.0062 (138)	—	—	—	1	0	0	1
168	lunar	0.0058 (141)	—	—	—	1	0	0	1

169	vejiga	0.0057 (142)	—	—	—	1	0	0	1
170	esternón	0.0056 (143)	—	—	—	1	0	0	1
171	sobaco	0.0056 (144)	—	—	—	1	0	0	1
172	intestino grueso	0.0055 (145)	—	—	Aparece solo <i>intestino</i> en 1.1. Partes del cuerpo (B2)	1	0	0	1
173	intestino delgado	0.0055 (146)	—	—	Aparece solo <i>intestino</i> en 1.1. Partes del cuerpo (B2)	1	0	0	1
174	lóbulo	0.0054 (147)	—	—	—	1	0	0	1
175	partes íntimas	0.0054 (148)	—	—	—	1	0	0	1
176	vaso sanguíneo	0.0054 (149)	—	—	—	1	0	0	1
177	rótula	0.0052 (152)	—	—	—	1	0	0	1
178	alveolo	0.0042 (155)	—	—	—	1	0	0	1
179	frenillo	0.0042 (156)	—	—	—	1	0	0	1
180	clítoris	0.0042 (157)	—	—	—	1	0	0	1
181	ternilla	0.0038 (159)	—	—	—	1	0	0	1
182	tronco superior	0.0038 (160)	—	—	Aparece solo <i>tronco</i> en 1.1. Partes del cuerpo (C2)	1	0	0	1
183	coxis	0.0037 (161)	—	—	—	1	0	0	1
184	vía respiratoria	0.0037 (162)	—	—	Aparece <i>vía</i> en 13.5. Medicina y medicamentos (C2) y <i>respiratoria</i> en 13.3. Síntomas (C1)	1	0	0	1
185	papila gustativa	0.0037 (163)	—	—	—	1	0	0	1
186	yema	0.0037 (164)	—	—	Aparece con otro significado en 5.3. Alimentos (B1)	1	0	0	1
187	aparato reproductor	0.0037 (165)	—	—	—	1	0	0	1
188	sistema respiratorio	0.0033 (171)	—	—	Aparece solo <i>respiratorio</i> como parte del bloque <i>trastorno respiratorio</i> en 13.3. Síntomas (C1)	1	0	0	1
189	meñique	0.0032 (172)	—	—	—	1	0	0	1
190	dedo gordo del pie	0.0032 (137)	—	—	—	1	0	0	1
191	dedo corazón	0.0032 (174)	—	—	—	1	0	0	1
192	bulbo raquídeo	0.0030 (175)	—	—	—	1	0	0	1
193	coño	0.0030 (176)	—	—	—	1	0	0	1
194	flexura	0.0030 (177)	—	—	—	1	0	0	1

195	aparato respiratorio	0.0030 (178)	—	—	Aparece solo <i>respiratorio</i> como parte del bloque <i>trastorno respiratorio</i> en 13.3. Síntomas (C1)	1	0	0	1
196	orificio nasal	0.0029 (184)	—	—	—	1	0	0	1
197	tripa	0.0029 (185)	—	—	—	1	0	0	1
198	anular	0.0029 (186)	—	—	Aparece como verbo en 3.1. Datos personales	1	0	0	1
199	mama	0.0029 (187)	—	—	—	1	0	0	1
200	batata	0.0029 (188)	—	—	—	1	0	0	1
201	polla	0.0029 (189)	—	—	—	1	0	0	1
202	espina dorsal	0.0029 (190)	—	—	—	1	0	0	1
203	papada	0.0029 (191)	—	—	—	1	0	0	1
204	cóclea	0.0029 (192)	—	—	—	1	0	0	1
205	dorso	0.0028 (193)	—	—	—	1	0	0	1
206	uretra	0.0028 (194)	—	—	—	1	0	0	1
207	sistema digestivo	0.0028 (195)	—	—	No aparece el bloque como tal, pero sí <i>digestivo</i> en solitario en 13.1. Salud y enfermedades (C1) y en combinación con <i>aparato digestivo</i> en 1.1. Partes del cuerpo (C1)	1	0	0	1
208	carrillo	0.0027 (200)	—	—	—	1	0	0	1
209	bronquio	0.0027 (201)	—	—	Aparece la palabra de la familia <i>bronquitis</i> en 13.1. Salud y enfermedades (B2)	1	0	0	1
210	bronquiolo	0.0027 (202)	—	—	—	1	0	0	1
211	omoplato	0.0027 (203)	—	—	—	1	0	0	1
212	bíceps femoral	0.0027 (204)	—	—	—	1	0	0	1
213	tripas	0.0027 (205)	—	—	—	1	0	0	1
214	cachete	0.0027 (206)	—	—	—	1	0	0	1
215	crystalino	0.0027 (207)	—	—	—	1	0	0	1
216	crisma	—	—	C2	1.1. Partes del cuerpo	0	0	1	1
217	mollera	—	—	C2	1.1. Partes del cuerpo	0	0	1	1
218	cervical	—	—	C2	1.1. Partes del cuerpo	0	0	1	1
219	cuero cabelludo	—	—	C2	1.1. Partes del cuerpo	0	0	1	1
220	epidermis	—	—	C2	1.1. Partes del cuerpo	0	0	1	1

221	extremidad superior	—	—	C2	1.1. Partes del cuerpo	0	0	1	1
222	extremidad inferior	—	—	C2	1.1. Partes del cuerpo	0	0	1	1