METODOLOGÍA Y ANÁLISIS DE LA DISPONIBILIDAD LÉXICA DE ESPAÑOL PARA GRUPOS ESPECÍFICOS: INMIGRANTES ADULTOS

Pablo Fernández-Merino Gutiérrez
Universidad Internacional de La Rioja

Resumen: El presente artículo supone una reflexión, desde el punto de vista práctico, destinada a facilitar la comprensión y el trabajo en el ámbito de la disponibilidad léxica cuando abordamos un grupo específico de informantes. Explicaremos el proceso de elaboración de un diccionario de léxico disponible fuera del marco del *Proyecto Panhispánico*; resaltaremos los pasos de la investigación previa, la preparación metodológica, la recogida y el análisis de datos y la exposición de los mismos.

Palabras clave: diccionario, inmigrante, léxico disponible, lingüística aplicada, enseñanza de lenguas.

Summary: This paper involves a consideration from a practical point of view which is set aside for making easier our comprehension and our labour in Lexical Availability when we work with a specific group of informants. We explain the elaboration process of a dictionary of lexical availability out of the *Panhispanic Project*. We also highlight the steps of the previous research, the methodological planning, the collection and analyses of data and its exposition.

Key words: dictionary, immigrant, Lexical Availability, applied linguistics, language teaching.

1. Introducción

Lo primero que debemos señalar es el arduo trabajo que hay detrás de estas páginas. Las afirmaciones y ejemplos aquí expuestos han sido obtenidos de la reflexión realizada durante la elaboración del texto *Disponibilidad léxica de inmigrantes en Castilla y León* (Fernández-Merino Gutiérrez), fruto de una investigación realizada a 326 inmigrantes adultos de las capitales de provincia de Castilla y León en el año 2012. Dicho texto toma a su vez referencia del estudio

preliminar *Disponibilidad léxica de inmigrantes en Castilla y León: Valladolid* (Fernández-Merino Gutiérrez). Ambas investigaciones fueron posibles gracias a la financiación de la Junta de Castilla y León y el Fondo Social Europeo[1].

Durante la elaboración de estos trabajos encontramos en el léxico disponible un valioso aliado para realizar una fotografía lingüística del vocabulario de un grupo concreto de informantes en un lugar determinado. Esta metodología nos permite obtener una lista de palabras para exponer en diccionarios de uso.

Este artículo pretende ser una pequeña guía vinculada a los datos obtenidos en los textos mencionados. De esta manera, facilitaremos el desarrollo de estudios posteriores en grupos específicos.

2. Conocer el léxico disponible

Cuando nos referimos al léxico disponible, hablamos de lo que Michéa denominó *palabras temáticas*. El autor francés mostró cómo las palabras ligadas al automatismo psicológico son verbos, adjetivos y elementos funcionales: "Il est des termes dont l'emploi, plus ou moins indépendant de circonstances et des besoins, est étroitement lié à l'automatisme psychologique" (Michéa 338-344); es decir que si observamos solo las palabras más presentes en un vocabulario fundamental, no tendrían cabida términos como *bicicleta* o *libro* pertenecientes a contextos específicos; por tanto, será necesario realizar un tipo de prueba que enmarque al hablante en el contexto que quiera estudiarse. Las *palabras temáticas* corresponden al léxico usado por los hablantes y que solo aparece cuando se producen determinadas asociaciones. López Morales (*Los estudios*) explora la potencialidad de los trabajos de léxico disponible y refleja su importancia a la hora de señalar ámbitos de estudio específicos que provoquen en el hablante los estímulos necesarios para expresar determinadas palabras (López Morales, *Léxico*); de esta manera los términos disponibles en el lexicón mental del individuo se actualizan en un contexto concreto. Así pues, el léxico disponible necesita un ámbito de estudio donde las actualizaciones léxicas se hagan efectivas; por

ello, se recurre a pruebas léxicas asociativas controladas que utilizan los centros de interés, o campos léxicos, para que actúen como estímulos:

> Las palabras que se usan habitualmente en cualquier situación comunicativa constituyen el léxico básico, mientras que las utilizadas solo cuando abarcamos un tema concreto integran el léxico disponible. (Rodríguez y Muñoz, 10)

El *Proyecto Panhispánico* engloba las investigaciones más relevantes en torno a la disponibilidad léxica en español. Como veremos a la hora de abordar la metodología, esta propuesta utiliza dieciséis centros de interés basados en los establecidos por los estudios franceses (Gougenheim *et al.*, 1964) con la intención de:

> 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, servir de punto de partida para análisis posteriores. (Bartol y Hernández)

El desarrollo y los avances del *Proyecto*, además de sus aplicaciones a las distintas ramas de la lingüística, han sido recogidos por varios investigadores de forma muy completa. Resultan fundamentales los trabajos ya mencionados de López Morales, las explicaciones de Borrego Nieto y Moreno Fernández, las aproximaciones cognitivas (Ferreira y Gary; Hernández; Ellis, Hernández e Izura) y las relaciones sociolingüísticas (Samper Padilla). El acercamiento a las recopilaciones de Bellón, Samper y Samper, por un lado y Samper y Samper, por otro, resulta muy útil para asimilar las principales líneas de investigación de esta metodología y observar los avances en cada una de ellas.

3. Seleccionar el grupo de estudio

La elección del grupo de estudio, en nuestro caso, partió de la necesidad concreta de la sociedad española en un

momento determinado: "la cuestión del aprendizaje de la lengua del país de acogida es un tema de capital importancia en las políticas migratorias y de integración de los inmigrantes en el conjunto de la Unión Europea. Varios textos de la U.E. desde 2003 así lo atestiguan" (Villarreal, 77). Nuestra investigación se encaminó a favorecer esta integración lingüística.

Cualquier estudio realizado a una comunidad concreta de hablantes puede resultar especialmente relevante; sin embargo, la experiencia señala que si las motivaciones y los objetivos primordiales se encuentran expuestos con claridad, el vínculo con el grupo de estudio y su acercamiento al mismo resultará más sencillo. En otras palabras, tener claro a quién queremos estudiar y por qué nos facilitará la labor posterior y nos dará una perspectiva real para adaptar la metodología. Si nuestros informantes son ejecutivos abordaremos el estudio de manera diferente que si son estudiantes de lenguas o miembros de una comunidad en peligro de exclusión.

3.1 Estudio previo

Las palabras de Slagter resaltan la importancia de esta parte de la investigación, en especial, de cara a una posible aplicación en la práctica docente del mismo: "sólo cuando se haya descrito el grupo-meta podremos intentar determinar lo más exactamente posible qué es lo que este necesitará hacer, en nuestro caso, con respecto al aprendizaje de un idioma extranjero" (Slagter, 3). La descripción del grupo de estudio y una investigación previa del mismo supone, por tanto, un paso previo clave en nuestra propuesta metodológica.

Aquí abordamos las características generales más relevantes de los informantes. En nuestro caso es interesante señalar los rasgos que describen al grupo, para lo cual nos apoyamos en la bibliografía pertinente. Nuestro grupo está compuesto por inmigrantes no hispanohablantes que proceden de países en vías de desarrollo y que, en numerosas ocasiones, encuentran dificultades en su integración (Andión *La inmigración*). Otro punto que debe señalarse es que todos los participantes se encuentran en edad labo-

ral que es, precisamente, la franja de edad más presente en España (Andión *El sexo y la edad*)

Nuestro estudio se realizó en los centros de español para inmigrantes de Castilla y León, lo que implica un acercamiento e investigación acerca del tipo de enseñanza que aquí se recibe y cómo se plantea; es decir, un acercamiento teórico a la enseñanza a adultos y a la enseñanza de español para inmigrantes. Gracias al planteamiento de un marco teórico sólido descubrimos, entre otras cosas, que trabajar con adultos la enseñanza de lenguas, si bien no resulta una limitación a la hora del aprendizaje, supone que será el propio alumno quien marque su propia experiencia de contacto con la lengua meta (Krashen, 10). Descubrimos también que la enseñanza del español a inmigrantes se concibe en este tipo de cursos como un elemento clave no solo en relación a la competencia lingüística, sino también a la integración (García Parejo, 7). Además, destacamos la heterogeneidad de los grupos (Miquel López, 2) o que las características intrínsecas de este tipo de enseñanza han servido como excusa para evitar abordar de una manera profesional este tipo de enseñanza (Hernández y Villalba, 28).

3.2 Estudio posterior

Si queremos plantear un trabajo, no solo general, sino descriptivo de nuestro grupo concreto, nuestra propuesta es ayudarnos de encuestas realizadas a los propios centros colaboradores[2] y de las variables cuestionadas a los informantes[3]. En otras palabras, el estudio posterior ayuda a reflejar la situación concreta de aquellos que han contestado a nuestra encuesta y prueba de disponibilidad y nos permite observar características específicas que resultarán muy útiles para abordar conclusiones posteriores.

En nuestro caso, las variables nos ayudan a ver que nuestro grupo de informantes se concentra principalmente entre los 16 y los 45 años de edad (80%) y que pertenecen en su mayoría a Marruecos (52%). La mayor parte de ellos (84%) no se encuentran trabajando en el momento de realización de la encuesta y su nivel de estudios es muy variado, siendo preponderantes aquellos con estudios pri-

marios (30%). La mayoría ha pasado más de tres años en España (52%), pese a que su asistencia a las clases de español suele ser reciente, pues llevan menos de un año (60%). Además suelen concentrarse en entidades privadas (80%), es decir, en asociaciones u organizaciones no gubernamentales.

Los centros de enseñanza de español para inmigrantes, es decir, los lugares donde desarrollamos la encuesta, fueron determinantes en el estudio que planteamos como muestra. Gracias a las conclusiones extraídas de las encuestas observamos cómo se organiza la docencia y las características específicas de los alumnos.

Las clases se desarrollan principalmente en los núcleos urbanos y, aunque existe docencia pública, en su mayoría es asumida por los centros privados y está impartida por voluntarios y trabajadores sin formación específica en segundas lenguas. La falta de especialización se ve reflejada en una falta de consenso en cuanto a criterios metodológicos y de organización en el número de alumnos; podemos también encontrar clases donde los alumnos se dividen por género, nivel o conveniencia horaria. Esta enseñanza forma parte del programa planteado para la integración global del individuo

Respecto a los alumnos, anotamos que son un grupo heterogéneo compuesto por personas en edad laboral, no hispanohablantes y provenientes, en su mayoría, de una realidad con posibilidades de exclusión. Su asistencia a las clases es irregular y sus niveles académicos resultan muy diferentes, desde el analfabetismo en su propia lengua hasta la posesión de títulos universitarios. La nacionalidad depende de la realidad concreta del barrio donde está asentado el centro.

En nuestra investigación, la exposición de estas ideas sirve como un pequeño análisis del medio que ayuda a observar no solo las dificultades de este tipo de enseñanza, sino también las posibilidades, las necesidades… Independientemente del carácter de nuestro estudio, hemos descubierto que resulta fundamental realizar un análisis que señale la situación real del grupo de informantes y la realidad de su centro de enseñanza o, en su defecto, del lugar donde se realice la prueba (si fuera, por ejemplo, una

empresa la que colaborara en nuestro estudio los parámetros de análisis serían diferentes, pero la realización de este apartado sería también muy interesante).

4. Cambios propuestos a la metodología del léxico disponible

4.1 Centros de interés

Es conveniente que nuestra propuesta esté sustentada en los estudios realizados y comparta los principios lingüísticos del *Proyecto Panhispánico*. Debemos tener en cuenta que todos los cambios que realicemos en relación a los centros de interés supondrán un distanciamiento con los estudios originales e impedirán una comparación con ellos. Pese a esto, la revisión de los centros de interés es necesaria para la elaboración de una propuesta propia basada en los informantes y en los objetivos que queramos conseguir. La reunión de Bilbao de 1999 y los encuentros posteriores de San Millán de la Cogolla en 2003 y 2005 fijan dieciséis centros de interés basados en la propuesta francesa:: 01 'Partes del cuerpo', 02 'La ropa', 03 'Partes de la casa (sin muebles)', 04 'Los muebles de la casa', 05 'Alimentos y bebidas', 06 'Objetos colocados en la mesa para la comida', 07 'La cocina y sus utensilios', 08 'La escuela: muebles y materiales', 09 'Calefacción, iluminación y medios de airear un edificio', 10 'La ciudad', 11 'El campo', 12 'Medios de transporte', 13 'Trabajos del campo y del jardín', 14 'Animales', 15 'Juegos y distracciones', 16 'Profesiones y oficios'.

Los investigadores encargados de fijar esta propuesta reconocen que los centros funcionan como estímulos artificiales para provocar las diferentes listas de palabras con las que luego trabajaremos. El hecho de describir el estímulo como artificial nos lleva a tratar de, en la medida de lo posible, naturalizar los centros de interés para que no resulten ajenos a nuestros informantes. Los cambios son comunes a muchos trabajos de léxico disponible, ya sea por experimentación o adaptación al grupo de estudio. El ya citado trabajo de recopilación de Samper Padilla y Samper Hernández muestra como distintos autores se han aproxi-

mado de diferente forma a los centros de interés. Cambiar el número de los centros, su nombre o su enfoque, no supone por tanto una práctica nueva. Se recomienda adaptarlos al tipo de estudio que se quiera realizar y a las características específicas de los informantes, lo que ya señalaron los investigadores franceses:

> Explorer le vocabulaire disponible, c'est rechercher comment les mots s'organisent entre eux par la seule force de l'association des idées (...) avant la sélection que nous impose l'adaptation à une situation concrète (...), l'influence de circonstances variées et de besoins divers. (Michéa, 340)

En nuestro caso, las decisiones tomadas a este respecto parten también de un estudio de la bibliografía referente al léxico disponible de español como segunda lengua, donde encontramos, por un lado, los que seleccionan los mismos centros que el *Proyecto Panhispánico*, por ejemplo Samper Hernández y Carcedo, junto con aquellos que plantean una revisión de los mismos para intentar adaptarlos a la situación del informante, Sánchez-Saus y Sifrar Kalan (*Leksikalna*). Nuestro estudio se enmarcaría dentro de este segundo grupo.

De esta manera, los cambios que realicemos servirán de puente entre los planteamientos teóricos y nuestro grupo concreto de informantes. A continuación ahondaremos en los principales criterios que, en nuestro caso, tuvimos en cuenta para la realización del estudio en inmigrantes adultos.

La primera decisión fue acortar el número de centros. Esta reducción es crucial debido a la disponibilidad horaria de los informantes y la organización de las clases en sesiones de una hora; en segundo lugar, el propio cansancio de quienes participan en la prueba induce a una menor compresión y/o despreocupación. Pese a que esta última afirmación se basa en la percepción del investigador, puede añadirse que en otros trabajos realizados con el mismo tipo de informantes (Fernández-Merino *Valladolid*) varias encuestas tuvieron que ser retiradas por falta de respuestas en los últimos centros.

Otro de los problemas que puede plantear esta cuestión es la peculiaridad del grupo y qué realidades queremos

reflejar. Nuestra investigación trabaja desde la necesidad comunicativa inmediata, lo cual mediatizará la utilización de unos centros de interés u otros.

De cara a proponer los centros de interés definitivos para el estudio del léxico disponible de inmigrantes adultos debemos reducir el número de centros y seleccionar aquellos que resulten más relevantes para nuestros informantes.

En un principio, se tiene en cuenta que el número de centros debería reducirse para dejar un tiempo prudencial a la explicación de los mismos por parte del investigador y poder encuadrar la prueba en sesiones de entre 50 y 60 minutos. Se eliminan los campos que, a priori, podrían resultar menos productivos y presentan una mayor dificultad de comprensión. Desde un primer momento se descartan el 13 'Trabajos de campo y de jardín' (con un 3,95 de palabras por informante en el trabajo de Samper Hernández, 45) y el 09 'Calefacción, iluminación y medios de airear un edificio' (con un promedio de 5,28 en dicho trabajo). Una vez eliminados los ámbitos donde las respuestas podrían tener una mayor dificultad para los informantes, centramos la actualización en las posibilidades de los estudiantes. Según las referencias del profesorado los informantes estarán dentro de los niveles A1, en su mayoría, y A2 del *Marco Común Europeo de Referencia para las Lenguas* (MCERL) desarrollado por el *Council of Europe*. En estos niveles la riqueza léxica se describe a través de un hablante que:

A2	Tiene suficiente vocabulario para desenvolverse en actividades habituales y en transacciones cotidianas que comprenden situaciones y temas conocidos.
A1	Tiene suficiente vocabulario para expresar necesidades comunicativas básicas. Tiene suficiente vocabulario para satisfacer necesidades sencillas de supervivencia.

Tabla 1: Descripción de la riqueza léxica en los niveles A1 y A2 del MCERL

En ambos descriptores encontramos la utilización de una palabra clave que ya hemos mencionado: necesidad; de esta manera, debemos volver la vista hacia las necesidades concretas del grupo para la realización de una encuesta que responda a la realidad.

Además del MCERL resultó fundamental en el desarrollo de nuestro estudio localizar manuales específicos de enseñanza de español destinados a nuestros informantes entre los que destaca: *Español como nueva lengua. Orientaciones para un curso de emergencia destinado a inmigrantes adultos* (Isabel). Aquí pudimos observar ámbitos concretos de aprendizaje que deberíamos tener en cuenta en nuestros centros.

La propuesta final, por tanto, supone una revisión del *Proyecto Panhispánico*, el *MCERL* (Council of Europe), propuestas concretas relacionadas con el grupo (Isabel) y el primer trabajo realizado en torno a inmigrantes adultos (Fernández-Merino *Valladolid*), del cual sacamos los datos que citamos a continuación. Decidimos eliminar en la propuesta definitiva los centros: 'Juegos y diversiones' y 'El campo'. El primero no está contemplado en la propuesta de Isabel, tiene una media muy baja de palabras por informante (6,86) y cuatro informantes no supieron qué responder. En cuanto a 'El campo', resulta interesante descubrir que tiene uno de los promedios de palabra por informante más bajo (6,53) y que hasta siete informantes lo dejaron en blanco. Aunque, los datos del centro 'Administración pública' (5,66 palabras por informante, siete abstenciones) indican una gran complejidad, decidimos contemplarlo por su pertinencia y relación directa con el grupo de informantes.

Respecto al nombre con los que se presentarán los centros en la propuesta final vemos que este se ha simplificado. Si bien el enunciado presenta un acortamiento que ayuda a que se identifique mejor con el campo léxico, la explicación del mismo a los informantes se verá ampliada cuando hablemos de la prueba.

En resumen, del total de los doce centros contemplados, ocho corresponden, directa o indirectamente, a los propuestos por el *Proyecto Panhispánico* (señalados en cursiva en la tabla 2). La nueva versión permite la comparación con varios centros del *Panhispánico* y soluciona los problemas expuestos relacionados con el tiempo de la prueba:

Proyecto Panhispánico	*Español como nueva lengua*(Isabel)	Valladolid 2010	Castilla y León 2013
01. Partes del cuerpo	Presentarse y dar datos personales.	*01. La ropa*	*01. Partes del cuerpo*
02. La ropa	Conocer la ciudad, el barrio o el pueblo.	*02. La ciudad*	*02. Alimentos y bebidas*
03. Partes de la casa (sin muebles)	Utilizar el transporte público.	*03. Alimentos y bebidas*	*03. La ciudad*
04. Los muebles de la casa	Hacer compras.	04. Familia y amigos	*04. La ropa*
05. Alimentos y bebida	Ir al médico.	*05. La casa: partes de la casa y muebles*	05. La salud
06. Objetos colocados en la mesa para la comida	Buscar una vivienda para alquilar.	*06. El campo*	*06. Transportes*
07. La cocina y sus utensilios	Buscar empleo.	*07. Partes del cuerpo*	07. Ir de compras
08. La escuela: muebles y materiales	Empezar a trabajar.	*08. Profesiones y oficios*	08. Administración pública
09. Calefacción, iluminación y medios de airear un edificio	Tramitar papeles y hacer gestiones.	09. Ir de compras: medidas, cantidades y tiendas	*09. La casa*
10. La ciudad		10. La salud	*10. Profesiones*
11. El campo		*11. Juegos y diversiones*	11. La familia
12. Medios de transporte		12. Administración pública: tramitar papeles y hacer gestiones	*12. Animales*
13. Trabajos del campo y del jardín		*13. Medios de transporte*	
14. Animales		*14. Animales*	
15. Juegos y distracciones			
16. Profesiones y oficios			

Tabla 2: Comparación de las propuestas de los centros de interés

4.2 Variables

Las características generales del grupo sirven para enmarcar el estudio en unos patrones socioculturales concretos. Planteamos un estudio al estilo de Parejo y Ambadiang: plural, que vaya más allá de lo exclusivamente lingüístico y personal. Reconocemos la importancia de los aspectos individuales, pero lo que más nos interesa es poder recoger variables de tipo general.

Lo primero que debemos tratar son las variables que otros estudios ya publicados de disponibilidad léxica en

contextos similares, en nuestro caso, en español como lengua extranjera. Samper Hernández, a raíz de los trabajos de Carcedo, centra su estudio en las variables: sexo, lengua materna, conocimiento de otras lenguas, nivel de español. Por otro lado, más recientemente, Sánchez-Saus decide contemplar las siguientes variables: sexo, nivel de español según el marco, frecuencia de uso, conocimiento de lenguas románicas. Contemplar unas u otras variables dependerá esencialmente de dos factores: la idiosincrasia del grupo y qué queremos estudiar.

Desde un punto de vista práctico es interesante señalar que un gran número de opciones en el análisis complica el estudio a la hora de realizar comparaciones más precisas y un estudio cualitativo; por otro lado, dificulta la labor de encontrar datos suficientes en determinadas opciones. Pese a ello, el aspecto positivo es que se realizan perfiles de informantes más adecuados a la realidad. Se llega por tanto a la conclusión de que, para el estudio estadístico concreto, será imprescindible reducir el número de opciones y asegurarse un número mínimo de informantes por cada una.

Otra propuesta sugestiva es que las variables utilizadas en la encuesta que contesta el informante no coincidan exactamente con las que luego se utilicen en el estudio (al igual que en otros trabajos como Samper Hernández); de esta manera, aunque se requiere un mayor esfuerzo por parte del investigador, se podrán perfilar adecuadamente los distintos grupos de las variables una vez que tengamos todos los datos.

El trabajo de Fernández-Merino (*Castilla y León*) realiza un estudio general y descubre las variables más productivas en la investigación de la disponibilidad léxica en inmigrantes adultos. Un mayor número de variables garantiza la caracterización del grupo al mismo tiempo que permite su posterior selección puesto que:

> No es posible reconocer de antemano qué tipo de variables sociales van a actuar sobre la variación lingüística (...). La razón de ello es doble: por una parte, por el hecho de que la variabilidad no tiene por qué manifestarse de la misma forma en comunidades de habla diferentes, dado que los factores sociales no

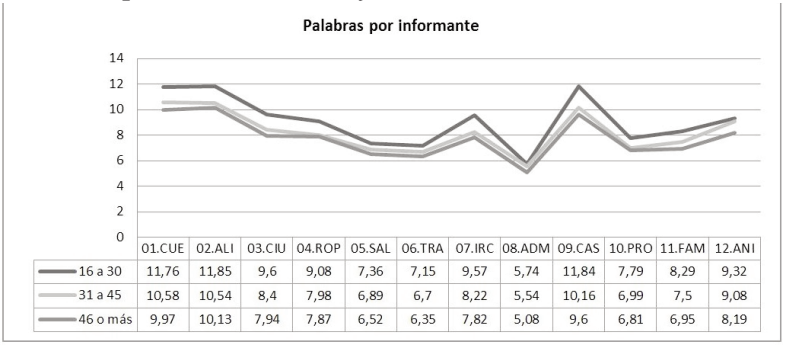
> actúan sobre la lengua de manera uniforme. Pero también, porque estos mismos factores no se configuran de la misma forma en sociedades distintas. (Blas Arroyo, 157)

Al igual que en la selección de centros, es importante ser conscientes y justificar la recogida y uso de unas u otras variables. El trabajo con inmigrantes adultos, por ejemplo, no contempla 'Nivel de español' o 'Frecuencia con la que los informantes emplean el español' puesto que no resultan prácticas en el desarrollo de nuestra encuesta; la primera porque, para la gran mayoría de los informantes, el nivel representa una constante, puesto que se encuentran en un A1/A2 según *el MCERL* (*Council of Europe*), aunque algunos presenten carencias en determinadas competencias que impide clasificarlos dentro del nivel inicial. La frecuencia de uso del español fuera del contexto escolar tampoco se considera oportuna debido a su subjetividad.

Sí contemplamos otras variables por su importancia en estudios sociolingüísticos, por su presencia en los trabajos del *Proyecto Panhispánico* o por su interés específico para esta investigación. A continuación realizamos una breve aproximación a las variables y opciones planteadas en Fernández-Merino (*Castilla y León*) y las conclusiones más relevantes obtenidas gracias al análisis cuantitativo de las mismas. La observación detenida de las variables a través del estudio de las diferentes opciones, en lo que a número de palabras, vocablos e índice de cohesión[4] se refiere, permite realizar una serie de aproximaciones teóricas que aquí recogemos; sin embargo, las conclusiones más visibles suelen corresponder al número de palabras por informante y centro de interés.

Sexo[5]. Las opciones de esta variable son, evidentemente, 'Hombre' o 'Mujer'. Se trata de una variable presente en la mayoría de estudios sociológicos debido a los papeles específicos de uno y otro género que, en nuestro caso, también serán relevantes (Andión, *Sexo y edad*). La variable 'Sexo' muestra un mayor número de palabras por informante en las mujeres; sin embargo, para comprobar si dicha diferencia es o no determinante, necesitaríamos someter los datos a pruebas matemáticas concretas[6].

Edad. Es una de las variables que no es necesario contemplar en otros estudios de léxico disponible porque los grupos son homogéneos en este aspecto. En nuestro caso, sin embargo, es un factor que puede ser productivo, puesto que la franja de edad adulta es muy amplia. La división de grupos etarios es muy variada en los estudios sociolingüísticos: "Trudgill, por ejemplo, no establece en Norwich los mismos grupos de edad que Martínez Martín en Burgos" (Williams, 48). En otras palabras, las opciones de los grupos se adaptan a las pretensiones concretas de los diferentes estudios. Nuestro estudio se decide por tres grupos etarios basados en la propuesta de Andión (*Sexo y edad*) y la experiencia de Fernández-Merino (*Valladolid*): de 16 a 30, de 30 a 45, 46 o más. Según los datos analizados en el último estudio el número de palabras por informante es ligeramente superior en los más jóvenes en todos los centros.

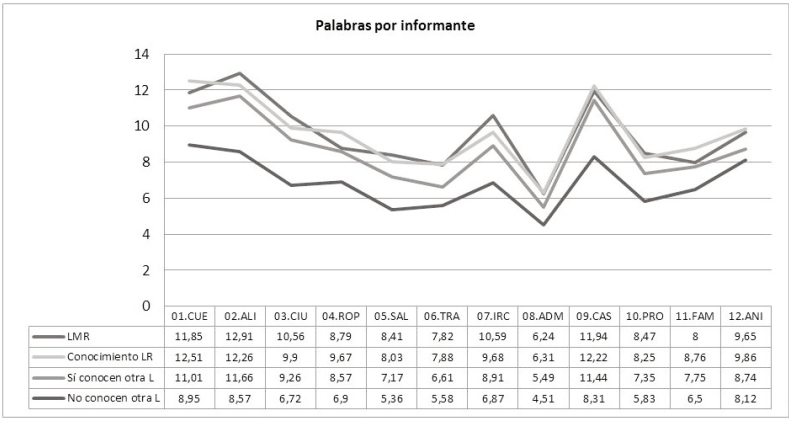


	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
16 a 30	11,76	11,85	9,6	9,08	7,36	7,15	9,57	5,74	11,84	7,79	8,29	9,32
31 a 45	10,58	10,54	8,4	7,98	6,89	6,7	8,22	5,54	10,16	6,99	7,5	9,08
46 o más	9,97	10,13	7,94	7,87	6,52	6,35	7,82	5,08	9,6	6,81	6,95	8,19

Gráfico 1: Palabras por informante según la variable "Edad"

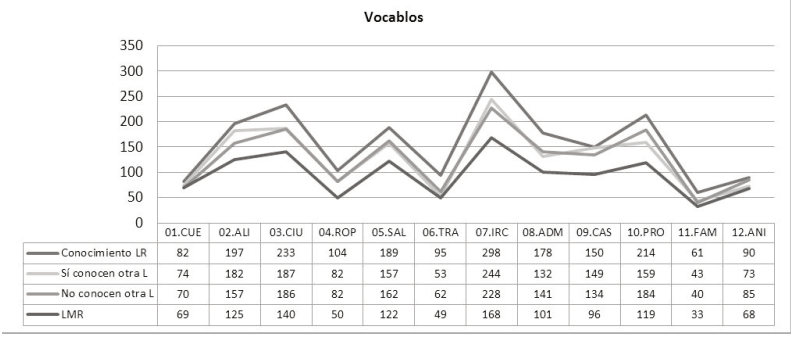
Lengua Materna. Recordemos que pretendemos enfocar este estudio de una manera general, para lo cual decidimos anotar la reflexión de Sánchez-Saus y señalar el conocimiento de lenguas románicas como un factor clave. De esta manera, dividiremos a los informantes en cuatro grupos en relación con su lengua materna y el conocimiento de otras lenguas: en primer lugar, aquellos que tienen como lengua materna una lengua románica (LMR) (por ejemplo portugués o rumano); en segundo lugar, aquellos que tienen conocimiento en cualquier grado de una lengua románica (LR), pero no la explicitan de manera clara como primera lengua (hablantes de soniké y francés o de árabe y francés); en ter-

cer lugar, aquellos cuya lengua materna (LM) no es románica y tienen conocimientos de otras lenguas que tampoco son románicas (hablantes de búlgaro y turco, por ejemplo); por último, anotaremos aquellos cuya LM no es románica y no tienen conocimiento de otras lenguas (informantes que hablan exclusivamente inglés o árabe). Las observaciones realizadas en torno al conocimiento de lenguas permiten ver que el hecho de tener lengua materna románica o conocer otra lengua (románica o no) favorece los datos de palabras por informante y vocablos respecto a aquellos que no conocen ninguna otra lengua.



	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
LMR	11,85	12,91	10,56	8,79	8,41	7,82	10,59	6,24	11,94	8,47	8	9,65
Conocimiento LR	12,51	12,26	9,9	9,67	8,03	7,88	9,68	6,31	12,22	8,25	8,76	9,86
Sí conocen otra L	11,01	11,66	9,26	8,57	7,17	6,61	8,91	5,49	11,44	7,35	7,75	8,74
No conocen otra L	8,95	8,57	6,72	6,9	5,36	5,58	6,87	4,51	8,31	5,83	6,5	8,12

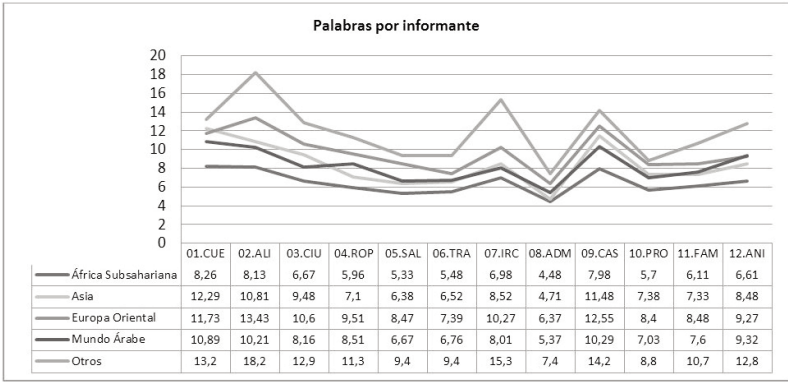
Gráfico 2: Palabras por informante según la variable "Lengua materna"



	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
Conocimiento LR	82	197	233	104	189	95	298	178	150	214	61	90
Sí conocen otra L	74	182	187	82	157	53	244	132	149	159	43	73
No conocen otra L	70	157	186	82	162	62	228	141	134	184	40	85
LMR	69	125	140	50	122	49	168	101	96	119	33	68

Gráfico 3: Número de vocablos según la variable "Lengua materna"

País de procedencia. Debido al carácter generalista de este estudio, decidimos realizar una división geográfica lo más adecuada posible en relación a los datos. El objetivo es observar la relevancia de la zona de procedencia aunque, como se comentó en el apartado anterior, los estudios específicos de un grupo de estudio muy numeroso (marroquíes, en este caso) también podrían resultar sumamente interesantes. Como máximo exponente anotamos el grupo 'Mundo Árabe', donde se ubican aquellos informantes procedentes de países que pertenecen a la Liga de Estados Árabes y se sitúan en el Norte de África; en segundo lugar, por representatividad, encontramos 'Europa Oriental', donde incluimos los países separados de la antigua Unión Soviética, además de aquellas naciones encuadradas en este grupo por la Organización de las Naciones Unidas; el tercer lugar lo ocupan aquellos que pertenecen a países del 'África Subsahariana'. Por último, se anotan los países del continente asiático y el cajón de sastre nombrado como 'Otros' donde ubicamos aquellos países que no entran en ninguna de las categorías anteriores y no tienen informantes suficientes para considerarse por separado. La procedencia de los informantes muestra una gran diferencia en palabras por informante entre los que proceden de países desarrollados y aquellos cuyo grupo incluye países en vías de desarrollo.



	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
África Subsahariana	8,26	8,13	6,67	5,96	5,33	5,48	6,98	4,48	7,98	5,7	6,11	6,61
Asia	12,29	10,81	9,48	7,1	6,38	6,52	8,52	4,71	11,48	7,38	7,33	8,48
Europa Oriental	11,73	13,43	10,6	9,51	8,47	7,39	10,27	6,37	12,55	8,4	8,48	9,27
Mundo Árabe	10,89	10,21	8,16	8,51	6,67	6,76	8,01	5,37	10,29	7,03	7,6	9,32
Otros	13,2	18,2	12,9	11,3	9,4	9,4	15,3	7,4	14,2	8,8	10,7	12,8

Gráfico 4: Palabras por informante según la variable "País de procedencia"

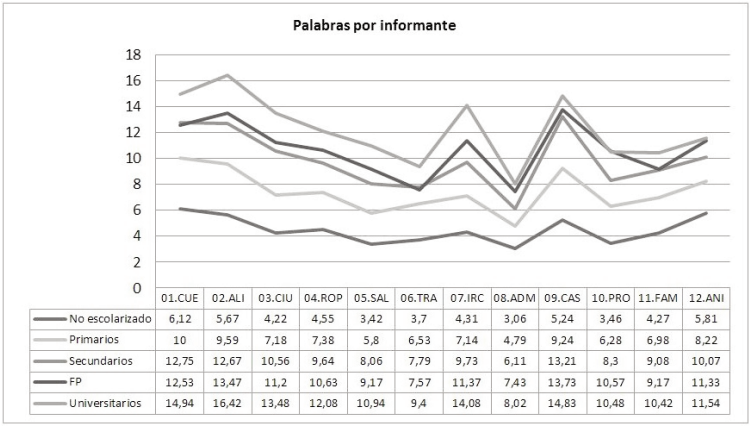
Nivel de estudios. Respecto a esta variable, se confirman las palabras de Blas Arroyo: "A juicio de numerosos sociolingüistas es, probablemente, el nivel educativo de los informantes el que contribuye a estratificar sociolectalmente de forma más clara las comunidades de habla" (227-228). En relación al grupo que nos interesa destacamos también que:

> Además, el nivel educativo constituye otro elemento a tener en cuenta con independencia del tipo de proyecto migratorio. Así, entre las mujeres de procedencia urbana se da en general un nivel más alto de estudios y un conocimiento mayor de una segunda lengua como el francés. Esto facilita el contacto con otros registros del español como el escrito. (Soto y El-Madkouri, 21).

Respecto a las opciones, debemos señalar que son aproximadas: 'No escolarizado' y 'Universitarios' son sin duda las más exactas; 'Primarios' reúne informantes con algún tipo de escolarización hasta alrededor de los 10 años, mientras que aquellos que permanecieron escolarizados entre los 11 y los 18 años, sin pasar a ninguna de las otras ramas propuestas, se encuentran situados en 'Secundarios'; la opción 'Formación Profesional' incluye a todos aquellos que hayan estudiado como realizar un oficio en concreto, ya sea a través de título oficial o no. Es importante señalar que, en los últimos casos, no hablamos de titulación, sino de aproximación formativa.

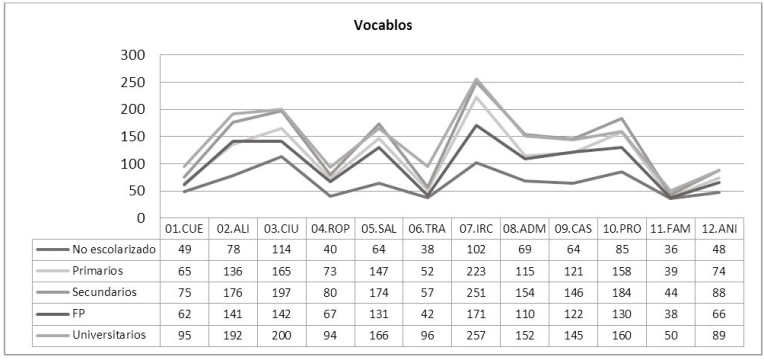
En relación a este punto resulta muy útil plantear una encuesta abierta de variables y, a partir de ahí, seleccionar las más interesantes y pertinentes teniendo en cuenta qué queremos estudiar así como la teoría sociolingüística en torno a la variable concreta. Si quisieramos profundizar más en este aspecto sería muy interesante plantear un cruce de variables para descubrir individuos coincidentes.

La variable más determinante y en la cual podemos observar una progresión más clara es el nivel de estudios. En esta ocasión podemos afirmar que los encuestados con un mayor nivel de estudios muestran un mayor número de palabras por informante. En cuanto al número de vocablos, aquellos no escolarizados serán quienes aporten las cifras más bajas.



	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
No escolarizado	6,12	5,67	4,22	4,55	3,42	3,7	4,31	3,06	5,24	3,46	4,27	5,81
Primarios	10	9,59	7,18	7,38	5,8	6,53	7,14	4,79	9,24	6,28	6,98	8,22
Secundarios	12,75	12,67	10,56	9,64	8,06	7,79	9,73	6,11	13,21	8,3	9,08	10,07
FP	12,53	13,47	11,2	10,63	9,17	7,57	11,37	7,43	13,73	10,57	9,17	11,33
Universitarios	14,94	16,42	13,48	12,08	10,94	9,4	14,08	8,02	14,83	10,48	10,42	11,54

Gráfico 5: Palabras por informante según la variable "Nivel de estudios"



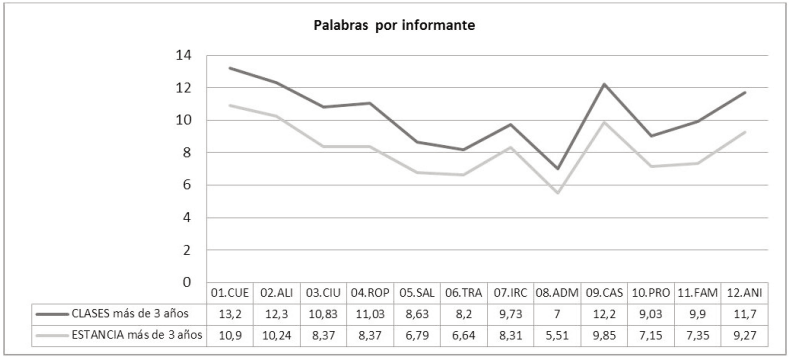
	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
No escolarizado	49	78	114	40	64	38	102	69	64	85	36	48
Primarios	65	136	165	73	147	52	223	115	121	158	39	74
Secundarios	75	176	197	80	174	57	251	154	146	184	44	88
FP	62	141	142	67	131	42	171	110	122	130	38	66
Universitarios	95	192	200	94	166	96	257	152	145	160	50	89

Gráfico 6: Vocablos según la variable "Nivel de estudios"

Tiempo de estancia en España y en clases de español. En relación al tiempo de estancia, gracias a los datos obtenidos en el primer trabajo: "se intuye que la ruptura lingüística en inmigrantes respecto a su léxico se produce pasados los tres años de estancia" (Fernández-Merino *Valladolid*, 117). Por esta razón, se plantean tres opciones (menos de 1 año, entre 1 y 3 años, más de 3 años). Respecto a la otra variable, la experiencia en los lugares de docencia y las entrevistas con los educadores señalan una tendencia de un determinado grupo de

informantes a permanecer durante varios años en un mismo centro. Decidimos plantear las mismas opciones que en el apartado anterior.

Ambas variables están mediatizadas por el número de informantes en las diferentes opciones y muestran datos similares en cuanto al número de palabras por informante; sin embargo, podemos apreciar cómo el hecho de llevar más de tres años en clases de español proporciona un mayor número de palabras por informante y supone también un punto de inflexión si lo relacionamos con el tiempo de estancia en España; lo que supone que, a partir de los tres años en las clases, la experiencia académica muestra unos valores superiores que la simple permanencia en el país.



	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07.IRC	08.ADM	09.CAS	10.PRO	11.FAM	12.ANI
CLASES más de 3 años	13,2	12,3	10,83	11,03	8,63	8,2	9,73	7	12,2	9,03	9,9	11,7
ESTANCIA más de 3 años	10,9	10,24	8,37	8,37	6,79	6,64	8,31	5,51	9,85	7,15	7,35	9,27

Gráfico 7: Palabras por informante comparando más de tres años de permanencia en clases y en España

Trabajo. En el estudio de Valladolid (Fernández-Merino, 2010) se tuvo en cuenta la *Clasificación Internacional Uniforme de Ocupaciones* desarrollada por la Organización Internacional del Trabajo (OIT: 2007) para catalogar las respuestas libres de los informantes. Esta variable suscitaba algunos problemas de concepto pues daba lugar a respuestas confusas, además de proporcionar un número excesivo de opciones; en la propuesta definitiva decidimos preguntamos, simplemente, si los informantes desarrollan o no algún oficio en el momento de la prueba.

Tipo de centro. Decidimos mantener la nomenclatura 'público', para referirnos a los Centros de Educación para

Adultos (CEPA) y 'privado', para señalar a las asociaciones y organizaciones. Pese a que el planteamiento de las clases y el carácter no oficial de la enseñanza puede ser común en ambas, resulta interesante comprobar si existe una relación directa con el léxico disponible.

Las variables 'Trabajo' y 'Tipo de centro' no son en este caso un factor determinante en los aspectos analizados, debido, principalmente, a las amplias diferencias de informantes en sus opciones. Si bien no son relevantes de cara al análisis de disponibilidad léxica, como hemos podido ver, sirven para precisar la caracterización del grupo (recordemos que hemos anotado que la mayoría de los informantes se encuentran en desempleo y acuden a centros privados para aprender español).

Respecto a este punto, resulta muy útil plantear una encuesta abierta de variables y, a partir de ahí, seleccionar las más interesantes y pertinentes teniendo en cuenta qué queremos estudiar así como la teoría sociolingüística en torno a la variable concreta. Si quisieramos profundizar más en este aspecto sería muy interesante plantear un cruce de variables para descubrir individuos coincidentes.

4.3 Desarrollo de la prueba

Tan importante como la fundamentación teórica es el desarrollo concreto de la prueba con el grupo de informantes. Debemos adaptarnos a su disponibilidad horaria y facilitar el buen hacer de los lugares que deseen colaborar con nuestro estudio. En nuestro caso, la labor y el trabajo conjunto con los responsables de las clases fueron fundamentales para el rápido y correcto desarrollo de esta parte de la investigación[7]. Podemos esbozar la secuencia de la prueba a través de los siguientes momentos:

Presentación del investigador, del proyecto y del uso que va a hacerse de los datos obtenidos. Ha sido importante recalcar el valor de la prueba no como un examen, sino, más bien, como un juego. La explicación de la elaboración de un diccionario y el carácter de investigación formal favorece la motivación de los informantes.

Realización de la encuesta de variables. Según el nivel del grupo puede enfocarse de varias maneras. Lo ideal es

que se vaya explicando pregunta por pregunta y respondiendo todos a la vez para agilizar el proceso; en ocasiones, el investigador y el responsable de la clase tuvieron que acercarse a los informantes individualmente para explicar los diferentes apartados. Pese a que la mayor parte de los informantes contestaron a todas las preguntas, aún hemos encontrado casos de *no sabe o no contesta* (NS/NC) que deben recogerse en el estudio de las variables.

Explicación detallada del procedimiento de realización de la prueba de léxico disponible. Se utiliza el centro '00 Los colores' para que el mecanismo de realización pueda asimilarse; de esta manera, podremos enumerar ejemplos que no interfieran con el desarrollo de la prueba.

Se responde a la prueba de disponibilidad. Entregamos a los informantes, junto con la encuesta, doce hojas[8]; cada una de ellas tiene el número del centro de interés correspondiente y está dividida en dos columnas de líneas paralelas numeradas con la intención de facilitar la explicación; se han diseñado un total de cuarenta espacios, suficientemente separados para facilitar la escritura; en ningún caso ha hecho falta escribir fuera de ellos. Hemos decidido ceñirnos lo más posible a la metodología del *Proyecto Panhispánico* y dedicar un total de dos minutos por dentro de interés (Samper y Samper). De esta manera, pese a los cambios que se desarrollarán en apartados posteriores, la propuesta metodológica coincide con otros trabajos de léxico disponible.

Enunciado de cada uno de los centros. En primer lugar se enunció el título del centro y luego se ofreció la explicación consecuente haciendo hincapié en los verbos, puesto que así no influimos en otras clases de palabra que puedan aparecer directamente relacionadas con el centro de interés. El lenguaje no verbal del investigador también facilitó la comprensión.

Orden de los centros. En lo que al orden de los centros se refiere, son relevantes las reflexiones de Tomé al respecto. Según la revisión de los centros que realiza la autora, es conveniente controlar las áreas temáticas afines para que queden separadas, así como diseñar varios órdenes para evitar que el cansancio no afecte siempre a los mismos centros.

En el grupo de informantes con el que trabajamos es también importante asegurar que la dinámica de la encuesta quede asimilada con lo que los primeros centros de interés deberán ser relativamente sencillos de contestar; además, trataremos de proponer unos órdenes que intercalen centros de mayor o menor dificultad de comprensión. El aspecto motivacional jugará por tanto un valor relevante a la hora de elaborar el orden de los centros, puesto que intentaremos realizar una prueba lo más asequible posible.

Centro	Ausencias	Campos en conflicto (más de tres palabras coincidentes).
La ciudad	3	'La salud' (hospital, farmacia, médico) 'Administración pública' (hospital, ayuntamiento, banco, policía, colegio, médico, comisaría) 'Ir de compras' (supermercado, tienda , farmacia , bar, mercado) 'Medios de transporte' (coche, (auto)bus , tren, taxista), 'Profesiones' (policía, médico, taxista)
La salud	6	'Partes del cuerpo' (cabeza, nariz, pie, pierna, espalda, corazón, estómago, tripa, barriga, garganta) 'La ciudad' (hospital, farmacia, médico) 'Administración pública' (hospital, médico, centro de salud) 'Profesiones' (profesor, médico, enfermero, doctor)
Profesiones	3	'La ciudad' (policía, médico, taxista) 'La salud' (profesor, médico, enfermero, doctor) 'Administración pública' (policía, médico, abogado, trabajar)
La casa	0	
La ropa	3	
Ir de compras	4	'Alimentos y bebidas' (pan, leche, arroz, carne, pescado, fruta, pollo, verdura, queso, lechuga, comida) 'La ciudad' (supermercado, tienda , farmacia , bar, mercado)
Administración pública	7	'La ciudad' (hospital, ayuntamiento, banco, policía, colegio, médico, comisaría) 'La salud' (hospital, médico, centro de salud) 'Profesiones' (policía, médico, abogado, trabajar)
Transportes	1	'La ciudad' (coche, (auto)bus , tren, taxista)
Alimentos y bebidas	0	'Ir de compras' (pan, leche, arroz, carne, pescado, fruta, pollo, verdura, queso, lechuga, comida)

Partes del cuerpo	0	'La salud' (cabeza, nariz, pie, pierna, espalda, corazón, estómago, tripa, barriga, garganta)
Animales	1	
Familia	1	

Tabla 3: Centros de interés, ausencias, y campos en conflicto según los datos de Valladolid

Los diferentes órdenes propuestos evitan juntar campos en conflicto (que compartan palabras), comienzan con centros que tuvieran cero o una ausencia en el trabajo de Valladolid y terminan con un centro de estas características e intercalan los centros, a priori, más complicados con otros más sencillos. Los diferentes órdenes propuestos evitan juntar campos en conflicto, comienzan con centros de cero y una ausencia, terminan con un centro de estas características e intercalan los centros, a priori, más complicados con otros más sencillos:

Orden 1	01.CUE	02.ALI	03.CIU	04.ROP	05.SAL	06.TRA	07IRC	08.ADM	9.CAS	10.PRO	11.FAM	12.ANI
Orden 2	01.ANI	02.FAM	03.PRO	04.CAS	05.ADM	06.IRC	07.TRA	08.SAL	09.ROP	10.CIU	11.ALI	12.CUE
Orden 3	01.TRA	02.CAS	03.ROP	04.ADM	05.IRC	06.FAM	07.SAL	08.ANI	09.CIU	10.CUE	11.PRO	12.ALI
Orden 4	01.ALI	02.PRO	03.CUE	04.CIU	05.ANI	06.SAL	07.FAM	08.IRC	09.ADM	10.ROP	11.CAS	12.TRA

Tabla 4: Órdenes de los centros de interés realizados

4.4 Edición

Los recursos en línea proporcionados por la Real Academia Española facilitan la labor de edición. La regularización de términos se ha realizado fundamentalmente a partir del *Diccionario de la lengua española*[9] de la Real Academia Española en su versión de acceso en línea de la 22ª edición. El elemento capital respecto a la presencia del léxico ha sido el *Corpus de Referencia del Español Actual* (CREA) que ha resuelto dudas en cuanto a incluir una forma de los términos u otra. También se ha recurrido a la versión en línea del *Diccionario panhispánico de dudas*[10] de la Real Academia Española. Otros recursos lexicográficos útiles son: el *Diccionario de uso del español*[11] (Moliner), el *Diccionario del español actual*[12] (Seco, Olimpia y Ramos)

y el *Gran diccionario de uso del español* actual[13] (Sánchez Pérez, 2001). Las cuestiones de índole gramatical se pueden consultar en la *Nueva gramática de la lengua española* de la Real Academia Española (2009).

Los criterios de edición del presente estudio toman como punto de partida la propuesta del *Proyecto Panhispánico* reflejada en el trabajo de Samper Padilla (*Criterios*). La especificidad del grupo de informantes no hispanohablantes nos orienta a contemplar también los trabajos específicos de Samper Hernández y Sánchez-Saus. La autora canaria ya planteó que el problema no reside únicamente en la limpieza de materiales o, lo que es lo mismo, la regularización ortográfica, sino también en el alcance de las asociaciones o la unificación de distintas variantes con ciertas unidades léxicas (Samper Hernández, 22). La edición adecuada y justificada del léxico es clave en la obtención de un corpus que sea útil desde el punto de vista de la investigación lingüística. Por esta razón, es muy importante mantener durante todo el trabajo los mismos criterios de edición y explicar las decisiones específicas. Para un correcto seguimiento de las realizaciones y los elementos regularizados decidimos expresar entre corchetes[14] la propuesta del informante y marcaremos en cursiva la opción recogida en el diccionario; los nombres de los centros de interés se marcarán, como hasta ahora, con comillas simples.

4.4.1Criterios comunes a todos los trabajos de disponibilidad léxica

4.4.1.1 Normas básicas

Eliminación de términos repetidos

Consideramos términos repetidos en un mismo informante no solo los que tienen varias apariciones, sino también aquellos que, tras haber realizado la regularización de variantes meramente flexivas y la corrección ortográfica pertinente, poseen la misma forma; lo mismo ocurre en el caso de los acortamientos, los grupos y locuciones nominales y en otras expresiones editadas. Encontraremos esta realidad, por ejemplo, con *moto* y *motocicleta* en 06 'Los transportes' (ambas regularizadas como *moto(cicleta)*).

Anotamos también como términos repetidos aquellos que se incluyen en el enunciado principal del centro. Desaparecen, por tanto, *compra* en 07 'Ir de compras' o *animal* en 12 'Los animales'. Se eliminan también los verbos de apoyo que se utilizaron para realizar la explicación del centro de interés. De esta manera, evitamos la mayor parte de estructuras oracionales que no tienen cabida en este tipo de estudio.

Corrección ortográfica

Durante la explicación de la prueba se especificó claramente que los informantes intentaran incluir las palabras que conocían sin importar si estaban o no correctamente escritas; es nuestra labor ajustar las expresiones a la norma. En algunos casos, tuvimos la oportunidad de charlar con ellos tras la prueba y se comprobó que realizaciones gráficas incomprensibles sí tenían su correspondencia en la expresión oral, por ejemplo [lyocomp] (no contemplado) en el centro 03 'La ciudad' para el informante refleja *Camp Nou*[15]; este hecho confirma que la competencia en expresión oral está más desarrollada que la competencia en expresión escrita. Algunos de los errores que analizaremos están directamente relacionados con las interferencias gráficas y fónicas, y supone una de las limitaciones de este tipo de estudios.

Unificación ortográfica

No hemos de olvidar que la unificación incidirá directamente en los datos de disponibilidad que obtengamos. El criterio fundamental es el utilizado por la RAE en su *Diccionario de la lengua española*. De esta manera, encontraremos los términos *güisqui*, *fútbol* o *chófer* tal y como son aceptados o recomendados por el DPD. En estos casos decidimos mantener la propuesta del DRAE. Siempre que el vocablo sea original de una lengua diferente al castellano, vendrá marcado en cursiva en los listados (*slip*, *leggings, snowboard*).

También se han hallado palabras que no están reflejadas en el DRAE y que, pese a la propuesta del DPD, se han regularizado teniendo en cuenta la forma recomendada por los diccionarios de uso o aquella más presente en el CREA. De esta manera, se prefiere la realización *parking* a [parquin] con la correspondiente cursiva.

Unificación de derivados regulares

Los casos de variación flexiva se han unificado en masculino singular, independientemente de la presencia mayoritaria en femenino o en plural (de esta manera [enfermera] será recogido como *enfermero* y [bomberos] como *bombero*).

Sin embargo, ha primado el criterio de uso para palabras que, pese a tener su entrada en el DRAE en singular, las encontramos principal o exclusivamente en plural (*gafas*, *gafas de sol*, *vacaciones*...); por el contrario, las formas de aquellos sustantivos que designan un objeto de partes simétricas (*pantalón*) o bien dos objetos simétricos (*media*), que pueden aparecer tanto en plural o en singular, se reflejan en singular (*braga*, *calzoncillo*…).

Se ha conservado la diferencia entre masculino y femenino siempre cuando conllevara un cambio de significado (*secador*, *secadora*, por ejemplo) o bien tuviera una repercusión directa en la grafía de la palabra, ya fuera por heteronimia (*toro*, *vaca*) o por lejanía ortográfica de sus pares sexuados (*actor*, *actriz*). Así pues, se reflejan aquellos pares con entrada propia en el DRAE (además de los señalados se incluyen también: *ternero*, *ternera* y *mulo*, *mula*).

Derivación apreciativa

Al igual que en otros trabajos de léxico disponible, eliminamos los sufijos apreciativos en casos como [colita] *cola*, o [pajarito] *pájaro*. Se mantienen aquellos con entrada propia en el DRAE que tienen un concepto concreto asignado, por ejemplo *alfombrilla* o *platillo*.

Unificación de formas plenas y acortamientos

Existen en nuestro idioma palabras susceptibles de ser acortadas. Hallamos, al igual que Marta Samper: [bici] y [bicicleta] o [tele] y [televisión]. Optamos por la misma solución que ella y la unificación se lleva a cabo con las formas *bici(cleta)* y *tele(visión)*, lo que indica que han aparecido ambas variantes en la prueba. En cada centro se reflejarán las adaptaciones oportunas.

Otros elementos de forma parentética

Encontramos grupos léxicos que pueden aparecer con presencia o ausencia de alguno de sus componentes. Al igual que en los acortamiento, utilizamos la forma parentética para señalarlos: *(cuarto de) baño*, *(vino) tinto*,

entre otros. Sánchez-Saus habla de elipsis originadas por combinatoria léxica para referirse a estas expresiones que "no siempre están totalmente lexicalizadas, por lo que en ocasiones el compuesto ha aparecido en las encuestas con su forma íntegra y, en otras, con algún elemento elidido" (Sánchez-Saus, 232).

Marcas comerciales

No hay unanimidad en lo referente a la incorporación de marcas comerciales a los estudios de disponibilidad. En el presente trabajo hemos decidido anotar todas las marcas que aparecen en el estudio. Esta misma opción ya ha sido tomada por otros investigadores, además de Gómez Devís, como Mateo García y Galloso. La disponibilidad léxica de las marcas comerciales muestra, en parte, el propio proceso de lexicalización de la marca y constituye un recurso fundamental en esta comunidad de habla concreta, puesto que incluimos en este grupo nombres propios de superficies comerciales. La forma de regularización se ha adaptado a la grafía original del nombre de la marca ya sea en castellano (*Zara**, *Dia** o *El árbol**), francés (*Carrefour**) o alemán (*Lidl**) con la inclusión de un asterisco al final de la palabra. Las marcas comerciales también pueden sufrir acortamientos, de tal manera señalamos *Coca(-Cola)** y *Pepsi (Cola)**. En el estudio original no consideramos conveniente plantear el problema de lexicalización de las marcas; por esta razón, nos limitamos a transcribir los nombres sin opinar acerca de este aspecto.

Tratamiento de los sintagmas nominales

En un primer momento se planteó dar cabida tan solo a los grupos nominales presentes en el DRAE o en los otros diccionarios de uso vistos anteriormente y expresar el resto de grupos a través de su núcleo; sin embargo, este planteamiento encierra varios problemas. En primer lugar, encontramos la dificultad de hacer referencia a determinados grupos léxicos que no aparecen en el DRAE, ni en los otros diccionarios consultados y que, por analogía, parece tener sentido admitir. EL DRAE admite *vino blanco* y *vino tinto*, pero no *judía verde* o *alubia blanca* como unidad léxica. Por otro lado, debemos anotar que las variaciones de un grupo nominal con un mismo núcleo; por ejemplo, *pimiento rojo*, *pimiento verde*, suelen ser realiza-

ciones de un mismo informante, con lo que su eliminación afectaría directamente a los datos cuantitativos. En último lugar, esto nos alejaría de la propuesta de otros trabajos de léxico disponible.

Decidimos, por tanto, señalar de manera explícita grupos nominales admitidos por el DRAE (*traje de chaqueta*) o por los diccionarios de uso consultados (*medio de transporte*, DUE); también reflejamos grupos nominales correctos en español que estén recogidos en el CREA.

Somos conscientes de la dificultad que esto conlleva y las limitaciones propias de los textos incluidos en el CREA; sin embargo, consideramos importante su aparición en la base de datos para incluirlos en el estudio en aras de utilizar un baremo externo de validez comprobada y no solo la propia competencia lingüística e intuición del investigador.

El núcleo del sintagma nominal es el elemento esencial que este tipo de estudios se propone recoger; por esta razón, decimos regularizar los siguientes elementos:

Se eliminan los **artículos** pese a que muchos de los encuestados, al igual que en Samper Hernández (2002), los hayan incluido sistemáticamente ([la boca], *boca*; [el armario] *armario*; [la escuela] *escuela*).

Adjetivos determinativos ya sean demostrativos ([este pantalón] *pantalón*), posesivos ([mi primo], *primo*), numerales ([un zapato], *zapato*), indefinidos ([mucha gente], *gente*) o interrogativos ([qué talla], *talla*).

Adjetivos calificativos ([iglesia bonita], *iglesia*) cuando no formen parte de un grupo léxico utilizado regularmente (CREA) ni reflejen una realidad específica (*bebida energética o alubia blanca*).

Sin embargo, anotaremos aquellos vocablos incluidos entre los anteriores que aparezcan como unidad léxica independiente. No encontraremos artículos, pero sí adjetivos calificativos de forma independiente (*caro*), indefinidos (*mucho*), o interrogativos (*dónde*). Del mismo modo, también daremos cuenta de las combinaciones más relevantes en la edición específica de cada centro de interés.

Tratamiento de los sintagmas verbales

Más complejo es sin duda el problema de los sintagmas verbales. Algunos informantes utilizan oraciones

para responder a los centros de interés: [antes de ir tengo que pedir una cita]. No podemos reflejar una expresión como esta en un trabajo de léxico disponible, por lo que decidimos priorizar los sintagmas nominales siguiendo las líneas anteriormente expuestas. Así pues, en esta realización registraremos *cita*, obviando los otros sintagmas que la acompañan.

Debemos señalar que las propias reglas de edición del *Proyecto Panhispánico* invitan a regular ciertos verbos que acompañen a un sustantivo que esté directamente relacionado con la acción del verbo ([beber zumo], sería *zumo*), así como los verbos atributivos se incluyen en el propio atributo ([estar enfermo], *enfermo*). Recodemos que en el trabajo con este tipo concreto de informantes el investigador debe repetir varias veces el enunciado y apoyarse en una serie de verbos para explicar el proceder de la prueba *ir, tener, escribir, haber,* (este último en forma impersonal "hay que escribir"). En la explicación de cada uno de los centros se reiteran una serie de verbos para asegurar la comprensión del mismo (por ejemplo *comer* y *beber* en 'Alimentos y bebidas').

Lo anteriormente expuesto nos lleva a tomar la decisión de regularizar el núcleo verbal junto con el sintagma nominal al que acompañan. Por supuesto, la combinación de los verbos y complementos puede resultar especialmente relevante, por ello, la señalaremos en el apartado específico de cada centro de interés.

Por último, es preciso añadir que, para que los datos cuantitativos no se vean afectados y poder reflejar con la mayor precisión posible la disponibilidad léxica del grupo, mantendremos verbos, adverbios y pronombres que aparezcan como elementos independientes en las pruebas. En compuestos de pronombre interrogativo más verbo, se regularizará el verbo en infinitivo. Por ejemplo, la aparición de [cuánto vale] será regularizada por *valer*, mientras que si encontramos [cuánto vale el pantalón] lo recogeremos como *pantalón*; estos casos responden a realidades particulares dentro de los datos que, si bien son las que más dificultan la edición, son un pequeño porcentaje del total y su aparición se ve limitada a determinados centros de interés.

4.4.1.2 Otros elementos regularizados

Uso de Mayúsculas. La regularización incluye las palabras o realidades léxicas que, por norma, deben tener en mayúscula la primera letra de cada vocablo (*Guardia Civil*, *Seguridad Social*). Lo mismo ocurre con los nombres de ciudades (*Valladolid*) o de otras entidades geográficas. Podemos también encontrar siglas referidas a conceptos de uso común en castellano (*DNI*, *TDT*)[16] o marcas comerciales (BMW*). Hemos optado por mantener las realizaciones y utilizar mayúsculas como mejor forma de hacer referencia a esta parte del léxico disponible.

Otros acortamientos que hemos decidido mantener son de las expresiones *C/* o *nº*.

Aparición de tallas, medidas y números. Por último, existen otros problemas que, debido a la peculiaridad de los centros, aún no han sido contemplados. Es frecuente la aparición de tallas (*xl*), medidas (*kg*, *litro*) y números ([42], [46]), especialmente en el centro 07 'Ir de compras'. Desde el criterio de reflejar el mayor número de realidades posibles, mantenemos las tallas para poder observar sus resultados. También optamos por reflejar las diferentes medidas tanto en la forma completa (*litro*) como en su abreviatura (*l*). Los números no entran dentro de los parámetros del léxico disponible; sin embargo, decidimos incluir *112* en 05 'La salud' por considerarlo un concepto específico.

Aparición de nombres propios. En los centros 10 'Profesiones' y 11 'La familia', algunos informantes deciden contestar con nombres propios de persona españoles o extranjeros. Al igual que en el trabajo de Valladolid, se decide regularizar este hecho mediante el término *nombres propios* para reflejar este comportamiento en la prueba y afectar lo menos posible a los datos cuantitativos; por supuesto, si un informante reitera la escritura de varios *nombres propios*, esta forma solo aparecerá una vez.

4.4.2 Criterios propios de la investigación en español como lengua extranjera

4.4.2.1 Principales criterios

Préstamos e interferencias con otras lenguas.

Sustentamos el criterio de mantener las realizaciones que aparecen en el DRAE de la forma en que este nos sugiere (por ejemplo, *slip*). Las palabras que son obtenidas de otras lenguas han sido eliminadas ([tram], [cake]).

Aun así, anotamos algunos vocablos que, si bien no son contemplados por el DRAE, aparecen en otros diccionarios de uso (*go-go*, *fitness*). Aunque no aparezca reflejado en el artículo, será conveniente señalar los casos más relevantes en los criterios específicos por centro de interés.

La verdadera encrucijada surge cuando las interferencias entre lenguas no son absolutas y no sabemos si estamos ante palabras extranjeras o errores ortográficos que hacen coincidir una palabra de otra lengua con otra de pronunciación parecida en castellano. Un claro ejemplo es [douleur] (centro 05 'La salud'), cuya pronunciación en lengua francesa[17] es similar a su pronunciación en castellano, con lo que puede llevar a confusión. Suponemos que, en un intento de escribir el mayor número de palabras posible y de asemejar la escritura a la idea que el propio informante tiene de la pronunciación de la palabra, surgen estas realizaciones que no siempre afectan a toda la expresión: [jaqueta] *chaqueta*, [administrateur] *administrador*, [livre familia] *libro de familia*, [doulour cabeza] *dolor de cabeza*. Decidimos contemplar estos casos dentro de la prueba castellanizándolos; consideramos que este léxico sí que está disponible para el hablante, aunque su competencia escrita no esté plenamente desarrollada.

Presentadores y modificadores

Ya hemos hablado de los criterios utilizados referentes a los sintagmas; sin embargo, resulta muy interesante observar cómo el modificador llega a formar parte del núcleo sintagmático, por ejemplo a través de la falsa separación del artículo: [lasperna] *pierna*, [lamario] [elmario] *armario*, [lafombra] *alfombra*, [lobra] *obra*. En otras realizaciones, la aparición del artículo hace que desparezcan elementos propios del núcleo: [la ros] *arroz*, [la spalda] *espalda*.

Amplitud de relaciones asociativas

Antes de abordar los criterios específicos conviene detenerse en el aspecto de las asociaciones individuales; aunque será importante acotar los márgenes del centro de interés eliminando asociaciones excesivamente alejadas,

Somos conscientes de la importancia de este segundo eje de asociación (por ejemplo, *ropa* aparece en 06 'Transportes' seguido de *maleta*), aunque en determinados centros decidimos no anotar estas asociaciones por estar muy alejadas o explícitamente señaladas en otro ámbito (en 01 'Partes del cuerpo' no contemplamos elementos de 'La ropa'); por otra parte, podemos añadir que campos léxicos especialmente abiertos (03 'La ciudad', 07 'Ir de compras') serán muy prolijos en este fenómeno. Las asociaciones semánticas contextualizadas son fundamentales para la interpretación correcta de los datos (*cabra* y *cobra* son confundidas por los informantes en el centro 12 'Los animales', pero al contemplar las palabras en su contexto podemos seleccionar una u otra).

Nuestra intención es mantener el mayor número posible de términos dentro del diccionario de léxico disponible; aun así, hemos encontrado ciertos vocablos especialmente alejados de los campos léxicos específicos (por ejemplo, *plástico* en 05 La salud). Estos elementos deben señalarse en el apartado 'criterios específicos por centro de interés'.

4.4.2.2 Errores

Desde nuestro punto de vista, la mayor parte de los errores responden directamente a los intentos de transmisión gráfica de una palabra conocida oralmente de la cual no se tienen nociones de escritura. El verdadero problema de la gran cantidad de errores reside en que el investigador debe hacer una labor de búsqueda e interpretación para poder extrapolar los datos; los propios términos nos ofrecen ciertos criterios comparativos más allá de la interpretación individual:

Tipo de centro en el que aparezca la palabra. En el centro 01 'Partes del cuerpo' encontramos la variante

[hombre] para lo que deducimos que es *hombro*, debido a la confusión entre vocales.

Conjunto de palabras asociadas. En el centro 12 'Animales', la realización [obeja], puede contemplarse como *abeja* u *oveja*; según las palabras que lo acompañen se interpretará como una o como otra; si aparece en enumeraciones con otros animales cuadrúpedos de granja, por ejemplo, concluiremos que es *oveja*.

Errores repetidos en un mismo informante que pueden deberse a interferencias con otra lengua conocida, o confusiones en la interpretación de la grafía; la encuesta de variables y la comparación con otras respuestas en diferentes centros de interés han resultado muy útiles en este aspecto. En el centro 01 'Partes del cuerpo' aparece la realización [cullo] que puede referirse a *culo* o a *cuello*; observando otras respuestas del informante encontramos [pello] refiriéndose a *pelo*, por ello, nos inclinamos por la regularización *culo*.

4.4.3 Criterios específicos por centros de interés: 'Partes del cuerpo'

Todos los estudios de disponibilidad léxica plantean criterios específicos relacionados con los centros de interés utilizados y su tipo de informantes. En este apartado recogemos las realizaciones particulares y explicamos las decisiones tomadas al respecto. A continuación exponemos un ejemplo tomado del centro 01 Partes del cuerpo:

No incluimos verbos de apoyo que utilizamos durante la explicación del centro ([ir]). Tampoco tendremos en cuenta expresiones especialmente alejadas del centro o que pueden haberse dado por confusión a la hora de rellenar la prueba ([falda], [camiseta], [ropa], [peine]).

Por otro lado, registramos los verbos *escuchar* y *tumbar* y los adjetivos *izquierdo*, *derecho*, *gordo* y *delgado* o el adverbio *dentro*. Pese a que uno de los criterios expuestos hace referencia a no incluir palabras que aparezcan en el nombre del centro, se anota *cuerpo*, por considerarse como un parte misma del cuerpo humano: "3. m. Tronco del cuerpo, a diferencia de la cabeza y las extremidades" (DRAE).

Algunos de los grupos léxicos más relevantes que se han mantenido debido a su presencia en el CREA son *planta del pie*, recogida también en otros trabajos de léxico disponible al igual que *dedo de la mano* o *dedo del pie*.

Decidimos regularizar el término *columna (vertebral)*, debido a la concreción del centro y a que aparece como un compuesto específico: "1. f. Eje del neuroesqueleto de los animales vertebrados, situado a lo largo de la línea media dorsal del cuerpo y formado por una serie de huesos cortos o vértebras, dispuestos en fila y articulados entre sí" (DRAE).

La expresión [organe genitale], fruto de la confusión entre las expresiones *órganos sexuales* y *genitales*, se regulariza como *genitales* por ser la más cercana a la propuesta del informante.

Es interesante la aparición de palabras totalmente alejadas del centro que, intuimos, se refieren a otras realidades que sí tienen cabida; tomamos la decisión de regularizar [oveja] a *oreja* debido a su similitud ortográfica y fonética; aunque entendemos que puede deberse a un fallo en el concepto léxico, no podemos asegurarlo, por lo que seguimos los principios de regulación. Lo mismo ocurre con el par [cadena] *cadera*. De manera similar sucede en [hombre] *hombro*; en este caso, la secuencia de vocablos en los que aparece apoya esta interpretación; podemos verlo en las series *codo*, *hombro* en uno de los informantes y *brazo*, *hombro* en otro.

5. Análisis

5.1 Programa informático

Lo ideal a la hora de plantear un estudio estadístico del léxico es poder colaborar con un matemático experto que ayude al lingüista a elaborar sus conclusiones. Bien es cierto que hay herramientas preparadas para que podamos llevar a cabo este trabajo nosotros mismos, para ello podemos usar programas de tratamiento estadístico (como el *IBM SPSS*) o la aplicación específica del *Proyecto Panhispánico* facilitada por Bartol y Hernández. Esta herra-

mienta, conocida como *Dispolex*[18], tiene su base en López Chávez y Strassburguer Frías, donde se concreta la fórmula del índice de disponibilidad. A continuación citamos la fórmula, y realizamos una escueta explicación de sus elementos. Para cualquier referencia más exhaustiva habrá que dirigirse al apartado teórico anteriormente citado.

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

D(Pj) = La disponibilidad de la palabra *j*
n = Máxima posición alcanzada en el centro de interés en la encuesta.
i = Número de posición de que se trata.
j = Índice de la palabra tratada.
e = Número natural *e* (2.718281828459045).
fji = Frecuencia absoluta de la palabra *j* en la posición *i*.
I1 = Número de informantes que participaron en la encuesta.

Si decidimos trabajar con este programa debemos tener en cuenta que todo cambio que realicemos respecto al *Proyecto Panhispánico* implicará un mayor trabajo en el volcado de datos y una menor base de comparación; sin embargo, dado que los estudios específicos requieren parámetros diferentes, será interesante realizar una revisión de centros y variables como la planteada en el capítulo anterior.

Es recomendable informatizar todas las realizaciones de los encuestados para poder tener acceso informático a sus propuestas y facilitar la tarea de edición; de esta manera se puede mantener una coherencia en los criterios de edición. Para esta fase la investigación de inmigrantes en Castilla y León utilizó el programa *Microsoft Excel* y un sistema de filtros para revisar los criterios de edición cuando fue necesario.

A continuación mostraremos algunos de los términos que, gracias a *Dispolex*, utilizaremos posteriormente para realizar los diferentes análisis y el *Diccionario general de léxico disponible en inmigrantes de Castilla y León*. Parte

de esta terminología es específica de los trabajos del *Panhispánico*, aunque otros pertenecen a la matemática estadística:

Aumento de vocablos o palabras: en las diferentes comparaciones haremos referencia a las diferencias halladas entre los estudios examinados mediante la resta de elementos y resolviendo el porcentaje de aumento o decrecimiento.

Disponibilidad: resultado de aplicar la fórmula de López Chávez y Strassburguer Frías a los datos del estudio.

Frecuencia relativa: para comprenderla hemos de definir primero frecuencia absoluta. La frecuencia absoluta de una variable estadística es el número de veces que aparece en la muestra dicho valor de la variable (un vocablo en este caso). La frecuencia absoluta es una medida que está influida por el tamaño de la muestra; al aumentarla, aumentará también el tamaño de la frecuencia absoluta. Esto hace que no sea una medida útil para la comparación y necesitemos introducir el concepto de frecuencia relativa, que es el cociente entre la frecuencia absoluta y el tamaño de la muestra.

Frecuencia acumulada: es la suma de las frecuencias relativas. Si *mano* (en primera posición del centro 01. 'Partes del cuerpo') tiene una frecuencia de 8.192 %, su frecuencia acumulada será también 8.192 %. Del mismo modo, *cabeza* (en segunda posición de este centro) tiene una frecuencia relativa de %, lo que indica que la frecuencia acumulada será 15.051 %. La siguiente frecuencia relativa se sumaría a 15.051 % y así sucesivamente.

Índice de cohesión: se halla dividiendo el promedio de respuestas por informante entre el número de palabras diferentes. Señala el carácter cerrado o abierto de los centros de interés. Un índice de cohesión alto reflejará que muchos de los vocablos son compartidos por un mayor número de informantes. Un índice de cohesión bajo mostrará que una gran parte de los vocablos son diferentes.

Palabras totales: suma de todas las palabras respondidas por centro de interés.

Palabras diferentes: es lo que se conoce como vocablo en los estudios de disponibilidad. Recoge el número total de términos por centro de interés.

Palabras por informante: promedio de respuestas por informante. Parámetro muy importante a la hora de analizar, entre otras cosas, la productividad de los centros. Se halla mediante la división del número total de palabras en un centro entre el número de informantes. En este trabajo utilizaremos también la expresión sinónima promedio de palabras.

Percentil: "1. m. *Mat.* Valor que divide un conjunto ordenado de datos estadísticos de forma que un porcentaje de tales datos sea inferior a dicho valor. Así, un individuo en el percentil 80 está por encima del 80% del grupo a que pertenece" (DRAE). Para estudiar la relación entre los diferentes centros de interés en cuanto a las palabras, vocablos e índice de cohesión, dividimos la muestra obtenida en segmentos que ayudan a clasificar los ámbitos en grupos: alto (mayor al percentil 66), medio (entre el percentil 33 y el 66) y bajo (menor al percentil 33).

Rango: Dentro de los estudios de léxico disponible se utiliza este término para referirse al puesto concreto dentro de una ordenación, por ejemplo de centros de interés según su número de vocablos. En nuestro análisis lo común será tener 12 rangos, puesto que contamos con 12 centros de interés. Por otro lado, rango, en sentido matemático, también hace referencia a todos los valores comprendidos entre el valor máximo y el mínimo; esta acepción será utilizada cuando abordamos el análisis cualitativo y nos referimos al conjunto de palabras con una disponibilidad mayor a 0,02.

5.2 Análisis cuantitativo

Los datos cuantitativos, palabras totales y palabras por informante, son aportados directamente por el programa *Dispolex*. Podremos describir el comportamiento de los centros de interés utilizados en relación al número de palabras, al de vocablos y al índice de cohesión. Estos datos son interesantes de cara a seleccionar unos centros de interés u otros para nuestro estudio y como elemento de comparación con otros trabajos de léxico disponible.

5.2.1 Palabras totales

Dentro de este aspecto realizamos una clasificación en tres grupos gracias a los percentiles:

Centros muy productivos. Muestran datos mayores al percentil 66 (2869,02). Destacan de manera especial aquellos que tienen un número de palabras totales superior a 3000 y presentan una media de palabras por informante mayor de 10: 02 'Alimentos y bebidas', 01 'Partes del cuerpo', 09 'La casa'. Dentro de este grupo se incluye también 12 'Animales', ligeramente inferior a los anteriores.

Centros de productividad media. Se sitúan entre el percentil 66 (2869, 02) y el 33 (2448,09), cerca de la media total de palabras por centro y su número de palabras por informante es mayor de 7,6 sin llegar a 8,8: 03 'La ciudad', 07 'Ir de compras', 04 'La ropa' y 11 'La familia'.

Centros de baja productividad. No olvidemos que estamos hablando siempre en relación a los datos de este estudio y según la perspectiva de los otros centros utilizados; aquí encontramos aquellos centros menores al percentil 33 (2448,09): 10 'Profesiones', 05 'La salud', 06 'Transportes'. Destaca dentro de este grupo el centro 08 'Administración pública' de muy baja productividad, con un número de palabras por informante considerablemente menor al resto (5,49).

Rango	Centro de interés	Palabras totales	Palabras por informante
2	01. Partes del cuerpo	3544	10,87
1	02. Alimentos y bebidas	3566	10,94
5	03. La ciudad	2849	8,74
7	04. La ropa	2730	8,37
10	05. La salud	2269	6,96
11	06. Transportes	2213	6,79
6	07. Ir de compras	2816	8,64
12	08. Administración pública	1792	5,49
3	09. La casa	3481	10,68
9	10. Profesiones	2358	7,23
8	11. La familia	2501	7,67
4	12. Animales	2926	8,98

Tabla 5: Resultados totales por centro de interés de palabras y palabras por informante

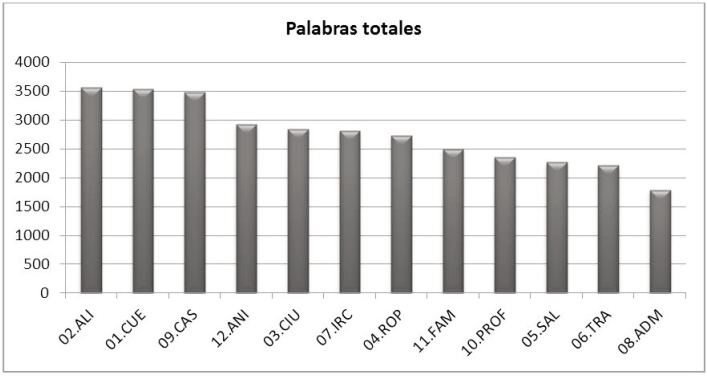


Gráfico 8: Palabras totales ordenadas por rango

5.2.2 Vocablos

Rango	Centro de interés	Vocablos
11	01. Partes del cuerpo	126
6	02. Alimentos y bebidas	308
2	03. La ciudad	382
9	04. La ropa	141
4	05. La salud	358
8	06. Transportes	141
1	07. Ir de compras	497
5	08. Administración pública	322
7	09. La casa	257
3	10. Profesiones	361
12	11. La familia	81
10	12. Animales	139

Tabla 6: Vocablos totales y rango por centro de interés

El gráfico 9 muestra, de mayor a menor, el número de vocablos aportados por cada uno de los centros de interés. De manera similar al trabajo realizado en el apartado anterior, podemos clasificar los centros de interés en cuatro grupos diferentes según su riqueza léxica. Comenzaremos señalando que la media de vocablos es de 259, 42.

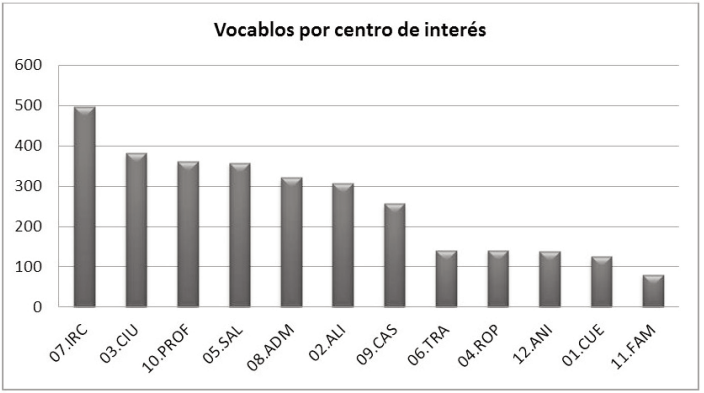


Gráfico 9: Vocablos totales ordenados por rango

En esta ocasión el más destacado es 07 'Ir de compras' que se encuentra en la posición de **mayor riqueza léxica** y constituye, además, el elemento más alejado de la media (237, 58). Los casi 500 vocablos que presenta hacen de este centro de interés el que más vocablos aporta a los datos finales. Este grupo, con unos datos mayores al percentil 66 (331,36) incluye también: 03 'La ciudad', 10 'Profesiones', y 05 'La salud'.

Después, podemos hablar de centros de **riqueza léxica media**, cuyo número de vocablos se encuentra son menores que el percentil 66 (331,36) y mayores que el 33 (141): 08 'Administración pública', 02 'Alimentos y bebidas' y 09 'La casa'.

Por último, anotamos los centros de **menor riqueza léxica**, aquellos que aportan un menor número de vocablos al diccionario. En este grupo se incluyen los centros con una cantidad igual o inferior al percentil 33 (141): 06 'Transportes', 04 'La ropa', 12 'Animales' y 01 'Partes del cuerpo'. Destaca especialmente 11 'La familia', puesto que no llega siquiera a los 100.

5.2.3 Índice de cohesión y densidad léxica

Ambos conceptos están íntimamente relacionados y reflejan el carácter de los diferentes centros. Mientras que el índice de cohesión se obtiene, como ya hemos señalado, de dividir la media de palabras por informante entre el número de vocablos obtenidos, la densidad léxica es el

resultado de dividir el total de las palabras entre el total de vocablos, lo que nos da la cantidad de repeticiones de un vocablo en la encuesta. El rango de ambas será, por tanto, coincidente.

Rango	Centro de interés	Índice de cohesión	Densidad léxica
2	01. Partes del cuerpo	0,086	28,13
7	02. Alimentos y bebidas	0,036	11,58
8	03. La ciudad	0,023	7,46
4	04. La ropa	0,059	19,36
10	05. La salud	0,019	6,34
5	06. Transportes	0,048	15,7
11	07. Ir de compras	0,017	5,67
12	08. Administración pública	0,017[19]	5,57
6	09. La casa	0,042	13,54
9	10. Profesiones	0,02	6,53
1	11. La familia	0,095	30,88
3	12. Animales	0,065	21,05

Tabla 7: Datos totales de índice de cohesión y densidad léxica por centro de interés

Del mismo modo, las apreciaciones realizadas sobre los centros más o menos cohesionados serán igualmente aplicables desde el punto de vista de la densidad léxica. El índice de cohesión medio de esta investigación es de 0,044 (un 10,62 de densidad léxica). Podemos dividir los centros en varios grupos:

En primer lugar, aquellos con un **índice de cohesión alto**, mayores al percentil 66 (0,051); aquí destacan 11 'La familia' y 01 'Partes del cuerpo' aunque también encontramos 12 'Animales' y 04 'La ropa'. Dichos centros se sitúan también por encima del percentil 66 de densidad léxica (16,652).

El segundo grupo lo constituyen aquellos cuyo **índice de cohesión** se aproxima más a la **media** total; hablamos de aquellos que se sitúan entre el percentil 66 (0,051) y el 33 (0,022): 06 'transportes', 09 'La casa' y 02 'Alimentos y bebidas' y 03, 'La ciudad'. Dichos centros se encuentran también entre los percentiles 66 (16,652) y 33 (7,116) correspondientes a la densidad léxica.

Por último, señalamos los centros con un **bajo índice de cohesión**, menor al percentil 33 (0,022) y una densidad

léxica más baja (menor de 7,116). Estos campos son, 10 'Profesiones', 05 'La salud', 07 'Ir de compras' y 08 'Administración pública'.

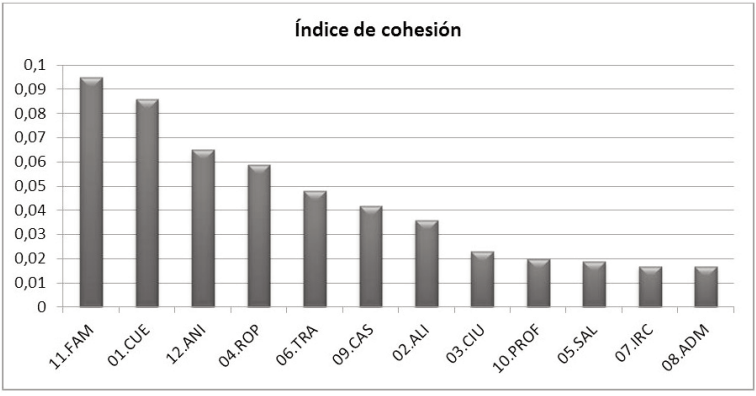


Gráfico 10: Índices de cohesión por centro de interés ordenados por rangos

5.3 Análisis por variables

Analizar las variables implica observar cómo se comportan las medias generales de las distintas variables en cuanto a palabras por informante, vocablos e índice de cohesión. Desde nuestro punto de vista, resulta particularmente útil, en especial si no se manejan conocimientos específicos de estadística matemática, plantear el análisis como una observación general de los datos que deberá ser contrastada con fórmulas específicas. Si decidimos mantener esta opción en nuestro trabajo, las aportaciones que realizamos serán subjetivas, por lo que nuestra labor científica debe centrarse fundamentalmente en la descripción.

De esta manera, anotaremos los resultados de las diferentes opciones y las relacionaremos con lo esperado teniendo en cuenta el número de encuestados en cada una de ellas; es decir, a mayor número de informantes, mayor número de vocablos y menor índice de cohesión, y a menor número de informantes, menor número de vocablos y mayor índice de cohesión; por tanto, realizaremos el análisis en relación a los datos más significativos; centraremos el desarrollo específico en el número de palabras por informante, aunque en determinadas varia-

bles podamos detenernos a realizar otro tipo de comparaciones que nos resulten útiles. A la hora de analizar los datos, decidimos obviar la opción 'no sabe o no contesta' (NS/NC).

5.3.1 Ejemplo de análisis: variable sexo

Lo primero que debemos recordar es que el número de hombres (151) y de mujeres (169) está bastante equilibrado en nuestro estudio. Las medias aquí expuestas muestran un mayor número de palabras por informante, un mayor número de vocablos y un índice de cohesión más alto para las mujeres. Debido a la igualdad en los encuestados, resulta interesante observar el desarrollo en los diferentes puntos.

	Palabras por informante		Vocablos		Índice de cohesión	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
01.CUE	10,09	11,68	95	95	0,106	0,123
02.ALI	9,35	12,52	195	254	0,048	0,049
03.CIU	7,98	9,54	241	289	0,033	0,033
04.ROP	7,14	9,54	92	118	0,078	0,081
05.SAL	6,04	7,83	212	266	0,028	0,029
06.TRA	6,74	6,89	105	91	0,064	0,076
07.IRC	7,89	9,39	311	374	0,025	0,025
08.ADM	5,18	5,84	207	221	0,025	0,026
09.CAS	9,31	12,01	170	205	0,055	0,059
10.PROF	6,94	7,59	228	249	0,03	0,03
11.FAM	7,05	8,33	54	62	0,131	0,134
12.ANI	8,26	9,67	95	117	0,087	0,083
MEDIA	7,66	9,24	167,08	195,08	0,059	0,062

Tabla 8: Variable sexo, datos por centro de interés y medias totales

En primer lugar hablaremos del número de palabras por informante cuya media general en la investigación es de 8,45. Las mujeres se sitúan por encima de la media, mientras que los hombres lo hacen por debajo. Si examinamos cada uno de los centros, encontramos que, en todos ellos, las mujeres obtienen datos mayores que los varones; destacan especialmente 02 'Alimentos y bebidas', donde el número de palabras es superior en 3,17, y 09 'La casa', con una diferencia de 2,7. También hay otros centros donde la diferencia es mínima: 08 'Administración pública' y 10

'Profesiones', con 0,66 y 0,65 respectivamente, o prácticamente inapreciable, 0,15 en 06 'Transportes'.

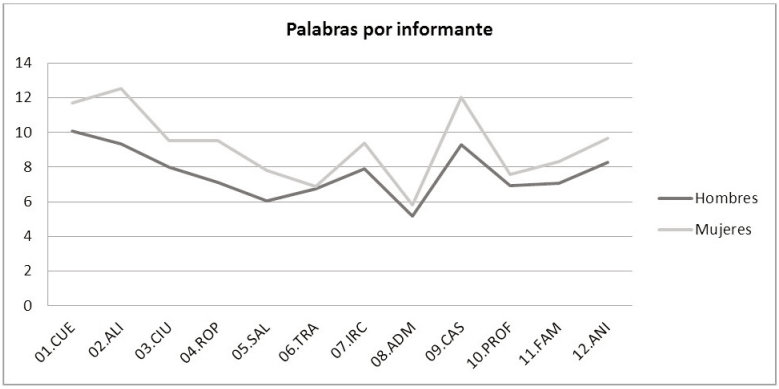


Gráfico 11: Variable sexo, palabras por informante

No es de extrañar que el número de vocablos sea mayor para las mujeres, puesto que el total de informantes en esta opción también lo es; sin embargo, algunos de los centros de interés destacan por su preponderancia de vocablos por parte de las informantes, es el caso de 07 'Ir de compras' (63 vocablos más) o 02 'Alimentos y bebidas' (59). Por otro lado, 06 'Transportes' muestra 14 vocablos más para los hombres. Destaca también el hecho de que en 06 'Partes del cuerpo' ambos grupos coincidan en el número de vocablos.

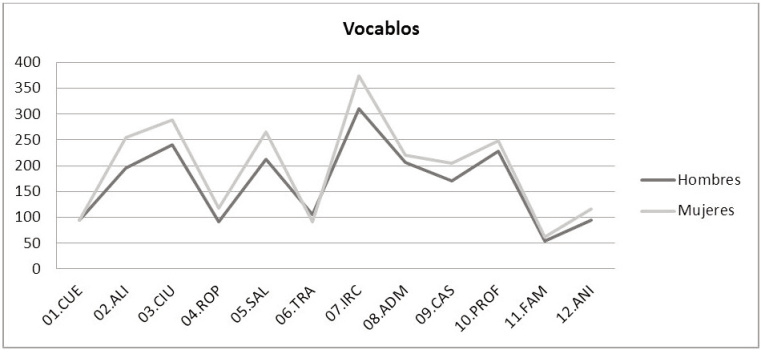


Gráfico 12: Variable sexo, vocablos totales

En lo referente al índice de cohesión, los datos son parejos en la mayoría de los centros; encontramos el mismo índice en 01 'La ciudad', 07 'Ir de compras' y 10 'Profesiones'. Las mayores diferencias aparecen en favor de las mujeres en los centros 01 'Partes del cuerpo' (0,017) y 06 'Transportes' (0,012). El único centro donde el índice de cohesión es superior en varones es 12 'Animales', con una diferencia de 0,004.

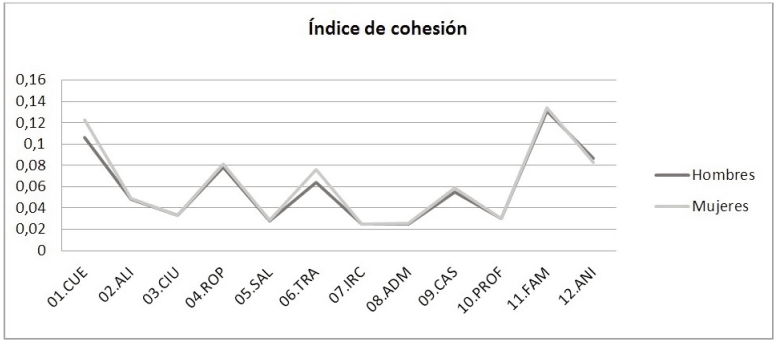


Gráfico 13: Variable sexo, índice de cohesión

En esta ocasión sí marcamos, como ya hemos dicho, una ligera superioridad en cifras por parte de las mujeres aunque no parece determinante. Según la observación, podemos afirmar que los datos donde las mujeres obtienen un mayor número de palabras y vocablos en relación a los hombres se obtienen en el centro 02 'Alimentos y bebidas'; por su parte, los varones, muestran sus mejores cifras en el centro 06 'Transportes'; sin embargo, lo más relevante en ambas opciones es la similitud en los datos obtenidos.

5.4 Análisis cualitativo

Puede resultar inviable plantear un análisis de todas las palabras recogidas en nuestro diccionario, por esta razón planteamos una selección de los vocablos más relevantes. El propio programa *Dispolex* propone tres maneras, las cuales son utilizadas para la comparación de conjuntos. Por un lado, encontramos la selección de las 100 palabras más disponibles de cada uno de los centros; por

otro, nos presentan el uso de vocablos cuya frecuencia acumulada sea menor del 80%, lo que aseguraría anotar las realizaciones más presentes. Mientras que la primera opción no tiene en cuenta la disponibilidad relativa de los vocablos, con lo que no es selectiva, la segunda se ve mediatizada por el porcentaje de aparición y deja fuera o dentro del rango determinados elementos dependiendo de la presencia de los anteriores. Decidimos por tanto utilizar la tercera vía que es la selección por disponibilidad. *Dispolex* propone tener en cuenta aquellos vocablos cuya disponibilidad sea mayor a 0,02. De cara a observar el léxico más disponible de los inmigrantes adultos de Castilla y León y proponer un diccionario con dichos datos, resulta mucho más interesante asumir la propuesta de *Dispolex* que nos dará un total de 509 palabras.

Centro de interés	Vocablos con disponibilidad mayor 0,02	Porcentaje de vocablos en relación al total
01. Partes del cuerpo	37	29,37%
02. Alimentos y bebidas	60	19,48%
03. La ciudad	67	17,54%
04. La ropa	33	23,40%
05. La salud	40	11,17%
06. Transportes	18	12,77%
07. Ir de compras	63	12,68%
08. Administración pública	38	11,80%
09. La casa	50	19,46%
10. Profesiones	42	11,63%
11. La familia	25	30,86%
12. Animales	36	25,90%

Tabla 9: Vocablos con disponibilidad mayor a 0,02 y su porcentaje respecto al total

5.4.1 Estudio de los centros comunes al proyecto panhispánico

Ya hemos señalado anteriormente que esta investigación comparte determinados centros con el *Proyecto Panhispánico*. Recordamos que dichos campos léxicos son: 01 'Partes del cuerpo', 02 'Alimentos y bebidas', 03 'La ciudad', 04 'La ropa', 06 'Transportes', 10 'Profesiones', 12 'Animales', además de contar con un centro que agrupa dos de las propuestas del *Proyecto*, 09 'La casa' correspon-

diente a 'Partes de la casa (sin muebles)' y 'Los muebles de la casa'.

Para el estudio cualitativo resulta muy útil la comparación con trabajos de léxico disponible que tengan elementos en común con el planteamiento base de nuestro estudio. En nuestro caso utilizaremos la tesis de Raquel Cruz Alonso, realizada en 550 escolares de último curso de Bachillerato en Castilla y León (similitud geográfica) y la investigación que Sánchez-Saus realiza con 322 informantes extranjeros procedentes de las universidades andaluzas (informantes extranjeros). Gracias a esta parte del estudio podremos observar si el cambio a la hora de realizar el diccionario de inmigrantes adultos supone también un cambio sustancial en las prioridades léxicas de los centros compartidos.

Decidimos elaborar el análisis a partir de los vocablos más disponibles (más de 0,02) de nuestro estudio y observar en qué posición quedan en ambas investigaciones y qué vocablos dentro de las 20 primeros expuestos en Sánchez-Saus y Cruz Alonso no aparecen dentro del rango marcado en nuestra propuesta. En aquellas palabras que no coincidan los criterios de edición se mantendrá la palabra más disponible; por ejemplo, *(carne de) cerdo* aparece como *cerdo* (0.01703) y *carne de cerdo* (0.00681) en Cruz Alonso, luego mantendremos *cerdo*. Iremos señalando las peculiaridades en cada uno de los centros y las variaciones en la nomenclatura de los centros de interés que afecten al estudio; es el caso de 'Alimentos y bebidas' que se enuncia como 'Comidas y bebidas' en Cruz Alonso.

En líneas generales, podemos decir que, normalmente, la disponibilidad de los vocablos suele ser menor en nuestro estudio tanto en el orden (nuestro primer vocablo será menos disponible que la primera palabra de Cruz Alonso, aunque no siempre en Sánchez-Saus), como en los vocablos específicos; es decir, las ratios de disponibilidad en los vocablos más disponibles en nuestra investigación son menores si los comparamos con el mismo vocablo en las otras investigaciones independientemente del rango que ocupen (*lengua* en 'Partes del cuerpo' ocupa en nuestro estudio la posición 19 con una disponibilidad de 0.09589, en Cruz Alonso y Sánchez-Saus ocupa el puesto 23 con una disponibilidad de 0.20998 y 0.11766 respectivamente). La especificidad de nuestro estudio radica en que encontramos determinadas

palabras con una disponibilidad léxica mayor en nuestra propuesta, incluso vocablos presentes en nuestra muestra que no tienen una correspondencia en los otros dos trabajos. Cuando la disponibilidad léxica de una palabra en inmigrantes es mayor que dicha palabra en nativos o en universitarios extranjeros, también supone un cambio notable en cuanto al puesto que ocupa en el diccionario. Los grandes cambios en el orden de algunas palabras fijan las prioridades léxicas del grupo de informantes. Resaltaremos especialmente aquellos vocablos que en nuestro estudio superen en disponibilidad léxica a las otras dos investigaciones.

5.4.1.1 Ejemplo de análisis: 'Partes de cuerpo'

Nº, Palabra y Disponibilidad (Fernández-Merino *Castilla y León*)			Nº y Disponibilidad (Cruz)		Nº y Disponibilidad (Sánchez-Saus)	
1	mano	0.73295	4	0.81760	2	0.65177
2	cabeza	0.66056	3	0.81930	3	0.61652
3	ojo	0.59085	2	0.82090	1	0.67221
4	nariz	0.39458	8	0.72136	5	0.50393
5	dedo	0.35305	7	0.73579	10	0.41528
6	boca	0.33455	10	0.58168	8	0.43203
7	pierna	0.32885	5	0.79496	6	0.49598
8	pelo	0.31758	15	0.33617	7	0.44966
9	oreja	0.30605	9	0.67462	11	0.36911
10	pie	0.26193	6	0.75586	4	0.51348
11	brazo	0.20618	1	0.84818	9	0.42036
12	espalda	0.17229	20	0.27733	12	0.24091
13	cara	0.17085	30	0.14416	17	0.19720
14	cuello	0.14708	12	0.45411	21	0.15171
15	diente	0.14237	21	0.27296	15	0.20015
16	uña	0.13732	13	0.41842	20	0.16709
17	rodilla	0.13263	11	0.49618	13	0.22676
18	pecho	0.12290	19	0.28130	22	0.13507
19	lengua	0.09589	23	0.20998	23	0.11766
20	estómago	0.09159	28	0.16943	14	0.20840
21	piel	0.08971	77	0.02699	19	0.16982
22	corazón	0.08858	26	0.17465	18	0.18119
23	muñeca	0.08660	18	0.31888	28	0.06561
24	tripa	0.07331	96	0.01729	111	0.00410
25	labio	0.06806	24	0.17958	16	0.19869
26	codo	0.06614	14	0.40565	29	0.06509
27	hombro	0.06025	16	0.32643	24	0.11181
28	frente	0.05856	43	0.07137	32	0.05048
29	culo	0.05658	31	0.14340	25	0.10414
30	barriga	0.04840	54	0.04663	27	0.07141
31	ceja	0.04619	27	0.17108	30	0.05918
32	cuerpo	0.03370	Eliminado por criterios de edición		Eliminado por criterios de edición	

33	garganta	0.03188	81	0.02375	26	0.09551
34	hígado	0.03071	36	0.11618	42	0.03006
35	pulmón	0.02960	33	0.13833	35	0.04652
36	muslo	0.02663	34	0.13526	49	0.01995
37	tobillo	0.02423	17	0.31924	31	0.05084

Tabla 10: Comparación cualitativa de 'Partes del cuerpo' entre Fernández-Merino (Castilla y León) con inmigrantes, Cruz con nativos en Castilla y León y Sánchez-Saus con extranjeros en Andalucía

El trabajo de Cruz Alonso muestra 86 vocablos con disponibilidad superior a 0,02; un número considerablemente mayor de en comparación a nuestro trabajo (37) o al de Sánchez-Saus (48), lo que concuerda con la afirmación expuesta de un número menor de vocablos para nuestros informantes.

En lo que concierne a la disponibilidad léxica de los diferentes términos, nuestro estudio muestra unos datos generalmente menores en comparación con los estudiantes de Castilla y León; tan solo determinadas palabras muestran una disponibilidad mayor en nuestra propuesta: *cara*, *piel*, *garganta* y *barriga*. Además de estas, destacan *cuerpo*, que no aparece en Cruz Alonso por criterios de edición, y *tripa* que quedaría fuera del rango de 0,02 en la investigación con nativos.

En Andalucía destacamos una mayor similitud en cuanto a los datos de disponibilidad, pues van a estar mucho más cercanos a los de nuestro estudio; pese a ello, aunque la disponibilidad de nuestra primera palabra (*mano* 0.73295) sea mayor (*ojo* 0.67221), lo cierto es que los inmigrantes adultos mostrarán más disponibilidad tan solo en algunas palabras específicas (*mano*, *cabeza*, *muñeca*, *hígado* y *codo*). Es conveniente señalar *cuerpo* que, al igual que en Cruz Alonso, se elimina por criterios de edición; además, los términos *tripa* y *muslo*, quedan fuera del rango de 0,02 en Sánchez-Saus.

Las 20 palabras más disponibles en uno y otro estudio se encuentran presentes en los vocablos más relevantes de nuestro trabajo.

En relación a la especificidad del léxico encontramos dos vocablos con una disponibilidad más alta en nuestro estudio. En primer lugar, tenemos *cuerpo*, cuya preponderancia responde a criterios de edición; sin embargo, *tripa*

se resalta como un vocablo relevante para nuestros informantes en relación a los de los otros dos estudios.

5.4.2 Centros específicos de este trabajo

Nuestra investigación posee, como hemos visto, centros de interés propios que merecen ser planteados desde un punto de vista cualitativo. Es un buen momento para señalar los tipos de palabras que aparecen, los vocablos referentes a otros centros…

5.4.2.1 Ejemplo de análisis: 'La salud'

Nº	Palabra	Disponibilidad	Nº	Palabra	Disponibilidad
1	médico	0.52699	21	tripa	0.04382
2	hospital	0.40226	22	paracetamol	0.04379
3	pastilla	0.33677	23	enfermedad	0.04076
4	farmacia	0.16925	24	garganta	0.03942
5	cabeza	0.15890	25	urgencias	0.03510
6	enfermero	0.14002	26	catarro	0.03489
7	jarabe	0.13272	27	aspirina	0.03315
8	fiebre	0.12293	28	dentista	0.02736
9	tos	0.11807	29	urgencia	0.02728
10	doctor	0.11633	30	espalda	0.02634
11	centro de salud	0.10710	31	diente	0.02585
12	medicamento	0.10263	32	receta	0.02495
13	dolor	0.10180	33	pierna	0.02426
14	enfermo	0.09834	34	mano	0.02344
15	estómago	0.09099	35	ojo	0.02330
16	gripe	0.06307	36	mal	0.02322
17	medicina	0.06029	37	antibiótico	0.02303
18	dolor de cabeza	0.05738	38	médico de cabecera	0.02157
19	cita	0.05424	39	corazón	0.02090
20	doler	0.05116	40	vomitar	0.02062

Tabla 11: Vocablos por encima de 0,02 de disponibilidad en el centro 'La salud'

Este ámbito muestra un total de 40 vocablos procedentes de otros centros de interés ya tratados, aunque aquí aparecen con mayor disponibilidad: 'Profesiones' (*médico* 0.26122, *enfermero* 0.13154…), 'La ciudad' (*hospital* 0.13662, *farmacia* 0.04303…). También aporta palabras novedosas como *pastilla* o *jarabe* correspondientes al campo léxico de los medicamentos. Es muy interesante observar la gran cantidad de compuestos que hallamos con el sustantivo dolor (*dolor*, *dolor de cabeza*) o el verbo *doler*; recordemos que las expresiones referentes a 'Partes del cuerpo' que aquí apare-

cen (*cabeza*, *estómago*, *tripa*, *garganta*, *espalda*…) se han regularizado también de esta manera cuando aparecían desempeñando la función de objeto directo de *doler*.

6. El diccionario de léxico disponible

La última parte que será, precisamente, la que sustente todo el trabajo (pues es de donde se obtienen todos los datos analizados) es el diccionario de léxico disponible. Como ya hemos comentado, los datos que aparecen aquí son obtenidos gracias a la aplicación *Dispolex*. Es recomendable, para no tener problemas con el procesador de textos que utilicemos, preparar el diccionario antes de incluirlo al trabajo definitivo, porque su densidad puede afectar al desarrollo de nuestro escrito.

Podemos ordenar y plantear el diccionario según los requerimientos de nuestro estudio. En nuestro caso, optamos por incluir solamente los centros de interés ordenados por disponibilidad, por un lado, y en orden alfabético, por otro. Si resultara interesante podríamos plantear el diccionario específico de determinadas variables, tal y como realiza Bartol en su publicación del léxico disponible de Soria, donde expone las palabras ordenadas según el sexo de los participantes.

La ubicación del diccionario en este tipo de trabajos suele darse en los anexos, detrás de la bibliografía; decidimos por tanto incluir dos ejemplos referentes al centro de interés 'Partes del cuerpo' extraídos del trabajo de Fernández-Merino (Castilla y León) en el lugar indicado[20].

7. Conclusiones

A lo largo de estas páginas hemos dibujado un pequeño itinerario metodológico que puede recorrerse sin dificultad para la elaboración de estudios de léxico disponible que difieran de las propuestas existentes y dediquen su análisis a un grupo específico de informantes.

En primer lugar, debemos conocer la base teórica de las investigaciones de léxico disponible. Esto supone trabajar

los textos de los investigadores franceses de los años 50, así como conocer el *Proyecto Panhispánico* y su recorrido.

Es conveniente seleccionar un grupo de estudio dispuesto a colaborar y unos espacios determinados para llevar a cabo la prueba de léxico disponible. Del mismo modo, resulta beneficioso plantear un estudio previo con base teórica para recoger los elementos más interesantes sobre nuestro grupo de informantes y su relación con el español.

Las variables esbozadas en la prueba de léxico disponible y una encuesta realizada a los educadores y coordinadores de los centros nos ayudarán a realizar un análisis específico de los informantes concretos que responderán a nuestra encuesta. De esta manera, podremos observar si se encuentran dentro de lo esperado, según la bibliografía consultada, o suponen una excepción.

El trabajo teórico previo a la visita a los centros colaboradores implica un conocimiento de los parámetros propuestos por el *Proyecto Panhispánico* y de la realidad concreta de nuestro grupo de estudio; gracias a ello, podremos realizar una selección de centros de interés y de variables que responda a los objetivos fijados en nuestro estudio, sea asequible para nuestros informantes y permita la comparación con otras investigaciones.

El desarrollo de la prueba supone un elemento clave para la obtención de datos y las decisiones que tomemos al respecto deben ir encaminadas a facilitar la comprensión y ejecución de la misma por parte de los informantes.

La edición de datos puede resultar compleja, en especial si trabajamos con sujetos que no dominen el español. Los diferentes estudios de disponibilidad léxica plantean una serie de aspectos que debemos tener muy en cuenta. Dicho esto, es necesario seguir los criterios del *Proyecto Panhispánico*, las anotaciones generales referidas a los estudios de español como lengua extranjera (si es el caso) y las apreciaciones específicas por centros de interés.

El análisis de datos parte de la informatización de los mismos. El programa *Dispolex* supone una herramienta de fácil acceso y uso intuitivo que nos permitirá obtener los elementos necesarios para el diccionario y su análisis. El análisis cuantitativo general y por variables nos permitirá un estudio en profundidad de los datos que puede verse

enriquecido si se enfoca este apartado desde la colaboración con expertos en otras disciplinas como la matemática estadística. Por otro lado, es interesante la observación cualitativa desde el enfoque aquí planteado (disponibilidad mayor a 0,02); el trabajo comparativo con otras investigaciones es también muy provechoso en este punto.

Gracias a los diferentes apartados proyectados en este artículo podemos realizar un estudio que acompañe a nuestro diccionario de léxico disponible, justifique nuestra propuesta y recoja los descubrimientos más importantes.

BIBLIOGRAFÍA

ANDIÓN HERRERO, Mª Antonieta. "La inmigración no hablante de español en España: situación y características". *Revista de lingüística y lenguas aplicadas* 1 (2006): 7-24. Impreso.

—. "El sexo y la edad en la población inmigrante en España. Consideraciones para E/LE". *Cuadernos Cervantes de la Lengua Española* 55 (2005): 18-25. Impreso.

BARTOL HERNÁNDEZ, José Antonio, and Natividad Hernández Muñoz. *Dispolex.*, 2003. Web.

BARTOL HERNÁNDEZ, José Antonio. *Léxico disponible de soria. Estudio y diccionarios*. Burgos: Instituto Castellano y leonés de la lengua, 2004. Impreso.

BELLÓN FERNÁNDEZ, Juan José, Marta SAMPER HERNÁNDEZ, and José Antonio SAMPER PADILLA. "El proyecto de estudio de la disponibilidad léxica en español". Ed. Gerd Wotjak. Frankfurt-Madrid: Vervuert Iberoamericana, 2003. 27-140. Impreso.

BLAS ARROYO, José Luis. *Sociolingüística del español. Desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Madrid: Cátedra, 2005. Impreso.

BORREGO NIETO, Julio. "Algunas preguntas con relación al concepto de léxico disponible". *Diccionario, léxico y cultura.* Eds. Josefina Prado Aragoneses and Mª Victoria Galloso Camacho. Huelva: Universidad de Huelva, 2004. 59-69. Impreso.

CARCEDO GONZÁLEZ, Alberto. "Desarrollo de la competencia léxica en español le: análisis de cuatro fases de disponibilidad". *Pragmalingüística* 5-6 (1999): 75-94. Impreso.

Council of Europe, Language Policy Division Strasbourg. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*. Madrid: Instituto Cervantes, Ministerio de Educación Cultura y Deporte, Anaya, 2002. Impreso.

CRUZ ALONSO, Raquel. "El léxico disponible de Castilla y León". Tesis. Universidad de Salamanca, 2011. Impreso.

ELLIS, Andrew W., Natividad Hernández Muñoz, and Cristina Izura. "Cognitive Aspects of Lexical Availability". *European Journal of Cognitive Psychology* 18 (5) (2006): 730-55. Impreso.

FERNÁNDEZ-MERINO GUTIÉRREZ, Pablo. "Diálogo sobre la realidad de la enseñanza de español a inmigrantes adultos". *Plurilingüismo y enseñanza de ELE en contextos multiculturales.* Eds. Beatriz Blecua, et al. Asociación para la Enseñanza del Español como Lengua Extranjera, 2013. 410-419. Impreso.

—. "Disponibilidad léxica de inmigrantes en Castilla y León". Tesis. Universidad de Valladolid, 2013. Impreso.

—. Disponibilidad léxica de inmigrantes en Castilla y León: Valladolid. Tesina. Valladolid: Universidad de Valladolid, 2010. Impreso.

—. *Encuestas y pruebas de disponibilidad en inmigrantes adultos*. Eds. Enrique Balmaseda and Emma Pérez. Monográfico ed. Disponibilidad léxica e inmigración Vol. Segundas lenguas e inmigración, 2014. En prensa.

Ferreira, Victor S., and Gary S. Dell. "Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production". *Cognitive Psychology* 40 (2000): 296-340. Impreso.

Galloso Camacho, Victoria. *El Léxico De Los Estudiantes Preuniversitarios En El Distrito Universitario De Salamanca*. Salamanca: Ediciones Universidad de Salamanca, 2002. Impreso.

García Parejo, Isabel. "La enseñanza del español a inmigrantes adultos". *Vademécum para la formación de profesores. enseñar español como segunda lengua (L2)/ lengua extranjera (LE)*. Eds. J. Sánchez and I. Santos. Madrid: SGEL, 2004. 1259-1277. Web.

Gómez Devís, Mª Begoña. *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Valencia: Universidad de Valencia, 2004. Impreso.

Gougenhejm, Georges, et al. *L'elaboration Du Français Élémentaire. Étude Sur l'établissement d'un Vocabulaire Et d'une Grammaire De Base*. 2ª ed. París: Didier, 1956. Impreso.

Hernández García, Mª Teresa, and Félix Villalba Martínez. "Los programas de L2 para adultos inmigrantes". *La Evaluación en la enseñanza de segundas lenguas. documentos de trabajo y conclusiones del III Encuentro de especialistas en la enseñanza de L2 a inmigrantes.* Ed. Catalina González Las. Granada: Port Royal, 2009. 27-37. Impreso.

Hernández Muñoz, Natividad. "La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en lingüística y psicología". *Estudios sobre adquisición del lenguaje IV Congreso internacional sobre la adquisición de las lenguas del estado.* Eds. M. A. Mayor, B. Zubianz, and E. Díez. Salamanca: Ediciones Universidad de Salamanca, 2005. 942-954. Impreso.

Isabel Sanz, Fuencisla. *Español como nueva lengua: orientaciones para un curso de emergencia destinado a inmigrantes adultos*. Madrid: Instituto Cervantes, 2005. Impreso.

Krashen, Stephen D. *Principles and Practice in Second Language Acquisition*. Nueva York: Prentice Hall, 1988. Impreso.

López Chávez, Juan, and Carlos Strassburguer Frías. "Un modelo para el cálculo del índice de disponibilidad léxica

individual". *La enseñanza del español como lengua materna.* Ed. Humberto López Morales. Río Piedras: Universidad de Puerto Rico, 1991. 91-112. Impreso.

López Morales, Humberto. "Los estudios de disponibilidad léxica: pasado presente". *Boletín de Filología de la Universidad de Chile (homenaje a Rodolfo Oroz)* 35 (1995): 245-59. Impreso.

—. *Léxico disponible en Puerto Rico.* Madrid: Arco/Libros, 1999. Impreso.

Mateo García, María Victoria. "Disponibilidad Léxica: Posibles Aplicaciones". *Actas de las I Jornadas sobre el estudio y la enseñanza del léxico.* Eds. Luque Durán, Juan de Dios and Antonio Pamies Bertrán. Granada: Universidad de Granada, 1994. 141-147. Impreso.

Michéa, René. "Mots Fréquents Et Mots Disponibles. Un Aspect Nouveau De La Statistique Du Langage". *Les langues modernes* 47 (1953): 338-44. Impreso.

Miquel López, L. (1995). Reflexiones previas sobres la enseñanza de ELE a inmigrantes y refugiados. *Didáctica,* 7, 241-270. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/antologia_didactica/inmigracion/miquel.htm.

Moliner, María. *Diccionario De Uso Del Español.* Madrid: Gredos, 2007. Impreso.

Moreno Fernández, Francisco. "Metodología del proyecto para el estudio sociolingüístico del español de España y América". *Lingüística* 8 (1996): 257-87. Impreso.

Parejo García, Isabel, and Théophile Ambadiang. "*Propuesta de análisis de variables que pueden influir en el proceso de enseñanza/aprendizaje del español en el colectivo de inmigrantes adultos*". *Didáctica* 7 (1995): 377-84. Impreso.

Real Academia Española. "Corpus de referencia del español actual". *Real Academia Española.* 2008.Web. 15/01/2013 <http://corpus.rae.es/creanet.html>.

—. "Diccionario de la lengua". *Real Academia Española.* 2014.Web. 15/01/2013 <http://lema.rae.es/drae/>.

—. "Diccionario panhispánico de dudas". *Real Academia Española.* 2014.Web. 15/01/2013 <http://lema.rae.es/dpd/>.

—. *Nueva gramática de la lengua española.* I: Morfología y sintaxis Vol. Madrid: Espasa-Calpe, 2009. Impreso.

Rodríguez Muñoz, Francisco José, and Ofelia Muñoz Hernández. "De la disponibilidad a la didáctica léxica". *Tejuelo: didáctica de la lengua y la literatura* 4 (2009): 8-18. Impreso.

Samper Hernández, Marta. *Disponibilidad léxica en alumnos de ELE.* Málaga: Asociación para la Enseñanza del Español como Lengua Extranjera, 2002. Impreso.

Samper Padilla, José Antonio, and Marta Samper Hernández. "Aportaciones recientes de los estudios de disponibilidad léxi-

ca". *Lynx: Panorámica de estudios lingüísticos* 5 (2006): 5-95. Impreso.

SAMPER PADILLA, José Antonio. "Criterios de edición del léxico disponible: sugerencias". *Lingüística ALFAL* 10 (1998): 311-33. Impreso.

—. "Disponibilidad Léxica y Sociolingüística". *Discurso y Sociedad: contribuciones al estudio de la lengua en contexto social.* Eds. José Luis Blas Arroyo, Manuela Casanova Ávalos, and Mónica Velando Casanova. Castellón: Publicaciones de la Universidad Jaume I, 2006. 99-120. Impreso.

SÁNCHEZ-SAUS LASERNA, Marta. "Bases Semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas". Tesis. Universidad de Cádiz, 2011. Impreso.

SCHUMANN, John H. "Social Distance as a Factor in Second Language Acquisition". *Language Learning* 26.1 (1976): 135-43. Impreso.

SECO, M., A. Olimpia, and G. Ramos. *Diccionario del español actual*. Madrid: Aguilar, 2011. Impreso.

ŠIFRAR KALAN, M. (2012). Análisis comparativo de la disponibilidad léxica en español como lengua extranjera (ele) y lengua materna (elm). *Marcoele, 15*. Web

—. "Leksikalna Razpoložljivost v Španščini Kot Tujem Jeziku". Tesis. Univerza v Ljubljani, 2011. Impreso. .

SLAGTER, P. J. *Un nivel umbral*. Eds. Juan Agustín and Yagüe Barredo. 2007th ed. Marco ELE, 5 Vol. Estrasburgo: Consejo de Europa, 1979. Impreso.

SOTO ARANDA, Beatriz, and Mohamed EL-MADKOURI. "La adquisición del español en la población inmigrada en España. Apuntes para una reflexión sobre el paradigma". *Estudios de lingüística Universidad de Alicante* 16 (2002). Impreso.

TOMÉ CORNEJO, Carmela. "Aspectos metodológicos de los estudios de disponibilidad léxica. Los centros de interés". Memoria de Grado. Salamanca: Universidad de Salamanca, 2010. Impreso.

V.V.A.A. *Gran diccionario de uso del español actual*. Ed. A. Sánchez Pérez. Madrid: Sociedad General Española de Librería, 2001. Impreso.

VILLARREAL, Fernando. *Enseñanza de la lengua a inmigrantes. Estudio de políticas de Integración Lingüística En Tres Países Europeos y Retos Para El Caso Español*. Madrid: Ministerio de Trabajo e Inmigración, 2009. Impreso.

WILLIAMS, Lynn. *Aspectos sociolingüísticos del habla de la sociedad de Valladolid*. Valladolid: Universidad de Valladolid, 1988. Impreso.

ANEXOS

ANEXO I: Encuesta a los centros y educadores

Nombre del centro (ONG/CEPA):
Ciudad:
Profesor/Contacto:
Fecha de la encuesta:

Educador

¿Quiénes imparten las clases?
a. Voluntarios
b. Voluntarios y trabajadores
c. Trabajadores

¿En qué elementos se basa la formación del profesorado? (se escoge a especialistas en enseñanza ELE, profesionales de la rama de lo social, se realizan cursos de formación…)

Características generales del alumnado

Número de alumnos en el centro:
Hombres/Mujeres:
Nacionalidades:
Niveles:

Otros aspectos:

Criterios de la división en grupos:
Metodología de enseñanza:
Regularidad de la asistencia:
Principales dificultades:

Percepción de la realidad social del grupo[21]

¿Qué pautas está siguiendo el grupo en su adaptación a la sociedad de acogida?
¿Crees que el grupo de aprendices se encuentra en una situación de aislamiento?
¿Considera el grupo /los grupos a los que pertenecen los estudiantes como un grupo que cubre sus propias necesidades?
¿Crees que existen estereotipos por parte de la sociedad de acogida y de los estudiantes?
¿Los estudiantes tienen intención de permanecer largo tiempo en el país?

ANEXO II: Encuesta a los informantes

Centro:_____
Ciudad:_____

1. Sexo:

a. Hombre
b. Mujer

2. Edad ¿Cuántos **años** tienes?

_____

3. Tiempo en **España**:

_____ años
_____ meses

4. País:

_____

5. Tu **primera lengua**

_____

6. Sabes hablar y escribir en **otras lenguas** que no sean español o tu primera lengua

a. Sí: _____
b. No

7. ¿**Trabajas**?:

a. Sí
b. No

8. Años de **estudio**:

a. **No** he ido al colegio
b. **hasta los 10 años** aproximadamente
c. **hasta los 18** aproximadamente
d. He estudiado cómo realizar un oficio (**Formación profesional**)
e. **Universitarios**

9. ¿Cuánto **tiempo** llevas en **clases de español**?

a.Menos de 1 año
b. Entre 1 y 3 años
c. Más de 3 años

ANEXO III: Prueba de disponibilidad léxica

Centro N

1._____ 21._____
2._____ 22._____
3._____ 23._____
4._____ 24._____
5._____ 25._____
6._____ 26._____
7._____ 27._____
8._____ 28._____
9._____ 29._____
10._____ 30._____
11._____ 31._____
12._____ 32._____
13._____ 33._____
14._____ 34._____
15._____ 35._____
16._____ 36._____
17._____ 37._____
18._____ 38._____
19._____ 39._____
20._____ 40._____

ANEXO IV. Ordenado por disponibilidad: 'Partes del cuerpo'

Nº	Palabra	Disponibilidad	Frecuencia relativa	% Aparición	Frecuencia acumulada
1	mano	0.73295	8.192 %	88.650 %	8.192 %
2	cabeza	0.66056	6.859 %	74.233 %	15.051 %
3	ojo	0.59085	7.143 %	77.301 %	22.194 %
4	nariz	0.39458	5.385 %	58.282 %	27.579 %
5	dedo	0.35305	5.130 %	55.521 %	32.709 %
6	boca	0.33455	4.705 %	50.920 %	37.414 %
7	pierna	0.32885	4.422 %	47.853 %	41.836 %
8	pelo	0.31758	4.592 %	49.693 %	46.428 %
9	oreja	0.30605	4.308 %	46.626 %	50.736 %
10	pie	0.26193	3.571 %	38.650 %	54.307 %
11	brazo	0.20618	2.948 %	31.902 %	57.255 %
12	espalda	0.17229	2.806 %	30.368 %	60.061 %
13	cara	0.17085	2.183 %	23.620 %	62.244 %
14	cuello	0.14708	2.239 %	24.233 %	64.483 %
15	diente	0.14237	2.353 %	25.460 %	66.836 %
16	uña	0.13732	2.466 %	26.687 %	69.302 %
17	rodilla	0.13263	2.409 %	26.074 %	71.711 %
18	pecho	0.12290	2.154 %	23.313 %	73.865 %
19	lengua	0.09589	1.672 %	18.098 %	75.537 %
20	estómago	0.09159	1.729 %	18.712 %	77.266 %
21	piel	0.08971	1.219 %	13.190 %	78.485 %
22	corazón	0.08858	1.446 %	15.644 %	79.931 %
23	muñeca	0.08660	1.389 %	15.031 %	81.320 %
24	tripa	0.07331	1.219 %	13.190 %	82.539 %
25	labio	0.06806	1.219 %	13.190 %	83.758 %
26	codo	0.06614	1.190 %	12.883 %	84.948 %
27	hombro	0.06025	0.964 %	10.429 %	85.912 %
28	frente	0.05856	1.049 %	11.350 %	86.961 %
29	culo	0.05658	1.077 %	11.656 %	88.038 %
30	barriga	0.04840	0.850 %	9.202 %	88.888 %
31	ceja	0.04619	0.822 %	8.896 %	89.710 %
32	cuerpo	0.03370	0.368 %	3.988 %	90.078 %
33	garganta	0.03188	0.510 %	5.521 %	90.588 %
34	hígado	0.03071	0.567 %	6.135 %	91.155 %
35	pulmón	0.02960	0.595 %	6.442 %	91.750 %
36	muslo	0.02663	0.482 %	5.215 %	92.232 %
37	tobillo	0.02423	0.397 %	4.294 %	92.629 %
38	barba	0.01946	0.397 %	4.294 %	93.026 %
39	mejilla	0.01846	0.368 %	3.988 %	93.394 %
40	hueso	0.01607	0.255 %	2.761 %	93.649 %
41	cerebro	0.01601	0.227 %	2.454 %	93.876 %
42	músculo	0.01562	0.340 %	3.681 %	94.216 %
43	cintura	0.01454	0.255 %	2.761 %	94.471 %
44	sangre	0.01400	0.283 %	3.067 %	94.754 %
45	tronco	0.01288	0.170 %	1.840 %	94.924 %
46	muela	0.01184	0.227 %	2.454 %	95.151 %
47	nuca	0.01180	0.198 %	2.147 %	95.349 %
48	cabello	0.01178	0.227 %	2.454 %	95.576 %
49	pestaña	0.01094	0.283 %	3.067 %	95.859 %
50	cadera	0.01063	0.198 %	2.147 %	96.057 %
51	bigote	0.01042	0.170 %	1.840 %	96.227 %
52	oído	0.00980	0.142 %	1.534 %	96.369 %
53	teta	0.00801	0.170 %	1.840 %	96.539 %
54	talón	0.00716	0.227 %	2.454 %	96.766 %
55	abdomen	0.00716	0.142 %	1.534 %	96.908 %
56	barbilla	0.00667	0.113 %	1.227 %	97.021 %
57	pata	0.00616	0.085 %	0.920 %	97.106 %
58	riñón	0.00613	0.170 %	1.840 %	97.276 %
59	ombligo	0.00583	0.113 %	1.227 %	97.389 %
60	costilla	0.00551	0.085 %	0.920 %	97.474 %
61	vena	0.00505	0.113 %	1.227 %	97.587 %
62	pene	0.00447	0.113 %	1.227 %	97.700 %
63	fémur	0.00407	0.057 %	0.613 %	97.757 %

64	columna (vertebral)	0.00398	0.085 %	0.920 %	97.842 %
65	cúbito	0.00375	0.057 %	0.613 %	97.899 %
66	intestino	0.00355	0.113 %	1.227 %	98.012 %
67	vientre	0.00338	0.057 %	0.613 %	98.069 %
68	vagina	0.00325	0.085 %	0.920 %	98.154 %
69	humano	0.00283	0.028 %	0.307 %	98.182 %
70	bello	0.00283	0.028 %	0.307 %	98.210 %
71	pupila	0.00277	0.057 %	0.613 %	98.267 %
72	palma	0.00264	0.057 %	0.613 %	98.324 %
73	extremidad	0.00260	0.028 %	0.307 %	98.352 %
74	oír	0.00260	0.028 %	0.307 %	98.380 %
75	amígdala	0.00240	0.028 %	0.307 %	98.408 %
76	esófago	0.00240	0.057 %	0.613 %	98.465 %
77	seno	0.00240	0.028 %	0.307 %	98.493 %
78	mandíbula	0.00240	0.028 %	0.307 %	98.521 %
79	dentro	0.00240	0.028 %	0.307 %	98.549 %
80	gemelo	0.00236	0.057 %	0.613 %	98.606 %
81	encía	0.00229	0.057 %	0.613 %	98.663 %
82	dedo del pie	0.00221	0.028 %	0.307 %	98.691 %
83	vértebra	0.00221	0.028 %	0.307 %	98.719 %
84	moflete	0.00221	0.028 %	0.307 %	98.747 %
85	derecho	0.00221	0.028 %	0.307 %	98.775 %
86	dedo de la mano	0.00221	0.028 %	0.307 %	98.803 %
87	puño	0.00221	0.028 %	0.307 %	98.831 %
88	izquierdo	0.00203	0.028 %	0.307 %	98.859 %
89	carrillo	0.00187	0.028 %	0.307 %	98.887 %
90	escuchar	0.00187	0.028 %	0.307 %	98.915 %
91	tórax	0.00187	0.028 %	0.307 %	98.943 %
92	perilla	0.00187	0.028 %	0.307 %	98.971 %
93	carne	0.00173	0.028 %	0.307 %	98.999 %
94	polla	0.00173	0.057 %	0.613 %	99.056 %
95	célula	0.00173	0.028 %	0.307 %	99.084 %
96	índice	0.00173	0.028 %	0.307 %	99.112 %
97	pulgar	0.00159	0.028 %	0.307 %	99.140 %
98	dorso	0.00146	0.028 %	0.307 %	99.168 %
99	nervio	0.00146	0.028 %	0.307 %	99.196 %
100	pectoral	0.00146	0.028 %	0.307 %	99.224 %
101	clavícula	0.00146	0.028 %	0.307 %	99.252 %
102	orina	0.00135	0.028 %	0.307 %	99.280 %
103	tibia	0.00135	0.028 %	0.307 %	99.308 %
104	arteria	0.00135	0.028 %	0.307 %	99.336 %
105	faz	0.00135	0.028 %	0.307 %	99.364 %
106	nuez	0.00135	0.028 %	0.307 %	99.392 %
107	trasero	0.00134	0.057 %	0.613 %	99.449 %
108	peroné	0.00124	0.028 %	0.307 %	99.477 %
109	esternón	0.00124	0.028 %	0.307 %	99.505 %
110	tumbar	0.00124	0.028 %	0.307 %	99.533 %
111	iris	0.00114	0.028 %	0.307 %	99.561 %
112	inclinar	0.00114	0.028 %	0.307 %	99.589 %
113	coño	0.00105	0.028 %	0.307 %	99.617 %
114	nudillo	0.00097	0.028 %	0.307 %	99.645 %
115	cola	0.00097	0.028 %	0.307 %	99.673 %
116	genitales	0.00097	0.028 %	0.307 %	99.701 %
117	planta del pie	0.00097	0.028 %	0.307 %	99.729 %
118	saliva	0.00089	0.028 %	0.307 %	99.757 %
119	víscera	0.00089	0.028 %	0.307 %	99.785 %
120	rótula	0.00082	0.028 %	0.307 %	99.813 %
121	espinilla	0.00082	0.028 %	0.307 %	99.841 %
122	pechuga	0.00070	0.028 %	0.307 %	99.869 %
123	vesícula	0.00064	0.028 %	0.307 %	99.897 %
124	gordo	0.00050	0.028 %	0.307 %	99.925 %
125	memoria	0.00046	0.028 %	0.307 %	99.953 %
126	delgado	0.00046	0.028 %	0.307 %	%

ANEXO V. Orden alfabético: Partes del cuerpo

Nº	Palabra	Disponibilidad	Frecuencia relativa	% Aparición	Frecuencia acumulada
1	abdomen	0.00716	0.142 %	1.534 %	0.142 %
2	amígdala	0.00240	0.028 %	0.307 %	0.170 %
3	arteria	0.00135	0.028 %	0.307 %	0.198 %
4	barba	0.01946	0.397 %	4.294 %	0.595 %
5	barbilla	0.00667	0.113 %	1.227 %	0.708 %
6	barriga	0.04840	0.850 %	9.202 %	1.558 %
7	bello	0.00283	0.028 %	0.307 %	1.586 %
8	bigote	0.01042	0.170 %	1.840 %	1.756 %
9	boca	0.33455	4.705 %	50.920 %	6.461 %
10	brazo	0.20618	2.948 %	31.902 %	9.409 %
11	cabello	0.01178	0.227 %	2.454 %	9.636 %
12	cabeza	0.66056	6.859 %	74.233 %	16.495 %
13	cadera	0.01063	0.198 %	2.147 %	16.693 %
14	cara	0.17085	2.183 %	23.620 %	18.876 %
15	carne	0.00173	0.028 %	0.307 %	18.904 %
16	carrillo	0.00187	0.028 %	0.307 %	18.932 %
17	ceja	0.04619	0.822 %	8.896 %	19.754 %
18	célula	0.00173	0.028 %	0.307 %	19.782 %
19	cerebro	0.01601	0.227 %	2.454 %	20.009 %
20	cintura	0.01454	0.255 %	2.761 %	20.264 %
21	clavícula	0.00146	0.028 %	0.307 %	20.292 %
22	codo	0.06614	1.190 %	12.883 %	21.482 %
23	cola	0.00097	0.028 %	0.307 %	21.510 %
24	columna (vertebral)	0.00398	0.085 %	0.920 %	21.595 %
25	coño	0.00105	0.028 %	0.307 %	21.623 %
26	corazón	0.08858	1.446 %	15.644 %	23.069 %
27	costilla	0.00551	0.085 %	0.920 %	23.154 %
28	cúbito	0.00375	0.057 %	0.613 %	23.211 %
29	cuello	0.14708	2.239 %	24.233 %	25.450 %
30	cuerpo	0.03370	0.368 %	3.988 %	25.818 %
31	culo	0.05658	1.077 %	11.656 %	26.895 %
32	dedo	0.35305	5.130 %	55.521 %	32.025 %
33	dedo de la mano	0.00221	0.028 %	0.307 %	32.053 %
34	dedo del pie	0.00221	0.028 %	0.307 %	32.081 %
35	delgado	0.00046	0.028 %	0.307 %	32.109 %
36	dentro	0.00240	0.028 %	0.307 %	32.137 %
37	derecho	0.00221	0.028 %	0.307 %	32.165 %
38	diente	0.14237	2.353 %	25.460 %	34.518 %
39	dorso	0.00146	0.028 %	0.307 %	34.546 %
40	encía	0.00229	0.057 %	0.613 %	34.603 %
41	escuchar	0.00187	0.028 %	0.307 %	34.631 %
42	esófago	0.00240	0.057 %	0.613 %	34.688 %
43	espalda	0.17229	2.806 %	30.368 %	37.494 %
44	espinilla	0.00082	0.028 %	0.307 %	37.522 %
45	esternón	0.00124	0.028 %	0.307 %	37.550 %
46	estómago	0.09159	1.729 %	18.712 %	39.279 %
47	extremidad	0.00260	0.028 %	0.307 %	39.307 %
48	faz	0.00135	0.028 %	0.307 %	39.335 %
49	fémur	0.00407	0.057 %	0.613 %	39.392 %
50	frente	0.05856	1.049 %	11.350 %	40.441 %
51	garganta	0.03188	0.510 %	5.521 %	40.951 %
52	gemelo	0.00236	0.057 %	0.613 %	41.008 %
53	genitales	0.00097	0.028 %	0.307 %	41.036 %
54	gordo	0.00050	0.028 %	0.307 %	41.064 %
55	hígado	0.03071	0.567 %	6.135 %	41.631 %
56	hombro	0.06025	0.964 %	10.429 %	42.595 %
57	hueso	0.01607	0.255 %	2.761 %	42.850 %
58	humano	0.00283	0.028 %	0.307 %	42.878 %
59	inclinar	0.00114	0.028 %	0.307 %	42.906 %
60	índice	0.00173	0.028 %	0.307 %	42.934 %
61	intestino	0.00355	0.113 %	1.227 %	43.047 %
62	iris	0.00114	0.028 %	0.307 %	43.075 %
63	izquierdo	0.00203	0.028 %	0.307 %	43.103 %

64	labio	0.06806	1.219 %	13.190 %	44.322 %
65	lengua	0.09589	1.672 %	18.098 %	45.994 %
66	mandíbula	0.00240	0.028 %	0.307 %	46.022 %
67	mano	0.73295	8.192 %	88.650 %	54.214 %
68	mejilla	0.01846	0.368 %	3.988 %	54.582 %
69	memoria	0.00046	0.028 %	0.307 %	54.610 %
70	moflete	0.00221	0.028 %	0.307 %	54.638 %
71	muela	0.01184	0.227 %	2.454 %	54.865 %
72	muñeca	0.08660	1.389 %	15.031 %	56.254 %
73	músculo	0.01562	0.340 %	3.681 %	56.594 %
74	muslo	0.02663	0.482 %	5.215 %	57.076 %
75	nariz	0.39458	5.385 %	58.282 %	62.461 %
76	nervio	0.00146	0.028 %	0.307 %	62.489 %
77	nuca	0.01180	0.198 %	2.147 %	62.687 %
78	nudillo	0.00097	0.028 %	0.307 %	62.715 %
79	nuez	0.00135	0.028 %	0.307 %	62.743 %
80	oído	0.00980	0.142 %	1.534 %	62.885 %
81	oír	0.00260	0.028 %	0.307 %	62.913 %
82	ojo	0.59085	7.143 %	77.301 %	70.056 %
83	ombligo	0.00583	0.113 %	1.227 %	70.169 %
84	oreja	0.30605	4.308 %	46.626 %	74.477 %
85	orina	0.00135	0.028 %	0.307 %	74.505 %
86	palma	0.00264	0.057 %	0.613 %	74.562 %
87	pata	0.00616	0.085 %	0.920 %	74.647 %
88	pecho	0.12290	2.154 %	23.313 %	76.801 %
89	pechuga	0.00070	0.028 %	0.307 %	76.829 %
90	pectoral	0.00146	0.028 %	0.307 %	76.857 %
91	pelo	0.31758	4.592 %	49.693 %	81.449 %
92	pene	0.00447	0.113 %	1.227 %	81.562 %
93	perilla	0.00187	0.028 %	0.307 %	81.590 %
94	peroné	0.00124	0.028 %	0.307 %	81.618 %
95	pestaña	0.01094	0.283 %	3.067 %	81.901 %
96	pie	0.26193	3.571 %	38.650 %	85.472 %
97	piel	0.08971	1.219 %	13.190 %	86.691 %
98	pierna	0.32885	4.422 %	47.853 %	91.113 %
99	planta del pie	0.00097	0.028 %	0.307 %	91.141 %
100	polla	0.00173	0.057 %	0.613 %	91.198 %
101	pulgar	0.00159	0.028 %	0.307 %	91.226 %
102	pulmón	0.02960	0.595 %	6.442 %	91.821 %
103	puño	0.00221	0.028 %	0.307 %	91.849 %
104	pupila	0.00277	0.057 %	0.613 %	91.906 %
105	riñón	0.00613	0.170 %	1.840 %	92.076 %
106	rodilla	0.13263	2.409 %	26.074 %	94.485 %
107	rótula	0.00082	0.028 %	0.307 %	94.513 %
108	saliva	0.00089	0.028 %	0.307 %	94.541 %
109	sangre	0.01400	0.283 %	3.067 %	94.824 %
110	seno	0.00240	0.028 %	0.307 %	94.852 %
111	talón	0.00716	0.227 %	2.454 %	95.079 %
112	teta	0.00801	0.170 %	1.840 %	95.249 %
113	tibia	0.00135	0.028 %	0.307 %	95.277 %
114	tobillo	0.02423	0.397 %	4.294 %	95.674 %
115	tórax	0.00187	0.028 %	0.307 %	95.702 %
116	trasero	0.00134	0.057 %	0.613 %	95.759 %
117	tripa	0.07331	1.219 %	13.190 %	96.978 %
118	tronco	0.01288	0.170 %	1.840 %	97.148 %
119	tumbar	0.00124	0.028 %	0.307 %	97.176 %
120	uña	0.13732	2.466 %	26.687 %	99.642 %
121	vagina	0.00325	0.085 %	0.920 %	99.727 %
122	vena	0.00505	0.113 %	1.227 %	99.840 %
123	vértebra	0.00221	0.028 %	0.307 %	99.868 %
124	vesícula	0.00064	0.028 %	0.307 %	99.896 %
125	vientre	0.00338	0.057 %	0.613 %	99.953 %
126	víscera	0.00089	0.028 %	0.307 %	99.981 %

NOTAS

[1] Programa de contratación PIRTU (Personal de Investigación de Reciente Titulación). El trabajo fue realizado en la Universidad de Valladolid y fue dirigido por la Dra. Carmen Hernández González.

[2] ANEXO I

[3] ANEXO II

[4] Para revisar el vocabulario específico de léxico disponible consultar el aparatado 5.1 referente al programa informático.

[5] Analizaremos detenidamente esta variable en el ejemplo 5.3.1

[6] Encontraremos un desarrollo específico de las variables en Fernández-Merino (*Castilla y León* y en *Encuestas y pruebas*)

[7] Esta labor, al igual que la informatización de los datos fueron llevados a cabo por un solo investigador.

[8] ANEXO III.

[9] DRAE

[10] DPD

[11] DUE

[12] DEA

[13] GDUEA

[14] Sin que tenga relación con símbolos de la fonética.

[15] Estadio del F.C. Barcelona.

[16] Documento Nacional de Identidad, Televisión Digital Terrestre.

[17] Fonéticamente [dulœʀ].

[18] El acceso a este programa informático se gestiona desde la propia página de referencia a través del banco de datos.

[19] Aunque al simplificar los decimales los datos de 07. IRC y 08. ADM coincidan, los datos completos muestran su correcta ubicación en el rango.

[20] ANEXOS IV y V

[21] Selección de preguntas extraídas de la reflexión de Schumann analizadas en Fernández-Merino (*Diálogo*).