

# Léxico disponible de 82 inmigrantes estudiantes de español en la ciudad de Alcalá de Henares

Diego Javier Gallego Gallego

La presente investigación se inscribe dentro de los estudios de disponibilidad léxica que surgieron en Francia hacia mitad del siglo XX gracias al trabajo desarrollado por el equipo de investigadores del *Centre d'Etude du Français élémentaire* a cargo de Gougenheim, y continuados con gran éxito en el mundo hispánico por Humberto López Morales, quien actualmente coordina las investigaciones sobre disponibilidad léxica dentro del *Proyecto panhispánico*. El trabajo que aquí presentamos forma parte de un estudio mucho más amplio en donde se analiza la disponibilidad léxica de los estudiantes de español como lengua extranjera en la Comunidad de Madrid. Dentro de los 300 informantes que participan en dicho estudio, contamos con una submuestra de 82 inmigrantes que en el momento de realización del estudio se encontraban estudiando español en el centro de idiomas Alcalingua de la Universidad de Alcalá y sobre cuyos resultados realizamos el siguiente trabajo investigativo.

**Palabras claves:** léxico disponible, disponibilidad léxica, español como lengua extranjera, inmigrantes, España, Comunidad de Madrid, palabras, vocablos.

This research is included in the studies of lexical availability which emerged in France in the half of the twentieth century by the researchers of the *Centre d'Etude du Français élémentaire*, by Gougenheim, and were continued with great success in the Hispanic world by Humberto López Morales, who currently coordinates several researches about lexical availability within *Panhispánico* Project. This work is part of a larger study where lexical availability of students of Spanish as foreign language in Madrid is analysed in which among the 300 respondents have participated. For this study, we have built a subsample of 82 immigrants which were studying Spanish in Alcalingua Language Center of the University of Alcalá. The following research is based on the results of this subsample.

**Keywords:** lexical available, lexical availability, Spanish as foreign language, immigrants, Spain, Community of Madrid, words, terms.

## 1. Introducción<sup>1</sup>

A pesar de ser una disciplina relativamente nueva dentro del campo de la lingüística, los estudios de disponibilidad léxica en español han logrado consolidarse como una línea de investigación sólida gracias a que cuentan con unos principios teóricos y metodológicos comunes que establecen unas pautas de actuación similares, aplicables tanto a los estudios que analizan la disponibilidad léxica en lengua materna, como a los cada vez más frecuentes estudios sobre el léxico disponible de los estudiantes de español como lengua extranjera. Con el fin de facilitar la tarea comparativa con otras investigaciones similares, hemos tomado algunos de los principales criterios metodológicos utilizados por las investigaciones pertenecientes al *Proyecto panhispánico*, a los que hemos incorporado algunas modificaciones que se adaptan al tipo de informante de nuestra investigación, como lo son los estudiantes extranjeros de español. En concreto, tomamos como referencia para nuestra investigación el trabajo pionero de López Morales (1999), quien investiga la disponibilidad léxica de los estudiantes puertorriqueños; igualmente, nos hemos basado en la investigación realizada por Carcedo González (2000) sobre la disponibilidad léxica de los estudiantes finlandeses de español y finalmente, en el trabajo investigativo de Samper Hernández (2002) donde analiza el léxico disponible de los alumnos de español como lengua extranjera en situación de inmersión lingüística en España.

En nuestro caso hemos desarrollado un estudio con una muestra representativa de 300 estudiantes de español como lengua extranjera de la Comunidad de Madrid, dentro de los que se incluyen 82 estudiantes inmigrantes de español. El propósito fundamental del presente artículo es conocer el léxico disponible de los 82 estudiantes inmigrantes de español que participaron en nuestro estudio. Junto al análisis global de los resultados alcanzados por el conjunto de informantes inmigrantes, queremos analizar si existen diferencias significativas entre los resultados obtenidos por la submuestra de estudiantes inmigrantes y la submuestra de estudiantes de ELE, si la condición de inmigrante influye de manera significativa en la producción léxica de los informantes o si por el contrario, este factor no es determinante. Igualmente, nos hemos planteado como objetivos de trabajo: 1. Establecer cuáles de los centros de interés resultan más productivos, en función del número de palabras mencionadas en cada uno de ellos. 2. Establecer los centros que presentan una mayor riqueza léxica de acuerdo al número de vocablos registrados en los listados de disponibilidad. 3. Definir el carácter compacto o difuso de los centros de interés en función del índice de cohesión, y 4. Comparar los resultados obtenidos por la submuestra de estudiantes

inmigrantes y la submuestra de informantes de ELE en función del promedio de respuestas.

## 2. Fundamentos teóricos

### 2.1. Léxico frecuente y léxico disponible<sup>2</sup>

El concepto de disponibilidad léxica surge durante la investigación llevada a cabo en Francia hacia mitades del pasado siglo por el grupo de investigadores del *Centre d'Etude du Français élémentaire*, a quienes se les encargó la tarea de confeccionar un método de enseñanza de la lengua francesa a los territorios que constituían por entonces la Union Française. Para ello, el grupo de investigadores a cargo de Gougenheim<sup>3</sup> debía seleccionar, dentro del gran caudal léxico de la lengua francesa, aquellos vocablos que debían formar parte de este nivel *élémentaire*. Como era de esperar, los investigadores utilizaron en un primer momento el factor frecuencia para seleccionar los términos, pues se pensaba que las palabras más frecuentes eran también las más usadas. El proceso consistía en seleccionar un corpus representativo de textos y aplicar un método de procesamiento informático que ordenaba las palabras contenidas en dichos textos según el factor frecuencia.

Sin embargo, pronto se dieron cuenta de que a través de este método se dejaban a un lado palabras que eran conocidas y usuales por los franceses, y que, por tanto, deberían aparecer en el léxico que estaban confeccionando. El problema de tal situación radicaba en que a través de los listados de frecuencia se obtenían sobre todo palabras gramaticales (verbos, adjetivos y sustantivos de carácter general), y se dejaban a un lado aquellas palabras que tenían un contenido semántico muy concreto, como es el caso por ejemplo de la palabra *roi* (rey), que era muy conocida y utilizada por los franceses en esa época (a propósito de la muerte del rey de Inglaterra Jorge V). Fue entonces cuando Michéa (1950) comenzó por diferenciar las *palabras atemáticas* de las *palabras temáticas*; estas últimas tenían un alto índice de aparición pero solo en determinados temas, al contrario de las primeras, que aparecían independientemente del tema tratado. En su artículo de 1953, *Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage*, Michéa no solo diferenciaba las *palabras frecuentes* de las *palabras disponibles*, sino que además definía el concepto de “disponibilidad”, iniciando así toda una nueva línea de investigación que hoy se conoce como estudios de disponibilidad léxica<sup>4</sup>:

Una palabra disponible es una palabra que, sin ser particularmente frecuente, está siempre lista para ser empleada y viene inmediata y naturalmente a la mente en el momento en que se tiene necesidad de ella. Es una palabra que, formando parte de las asociaciones de ideas usuales, existe en potencia en el sujeto hablante en cuanto estas asociaciones entran en juego. (1953: 340)<sup>5</sup>.

Esta nueva diferenciación entre *palabras frecuentes* y *palabras disponibles*, llevó al grupo de investigadores franceses a desarrollar pruebas que validaran las nuevas hipótesis, dando como resultado la aparición de *L'Elaboration du Français Élémentaire* en 1956, que más adelante sería modificado en algunos de sus postulados en la edición de 1964, en la que recibió además el nombre de *L'Elaboration du Français Fondamental*. Para su desarrollo, el grupo de investigadores franceses aplicó pruebas de disponibilidad a estudiantes con edades entre los 9 y los 12 años de edad, en donde a partir de unos estímulos verbales llamados centros de interés los estudiantes producían unidades léxicas relacionadas que constituirían el léxico disponible.

A diferencia del *léxico frecuente* que trabaja con palabras que se actualizan independientemente de un tema concreto y que recoge palabras de mayor estabilidad como son los verbos o los adjetivos, en el *léxico disponible* las palabras aparecen en función de un tema concreto que les sirve de estímulo (centros de interés), siendo los sustantivos la categoría gramatical predominante dentro de los listados de léxico disponible. Al respecto, López Chávez (1994: 69) señala que el léxico disponible “es el conjunto de vocablos que utiliza una comunidad de hablantes como parte modular de su léxico patrimonial pero que únicamente aparecen en situaciones precisas, por lo que no son muy frecuentes pero sí muy conocidas. Son inestables debido a que es necesario que se toque un tema específico en el discurso para que aparezcan”.

## 2.2. Palabras temáticas y palabras aтемáticas

A partir de esta constatación, el grupo de investigadores a cargo de Gougenheim empezó entonces a diferenciar conceptos que hasta entonces se utilizaban como sinónimos: palabras frecuentes, palabras comunes y palabras usuales. Esto dio pie a que Michéa propusiera en su artículo sobre *Vocabulaire et culture* publicado en 1950, la diferencia entre palabras temáticas y palabras aтемáticas:

Ce sont des mots qui servent à nous exprimer au sujet des choses plutôt qu'à exprimer les choses elles-mêmes, des termes plus ou

moins communs à tous les sujets, à toutes les situations. (Michéa 1950: 188).

Las *palabras atemáticas* eran precisamente las que aparecían en los listados obtenidos a partir del factor frecuencia, donde independientemente del tema tratado, aparecían con gran regularidad; por el contrario, las *palabras temáticas* eran aquellas que tenían un contenido semántico muy concreto y por lo tanto su aparición estaba condicionada por el tema que se estuviera tratando. Esta diferenciación permitió asociar rápidamente a las palabras atemáticas con los listados de frecuencia, mientras que las palabras temáticas integrarían lo que los estudios franceses llamaron “el léxico disponible”.

El paso siguiente en el crecimiento de esta disciplina, fue desarrollar pruebas que sirvieran para comprobar los nuevos conceptos hallados hasta el momento. Para ello, Michéa aplicó pruebas asociativas en las que les pedía a sus estudiantes de liceo que escribieran las palabras que estuvieran relacionadas con un tema concreto, y las palabras resultantes conformarían el léxico disponible. Es por esto que a diferencia de las palabras frecuentes (o atemáticas), las palabras disponibles (o palabras temáticas) están restringidas a un campo asociativo o tema en concreto. Para que estas palabras se produzcan en nuestra mente, es necesario que se “activen” por medio de un estímulo concreto, que en el caso de los estudios de disponibilidad léxica corresponde a los centros de interés o unidades temáticas que sirven de estímulo durante la realización de las pruebas.

### 2.3. Palabra y vocablo

A pesar de que en el terreno de la disponibilidad léxica en que nos movemos es bastante conocida la diferenciación entre *palabra* y *vocablo*, hemos querido definir de manera breve ambos términos, con el fin de facilitar el acercamiento teórico a aquellas personas no especializadas en la materia. Y es que la diferenciación entre ambos conceptos resulta sobre todo necesaria en el momento de entender los datos presentados en las distintas investigaciones sobre disponibilidad. Por ejemplo, en uno de los primeros estudios realizados en el mundo hispánico, López Morales (1999: 41) señala que los 558 informantes que conforman la muestra puertorriqueña produjeron un total de “[...] 107.276 palabras, de las cuales, 7.311 eran vocablos (palabras diferentes)”, y a continuación muestra una tabla comparando el número de palabras producidas en cada uno de los centros de interés analizados, 16 en total. La gran diferencia numérica entre ambos índices se entiende si consideramos que al computar el total de palabras se registran todas y cada una de las

unidades léxicas (que también podríamos llamar “actualizaciones”) que los informantes escriben en los listados de disponibilidad, mientras que en el cómputo de los vocablos solo se contabilizan aquellas unidades léxicas no repetidas, es decir, las que solo aparecen una vez.

En el *Léxico disponible de los estudiantes preuniversitarios de Málaga*, Ávila Muñoz (2006) propone otra denominación para diferenciar las *palabras* y los *vocablos*. Al presentar los resultados de su investigación, este autor señala que los 458 sujetos que conforman la muestra aportan un total de 129.426 palabras o *tokens* y 9.507 vocablos o *lemas*. Consideramos que en esta propuesta de nomenclatura está latente la necesidad de diferenciar ambos términos, pues como hemos mencionado anteriormente, es necesaria para poder interpretar adecuadamente los datos obtenidos a través de las pruebas de disponibilidad. Más adelante, este mismo autor nos recuerda que la distinción entre *palabra* y *vocablo* es utilizada habitualmente dentro de los estudios de léxico-estadística, y nos presenta una definición bastante clara de ambos conceptos:

En el ámbito de la lexicoestadística es común la distinción entre *palabra* o cada una de las formas –repetidas o no– con las que se trabaja y *vocablo*, término que designa a cada una de las palabras diferentes. (Ávila Muñoz, 2006: 82).

### 3. Fundamentos metodológicos

Como se ha mencionado más arriba, el presente estudio hace parte de una investigación mucho más amplia en la que se analiza la disponibilidad léxica de los estudiantes de español como lengua extranjera de la Comunidad de Madrid. Dentro de los 300 informantes que conforman el total de la muestra, contamos a su vez con dos submuestras de diferente tamaño y características: la primera está constituida por 218 estudiantes de Español Lengua Extranjera (en adelante ELE) y la segunda por 82 estudiantes inmigrantes de español, sobre la cual nos centraremos en este artículo. La principal diferencia entre los dos grupos de informantes radica fundamentalmente en que los primeros son estudiantes de español que han seguido un proceso de escolarización en lengua española, ya sea en su país de origen o en España, y que en el momento de realización de la prueba se encontraban estudiando español en alguna academia o escuela de idiomas de la Comunidad de Madrid. Los informantes inmigrantes por su parte, son por lo general personas adultas que no han seguido un proceso de aprendizaje reglado de la lengua española y que se encontraban en el territorio español por razones principalmente de trabajo, es decir, que han aprendido la lengua debido a la situación de contacto y no por un proceso de enseñanza continuada<sup>6</sup>.

Al igual que la mayoría de las investigaciones pertenecientes al *Proyecto panhispánico*, hemos utilizado pruebas de disponibilidad abiertas con un tiempo de respuesta de dos minutos por centro de interés. En nuestro caso, además de los dieciséis centros tradicionales utilizados en las primeras investigaciones francesas<sup>7</sup>, hemos incluido cuatro centros más: CI17: ‘Los colores’, que viene siendo habitual en las actuales investigaciones hispánicas sobre disponibilidad léxica; CI18: ‘Adjetivos’, que se ha utilizado en las investigaciones madrileñas sobre disponibilidad; CI19: ‘Servicios, documentos, lugares necesarios para permanecer en España’<sup>8</sup>, y CI20: ‘España’, propios de esta investigación. Con la inclusión de estos dos últimos centros queremos observar de manera general si las nuevas áreas temáticas se refieren a realidades “cercanas” a los estudiantes extranjeros de español, y de manera específica, conocer el léxico que diferencia a los estudiantes inmigrantes de español de los demás estudiantes de ELE.

En cuanto a los factores sociológicos tenidos en cuenta en nuestro estudio, hemos incluido además de las variables “sexo” y “nivel sociocultural”, consideradas comunes a muchos trabajos sobre disponibilidad, algunas otras que se adecuan más al objetivo de esta investigación, como lo es el conocer los factores sociales y culturales que influyen en el léxico de los hablantes extranjeros de español, y que han sido utilizadas por autores como Carcedo González (2000) y Samper Hernández (2002) en sus investigaciones sobre disponibilidad léxica en ELE. Las variables sociológicas utilizadas en nuestra investigación son: sexo, edad, lengua materna, nivel de español, motivación para aprender español<sup>9</sup>, conocimiento de otras lenguas extranjeras (además del español) y nivel sociocultural. Vale la pena señalar que el nivel de español se adjudicó de acuerdo al nivel expresado por los informantes en las mismas pruebas de disponibilidad, y que fue corroborado con los grupos o niveles donde se aplicaron las pruebas (inicial, intermedio y avanzado). Por otra parte, para determinar el nivel sociocultural de los informantes, hemos considerado los siguientes factores: nivel de estudios del informante, nivel económico del informante, nivel de estudios de los padres y nivel económico de los padres.

Debido a la cantidad de variables utilizadas en nuestro estudio, siete en total, así como a las diferentes variantes de cada una de ellas, nos vimos en la necesidad de utilizar el programa informático *SPSS (Statistical Package for the Social Sciences)* que se adapta a los requerimientos propios de nuestra investigación. A diferencia de los programas tradicionales para el cálculo de la disponibilidad (*Lexidisp* y *Dispolex*) que trabajan con una serie de variables fijas, con sus respectivas variantes, que no permiten modificación, el paquete estadístico *SPSS* presenta, entre otras ventajas, la de establecer parámetros de análisis individualiza-

dos sin límite de variables, lo cual facilitó el tratamiento informático de los resultados léxicos.

## 4. Perfil de los estudiantes inmigrantes de español

De los 82 inmigrantes que conforman esta submuestra, 49 son hombres (59,76%) y 33 mujeres (40,24%). Por edades, encontramos que la gran mayoría, 63 informantes, tienen entre 20 y 34 años; 16 informantes tienen entre 35 y 55 años; dos tienen más de 55 años y solo un informante es menor de 20 años. Al analizar la lengua materna de los estudiantes inmigrantes de acuerdo a la tipología de lenguas, encontramos que las lenguas eslavas y las niger-congolesas conforman el 57,31% del total de la muestra (26 y 21 informantes respectivamente), seguidas por las lenguas afroasiáticas (12 informantes; 14,63%) y por las lenguas romance (11 informantes; 13,41%). Las lenguas germánicas por su parte constituyen el 8,54% del total de la muestra (7 informantes) mientras que el 6,1% (5 informantes) habla alguna de las lenguas asiáticas. Es importante señalar que dada la gran cantidad de lenguas maternas encontradas durante la realización de nuestra investigación (hasta 26 lenguas distintas), hemos decidido agruparlas por tipos de lenguas según compartieran unas características comunes, independientemente de la familia a la que pertenecieran. De esta manera, pudimos reducir significativamente el número de lenguas hasta quedarnos con seis tipos de lenguas diferentes: lenguas eslavas (ruso, polaco, ucraniano, búlgaro y checo), lenguas germánicas (inglés, alemán, danés, sueco y holandés), lenguas romance o románicas (rumano, italiano, francés y portugués), lenguas niger-congolesas (ibo, manchako, wolof o volof y suajili o suahilí), lenguas afroasiáticas (árabe), y lenguas asiáticas (chino mandarín<sup>10</sup>, coreano<sup>11</sup>, japonés<sup>12</sup> y taiwanés<sup>13</sup>).



Variantes	Variables						
	Sexo	Edad	Lengua materna	Nivel de español	Motivación aprender español	Conocimiento de otras lenguas	Nivel sociocultural
Hombres	49						
Mujeres	33						
0 y 19 años		1					
20 y 34 años		63					
35 y 55 años		16					
Más de 55 años		2					
Eslavas			26				
Germánicas			7				
Románicas			11				
Níger-congolesas			21				
Afroasiáticas			12				
Asiáticas			5				
Inicial				20			
Intermedio				50			
Avanzado				12			
Por estudio					46		
Por trabajo					34		
Ambos					2		
No						11	
Sí						71	
Bajo							29
Medio							33
Alto							20
Totales	82	82	82	82	82	82	82

Tabla 1. *Distribución de la muestra de inmigrantes de acuerdo a las variables de estratificación*

Por otra parte, encontramos que más de la mitad de los informantes (60,98%) tienen un nivel intermedio de español, seguidos por los del nivel inicial (24,39%) y finalmente por los estudiantes del nivel avanzado, que constituyen el 14,63% del total de la muestra. El 56,01% de los estudiantes inmigrantes (46 informantes) tienen como principal motivación para estudiar español el estudio, mientras que el 41,46% (34 informantes) lo hace por razones de trabajo. De los 82 informantes que

conforman la submuestra, solo dos (2,44%) manifiestan estudiar español por ambas razones, es decir, por estudio y por trabajo. Igualmente, observamos que un amplio porcentaje de los estudiantes inmigrantes de español (86,59%) tienen conocimientos de otras lenguas extranjeras además del español, frente al 13,41% que no lo tienen. Finalmente, observamos que el 40,24% de la muestra (33 informantes), pertenecen al nivel sociocultural medio, seguidos por el 35,37% que pertenecen al nivel bajo (29 informantes), mientras que el restante 24,39% (20 informantes) pertenecen al nivel alto.

En la Tabla 1 presentamos la distribución sociológica de la muestra de acuerdo a las distintas variables de estratificación consideradas en nuestra investigación.

## 5. Léxico disponible de los estudiantes inmigrantes de español: resultados generales

Los 82 estudiantes inmigrantes que conforman la submuestra producen un total de de 13.556 palabras, con un promedio de 165,32 palabras por informante y 677,8 palabras por centro de interés. Igualmente, encontramos que los estudiantes inmigrantes producen en promedio 8,27 palabras por centro de interés, dato equiparable al alcanzado por los inmigrantes de la investigación de Fernández-Merino Gutiérrez (2011), que presentan una media de 8,25 palabras por centro. Si comparamos estos resultados con otras investigaciones con hablantes no nativos, encontramos que los estudiantes inmigrantes presentan un mayor promedio que el alcanzado por los estudiantes finlandeses de español (Carcedo González 2000), quienes producen una media de 5,55 palabras por centro de interés, mientras que los estudiantes de español como lengua extranjera de la investigación de Samper Hernández (2002) producen en promedio 11,29 palabras por centro.

A continuación presentamos los principales resultados cuantitativos obtenidos por los estudiantes inmigrantes de español, dentro de los que incluimos la frecuencia media de respuestas (FMR)<sup>14</sup>, el número de total de vocablos y el índice de cohesión (ICOH), con el fin de conocer el carácter abierto o cerrado del léxico disponible de nuestros informantes. Finalmente, realizaremos un análisis comparativo entre los resultados obtenidos por los dos grupos que conforman nuestra muestra: 218 estudiantes de ELE y 82 estudiantes inmigrantes de español.

## 5.1. Frecuencia media de respuestas (FMR)

Al dividir el número de palabras aportadas por los estudiantes inmigrantes de español (13.556) por el número de sujetos que conforman esta submuestra (82), obtenemos la frecuencia media de respuestas (FMR), que en este caso es de 165,32. Como se observa en la Tabla 2, el centro con mayor FMR es ‘Partes del cuerpo’ (FMR: 13,34), seguido por ‘Comidas y bebidas’ (FMR: 13,05) y ‘La ciudad’ (FMR: 10,59). Por el contrario, los centros con menor FMR son ‘Servicios, documentos, trámites o lugares necesarios para permanecer en España’, con un frecuencia media de respuestas por informante de 6,37; ‘Iluminación, calefacción y ventilación’ (FMR: 4,52), y ‘Trabajos del campo y del jardín’, que presenta una FMR de 3,67.

Rango	Centros de interés	NPal	FMR
1	01. Partes del cuerpo	1.094	13,34
2	05. Comidas y bebidas	1.070	13,05
3	10. La ciudad	868	10,59
4	18. Adjetivos	819	9,99
5	14. Los animales	808	9,85
6	02. La ropa	765	9,33
7	16. Profesiones y oficios	741	9,04
8	17. Los colores	732	8,93
9	20. España	694	8,46
10	04. Los muebles de la casa	676	8,24
11	08. La escuela	676	8,24
12	03. La casa	665	8,11
13	12. Medios de transporte	589	7,18
14	15. Juegos y distracciones	556	6,78
15	11. El campo	552	6,73
16	06. Objetos colocados en la mesa...	529	6,45
17	07. La cocina	528	6,44
18	19. Servicios, documentos...	522	6,37
19	09. Iluminación, calefacción...	371	4,52
20	13. Trabajos del campo...	301	3,67

Tabla 2. Rango que alcanzan los centros de interés respecto al número de palabras y a la frecuencia media de respuestas por informante

Consideramos que la mayor o menor producción de palabras en cada centro, está relacionada con la naturaleza propia de cada área temática, con las que los estudiantes inmigrantes establecen una relación de distancia o cercanía que influye en el número de respuestas mencionadas en cada una de ellas. Si analizamos los centros con mayor número de palabras ('Partes del cuerpo', 'Comidas' y 'bebidas', y 'La ciudad'), encontramos que son áreas temáticas que se refieren a realidades más cercanas a los estudiantes que aquellos que ocupan las últimas posiciones ('Servicios, documentos, trámites o lugares necesarios para permanecer en España', 'Iluminación, calefacción y ventilación' y 'Trabajos del campo y del jardín'). Si analizamos las palabras mencionadas por los informantes en el centro 05, encontramos que se refieren a una necesidad básica para todos, como lo es la alimentación, lo cual explicaría la alta producción de palabras en este centro. Por otra parte, encontramos que cada vez es más frecuente el uso de un vocabulario específico y/o especializado para referirnos al cuerpo humano (centro 01), gracias a los constantes *inputs* recibidos de los medios de comunicación (prensa, cine, televisión). Por otro lado, teniendo en cuenta la naturaleza inmigrante de estos informantes, consideramos que el conocer un vocabulario amplio sobre el cuerpo humano puede facilitar su acercamiento a algunos servicios básicos como lo es la salud, lo que explicaría en parte la gran cantidad de actualizaciones producidas en este centro. Igualmente, consideramos que los estudiantes inmigrantes tienen una relación directa con la ciudad (centro 10), ya sea porque hayan vivido en ella en sus países de origen, o porque se encontraban viviendo en una ciudad en el momento de realización de la prueba. Por el contrario, consideramos que las áreas temáticas que alcanzan los últimos lugares en producción de palabras remiten a campos semánticos más lejanos o desconocidos para el tipo de informantes de nuestra investigación, con los que tienen una relación menos familiar y por tanto un menor conocimiento léxico<sup>15</sup>.

## 5.2. Número total de vocablos

Los estudiantes inmigrantes de español presentan una producción total de vocablos de 2.327, con un promedio de 116,35 vocablos por centro de interés. Como se observa en la Tabla 3, el centro con mayor número de vocablos es 'España', con 245, seguido por 'Adjetivos' y 'Profesiones y oficios', con 187 y 185 vocablos respectivamente. En el otro extremo de la tabla se ubican los centros 'La casa' (67 vocablos), 'Medios de transporte' (53 vocablos) y 'Los colores', con 52 vocablos.

Rango	Centros de interés	NVoc
1	20. España	245
2	18. Adjetivos	187
3	16. Profesiones y oficios	185
4	15. Juegos y distracciones	175
5	10. La ciudad	157
6	05. Comidas y bebidas	156
7	19. Servicios, documentos...	125
8	11. El campo	120
9	14. Los animales	112
10	08. La escuela	111
11	13. Trabajos del campo...	96
12	04. Los muebles de la casa	90
13	07. La cocina	86
14	02. La ropa	84
15	01. Partes del cuerpo	81
16	06. Objetos colocados en la mesa...	76
17	09. Iluminación, calefacción...	69
18	03. La casa	67
19	12. Medios de transporte	53
20	17. Los colores	52

Tabla 3. *Rango que ocupan los centros de interés respecto a la producción de vocablos*

Por otro lado, atendiendo al promedio de vocablos por centro de interés (116,35), encontramos que solo 8 de los 20 centros encuestados superan dicho promedio (centros 05, 10, 11, 15, 16, 18, 19 y 20), mientras que los restantes 12 centros se encuentran por debajo de la media en cuanto a producción de vocablos (Gráfico 1). Vale la pena señalar la gran diferencia que existe entre el centro ‘España’, que alcanza la primera posición en producción de vocablos (245) frente a ‘Los colores’, con solo 52 vocablos y que ocupa la última posición. La diferencia entre ambos centros es de 193 vocablos, donde el centro 20 supera la media en más de 128 vocablos, mientras que el 17 se encuentra 64 vocablos por debajo del promedio.

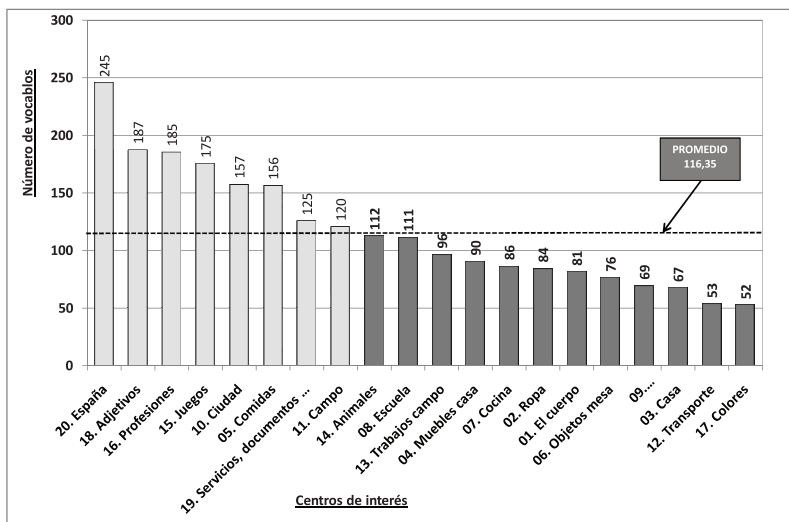


Gráfico 1. *Distribución de los centros de interés en función del número de vocablos*

Una posible explicación sobre por qué los informantes producen un menor número de vocablos en algunos centros de interés, la encontramos en las posibilidades asociativas que presentan unos centros y otros. Si observamos los centros que ocupan las últimas posiciones en cuanto a producción de vocablos (‘La casa’, ‘Medios de transporte’ y ‘Los colores’), encontramos que a diferencia de los demás centros, estas tres áreas temáticas presentan un número más reducido de posibilidades de actualización de vocablos, ya que son campos asociativos más restringidos.

Por ejemplo, en el centro 17, además de los nombres de los colores, los estudiantes pueden actualizar palabras referidas a las cualidades de la luz, como el tono, el brillo o la intensidad de dichos colores, como en efecto lo han hecho, pero agotadas estas asociaciones, las posibilidades de evocación son escasas para este centro. Igualmente, el centro 12 ofrece la posibilidad de actualizar diferentes medios de transporte: terrestres, aéreos, marítimos, así como nombres de objetos o lugares relacionados con dichos transportes; sin embargo, una vez mencionadas estas posibilidades asociativas, este núcleo temático tiende a ser limitado. Una situación similar ocurre con el centro referido a las partes de la casa, donde las posibilidades de asociación están restringidas a la idea o percepción que cada inmigrante tenga de una casa, y en donde a pesar de las indicaciones hechas antes de la prueba, los estudiantes terminan mencionando muebles y electrodomésticos que técnicamente no pertenecen a esta área temática.

Por otra parte, al comparar el número de palabras y de vocablos producidos por la submuestra de estudiantes inmigrantes de español, encontramos que a excepción del centro 06 ‘Objetos colocados en la mesa para la comida’ -que ocupa el mismo rango en ambos listados (16.<sup>a</sup> posición)-, todos los demás centros sufren alguna variación en el lugar que alcanzan en cuanto a producción de palabras y de vocablos. Como se observa en la Tabla 4, algunos centros como ‘La escuela’, conservan una posición casi idéntica, pasando del rango 11.<sup>o</sup> en producción de palabras, a la posición 10.<sup>a</sup> en número de vocablos. Otros centros como ‘Los muebles de la casa’, ‘Iluminación, calefacción y ventilación’, ‘La ciudad’ y ‘Adjetivos’, han modificado su rango en dos posiciones (arriba o abajo) respecto a su producción de palabras y de vocablos. En cuatro posiciones han variado su rango los centros ‘La cocina’, ‘Comidas y bebidas’, ‘Los animales’ y ‘Profesiones y oficios’, mientras que los centros ‘La casa’ y ‘Medios de transporte’, lo han hecho en seis posiciones.

Rango	Centros de interés	NPal	Rango	Centros de interés	NVoc
1	01. Partes del cuerpo	1.094	1	20. España	245
2	05. Comidas y bebidas	1.070	2	18. Adjetivos	187
3	10. La ciudad	868	3	16. Profesiones y oficios	185
4	18. Adjetivos	819	4	15. Juegos y distracciones	175
5	14. Los animales	808	5	10. La ciudad	157
6	02. La ropa	765	6	05. Comidas y bebidas	156
7	16. Profesiones y oficios	741	7	19. Servicios, documentos...	125
8	17. Los colores	732	8	11. El campo	120
9	20. España	694	9	14. Los animales	112
10	04. Los muebles de la casa	676	10	08. La escuela	111
11	08. La escuela	676	11	13. Trabajos del campo...	96
12	03. La casa	665	12	04. Los muebles de la casa	90
13	12. Medios de transporte	589	13	07. La cocina	86
14	15. Juegos y distracciones	556	14	02. La ropa	84
15	11. El campo	552	15	01. Partes del cuerpo	81
16	06. Objetos colocados en la mesa ...	529	16	06. Objetos colocados en la mesa...	76
17	07. La cocina	528	17	09. Iluminación, calefacción...	69
18	19. Servicios, documentos...	522	18	03. La casa	67
19	09. Iluminación, calefacción...	371	19	12. Medios de transporte	53
20	13. Trabajos del campo...	301	20	17. Los colores	52

Tabla 4. *Rango que ocupan los centros de interés respecto a la producción de palabras y vocablos*

Otros cambios significativos los encontramos en el centro ‘Partes del cuerpo’, que ha pasado de ocupar el primer lugar en producción de palabras, a situarse en la 15.<sup>a</sup> posición en cuanto al número vocablos, mientras que ‘España’, que ocupa la primera posición en producción de vocablos, ha descendido hasta la 9.<sup>a</sup> posición en número de palabras. Igualmente, resulta significativa la variación que ha sufrido el centro ‘Juegos y distracciones’ y ‘Los colores’, que se han desplazado en 10 y 12 posiciones respecto al rango que ocupan en uno u otro listado.

Finalmente, en el Gráfico 2 se muestra la distribución alcanzada por los centros de interés respecto a la producción de palabras y de vocablos por parte de los estudiantes inmigrantes de español. Igualmente, se indica el número total de palabras y de vocablos producidos en cada uno de ellos.

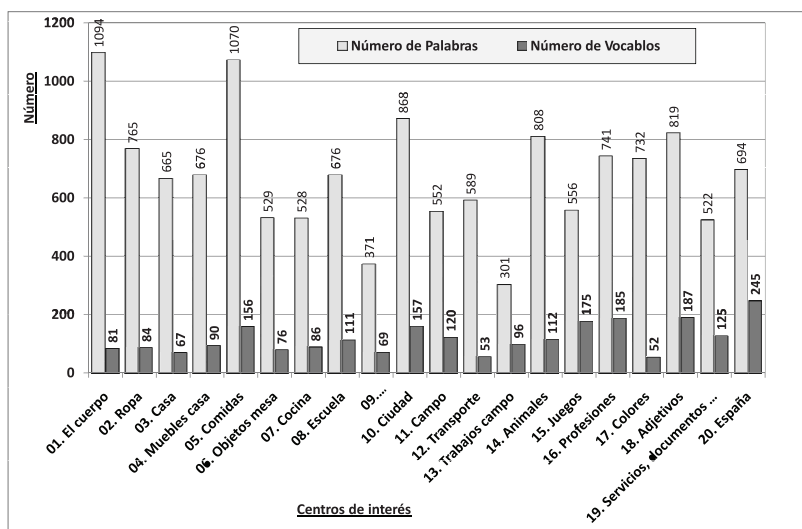


Gráfico 2. *Número de palabras y vocablos por centros de interés*

### 5.3. Índice de cohesión (ICOH)

El índice de cohesión (ICOH) por su parte permite observar el grado de coincidencia en las respuestas de los informantes dentro de un centro de interés y determinar si se trata de un centro compacto (es decir, cuando las respuestas de los informantes coinciden en gran medida) o difuso, cuando existe gran cantidad de palabras diferentes entre las respuestas de los informantes. Este índice, que se obtiene dividiendo el promedio de respuestas de cada sujeto por el número de vocablos o palabras diferentes, sitúa las respuestas en una escala cerrada entre 0 y 1, y será menor cuan-



do exista mayor variación en las respuestas. Es decir, si todos los informantes aportan las mismas unidades léxicas, el índice de cohesión será 1, alejándose de esta cifra a medida que las respuestas varían.

Atendiendo al ICOH, los centros de interés más compactos (o cerrados) para la submuestra de estudiantes inmigrantes de español son ‘Los colores’ (ICOH: 0,172), ‘Partes del cuerpo’ (ICOH: 0,165) y ‘Medios de transporte’, con un índice de cohesión de 0,136 (Tabla 5). Por el contrario, los centros que presentan un carácter más difuso (o abierto) son ‘Juegos y distracciones’ (ICOH: 0,039), ‘Trabajos del campo y del jardín’ (ICOH: 0,038) y ‘España’ (ICOH: 0,035), que alcanzan las últimas posiciones respecto al índice de cohesión.

Rango	Centros de interés	ICOH
1	17. Los colores	0,172
2	01. Partes del cuerpo	0,165
3	12. Medios de transporte	0,136
4	03. La casa	0,121
5	02. La ropa	0,111
6	04. Los muebles de la casa	0,092
7	14. Los animales	0,088
8	06. Objetos colocados en la mesa...	0,085
9	05. Comidas y bebidas	0,084
10	07. La cocina	0,075
11	08. La escuela	0,074
12	10. La ciudad	0,067
13	09. Iluminación, calefacción...	0,066
14	11. El campo	0,056
15	18. Adjetivos	0,053
16	19. Servicios, documentos...	0,051
17	16. Profesiones y oficios	0,049
18	15. Juegos y distracciones	0,039
19	13. Trabajos del campo...	0,038
20	20. España	0,035

Tabla 5. *Rango que alcanzan los centros de interés en función del índice de cohesión (ICOH)*

Por otra parte, en el Gráfico 3 se observa de manera clara la distribución que alcanzan los centros de interés en la submuestra de estudiantes inmigrantes de español en función del índice de cohesión (ICOH), la frecuencia media de respuestas (FMR) y el número de vocablos (Nvoc).

A pesar de que el centro ‘Los colores’ es el que alcanza el mayor ICOH y a su vez presenta el menor número de vocablos, y por el contrario, el centro ‘España’ tiene el mayor número de vocablos y presenta el menor índice de cohesión, no encontramos en los demás centros la misma correspondencia entre ambos índices. Por ejemplo el centro ‘Adjetivos’, ocupa el segundo lugar en producción de vocablos (187) y sin embargo se ubica en la posición 15.<sup>a</sup> de acuerdo al ICOH. De igual manera, si comparamos el índice de cohesión con la frecuencia media de respuestas de cada uno de los centros, no encontramos una relación directa entre estos índices que nos permita definir que los centros con mayor ICOH (es decir, los más compactos) son aquellos que presentan un menor promedio de respuestas, y los centros más difusos (con menor ICOH) como aquellos que tienen una frecuencia media de respuestas por informante superior.

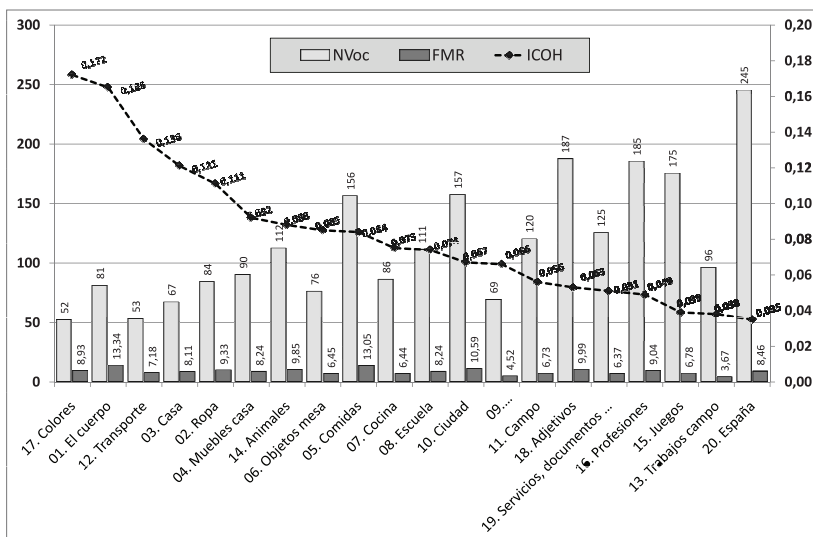


Gráfico 3. Rango que ocupan los centros de interés respecto al índice de cohesión (ICOH). Comparación con el número de vocablos (NVoc) y la frecuencia media de respuesta (FMR)

Para finalizar, presentamos en la Tabla 6 los resultados generales de la submuestra de estudiantes inmigrantes de español, en función de los diferentes índices utilizados durante el análisis cuantitativo: número de palabras (NPal), número de vocablos (NVoc), frecuencia media de respuesta (FMR) e índice de cohesión (ICOH).

Centros de interés	NPal	NVoc	FMR	ICOH
01. Partes del cuerpo	1.094	81	13,34	0,165
02. La ropa	765	84	9,33	0,111
03. La casa	665	67	8,11	0,121
04. Los muebles de la casa	676	90	8,24	0,092
05. Comidas y bebidas	1.070	156	13,05	0,084
06. Objetos colocados en la mesa . . .	529	76	6,45	0,085
07. La cocina	528	86	6,44	0,075
08. La escuela	676	111	8,24	0,074
09. Iluminación, calefacción . . .	371	69	4,52	0,066
10. La ciudad	868	157	10,59	0,067
11. El campo	552	120	6,73	0,056
12. Medios de transporte	589	53	7,18	0,135
13. Trabajos del campo . . .	301	96	3,67	0,038
14. Los animales	808	112	9,85	0,088
15. Juegos y distracciones	556	175	6,78	0,039
16. Profesiones y oficios	741	185	9,04	0,049
17. Los colores	732	52	8,93	0,172
18. Adjetivos	819	187	9,99	0,053
19. Servicios, documentos . . .	522	125	6,37	0,051
20. España	694	245	8,46	0,035

Tabla 6. *Resultados generales de la submuestra de estudiantes inmigrantes*

## 6. Comparación entre los resultados obtenidos por los estudiantes inmigrantes de español y los resultados de los estudiantes no inmigrantes

Como adelantábamos en la introducción del presente artículo, la investigación que aquí se presenta forma parte de un estudio más amplio realizado en la Comunidad de Madrid con 300 estudiantes extranjeros, dentro de los que se incluyen los 82 estudiantes inmigrantes de español. En este apartado queremos establecer un paralelismo entre los resultados alcanzados por los dos grupos de estudiantes de español que conforman la muestra general: 218 estudiantes no inmigrantes y 82 estudiantes inmigrantes. Teniendo en cuenta que el tamaño de ambos subgrupos no son iguales, basaremos el presente análisis comparativo en los promedios de respuestas que alcanzan dichos grupos en cuanto a

producción de palabras, así como en los valores alcanzados por los dos tipos de informantes en el índice de cohesión (ICOH). El promedio de palabras nos permite establecer la capacidad léxica de ambos tipos de informantes, medible a partir de la producción de palabras de cada uno de ellos. El índice de cohesión por su parte, nos permite conocer cuál de los dos grupos presenta una mayor coincidencia en las respuestas aportadas en cada centro de interés, o si por el contrario la condición de estudiante inmigrante no influye en el carácter abierto o cerrado de los centros de interés.

### 6.1. Frecuencia media de respuestas (FMR): comparativa entre ambas submuestras

La frecuencia media de respuesta (FMR) alcanzada por los estudiantes no inmigrantes (240,12) es superior que la FMR de la submuestra de estudiantes inmigrantes (165,32). De igual manera, la FMR de cada uno de los centros de interés es superior en la submuestra de estudiantes de español no inmigrantes frente a los valores obtenidos en la submuestra de inmigrantes.

Como se observa en la Tabla 7, la FMR de ambas muestras presenta una idéntica correspondencia en las posiciones 4.<sup>a</sup> y 7.<sup>a</sup> (centros 18 y 16 respectivamente). Igual situación se presenta con los centros ‘Servicios, documentos, trámites o lugares necesarios para permanecer en España’, ‘Iluminación, calefacción y ventilación’ y ‘Trabajos del campo y del jardín’, que se ubican en las posiciones 18.<sup>a</sup>, 19.<sup>a</sup> y 20.<sup>a</sup>, tanto en la distribución de acuerdo a la frecuencia media de respuesta, como en el rango que alcanzan los centros en función del número de palabras.

El Gráfico 4 presenta la distribución de los centros de interés respecto a la frecuencia media de respuesta (FMR), tanto para la submuestra de estudiantes de español no inmigrantes como en la submuestra de estudiantes inmigrantes. Como hemos mencionado más arriba, los estudiantes de ELE presentan una mayor FMR en todos los centros y posiciones que los estudiantes inmigrantes. Igualmente, en el Gráfico 25 se refleja la FMR alcanzada en cada centro, tanto en la submuestra estudiantes de ELE (no inmigrantes) como en la submuestra de estudiantes INM.

Estudiantes no inmigrantes			Estudiantes inmigrantes		
Rango	Centros de interés	FMR	Rango	Centros de interés	FMR
1	05. Comidas y bebidas	22,24	1	01. Partes del cuerpo	13,34
2	10. La ciudad	17,79	2	05. Comidas y bebidas	13,05
3	01. Partes del cuerpo	17,19	3	10. La ciudad	10,59
4	18. Adjetivos	16,94	4	18. Adjetivos	9,99
5	20. España	15,6	5	14. Los animales	9,85
6	14. Los animales	14,67	6	02. La ropa	9,33
7	16. Profesiones y oficios	13,38	7	16. Profesiones y oficios	9,04
8	02. La ropa	13,22	8	17. Los colores	8,93
9	17. Los colores	11,80	9	20. España	8,46
10	08. La escuela	11,51	10	04. Los muebles de la casa	8,24
11	15. Juegos y distracciones	11,48	11	08. La escuela	8,24
12	11. El campo	10,43	12	03. La casa	8,11
13	03. La casa	9,84	13	12. Medios de transporte	7,18
14	12. Medios de transporte	9,61	14	15. Juegos	6,78
15	04. Los muebles de la casa	9,57	15	11. Campo	6,73
16	07. La cocina	8,37	16	06. Objetos colocados en la mesa	6,45
17	06. Objetos colocados en la mesa	8,03	17	07. Cocina	6,44
18	19. Servicios, documentos ...	7,74	18	19. Servicios, documentos ...	6,37
19	09. Iluminación, calefacción...	5,43	19	09. Iluminación, calefacción...	4,52
20	13. Trabajos campo	5,28	20	13. Trabajos campo	3,67

*Tabla 7. Comparativa entre el rango que alcanzan los centros de interés respecto a la frecuencia media de respuesta (FMR) (submuestra de estudiantes no inmigrantes y submuestra de estudiantes inmigrantes)*

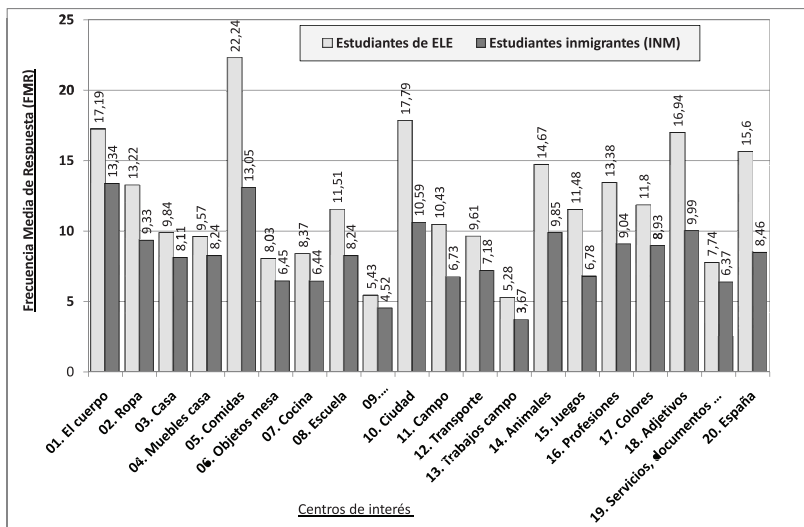


Gráfico 4. Frecuencia media de respuestas (FMR) (comparativa entre ambas submuestras)

## 6.2. Índice de cohesión (ICOH): comparativa entre ambas submuestras

Al comparar los resultados obtenidos en la submuestra de estudiantes de español no inmigrantes con los del grupo de estudiantes inmigrantes respecto al índice de cohesión (ICOH), observamos que a excepción del centro ‘Los muebles de la casa’, que se ubica en la posición 6.<sup>a</sup> en ambos listados, no existen coincidencias entre los rangos obtenidos por los centros de interés en una y otra muestra en función del ICOH.

Como se observa en la Tabla 8, el centro ‘Partes del cuerpo’, que según el índice de cohesión ocupa la 1.<sup>a</sup> posición en la submuestra de estudiantes no inmigrantes, pasa a ocupar el segundo lugar en la submuestra de estudiantes inmigrantes y viceversa: el centro ‘Los colores’, que se ubica en la segunda posición en la submuestra de estudiantes de español no inmigrantes, es el que presenta el mayor ICOH en la submuestra de estudiantes inmigrantes. De la misma manera, los demás centros varían el rango que ocupan de acuerdo al ICOH tanto en la muestra de estudiantes no inmigrantes como en la de estudiantes INM, desplazándose una, dos, tres y hasta cinco posiciones de un listado a otro, como es el caso de ‘Medios de transporte’, que pasa de la 3.<sup>a</sup> posición en la muestra de los estudiantes inmigrantes, a la 8.<sup>a</sup> en la muestra de estudiantes no inmigrantes.

Estudiantes no inmigrantes			Estudiantes inmigrantes		
Rango	Centros de interés	ICOH	Rango	Centros de interés	ICOH
1	01. Parte del cuerpo	0,143	1	17. Los colores	0,172
2	17. Los colores	0,105	2	01. Parte del cuerpo	0,165
3	03. La casa	0,083	3	12. Medios de transporte	0,136
4	02. La ropa	0,080	4	03. La casa	0,121
5	14. Los animales	0,069	5	02. La ropa	0,111
6	04. Los muebles de la casa	0,059	6	04. Los muebles de la casa	0,092
7	05. Comidas y bebidas	0,059	7	14. Los animales	0,088
8	12. Medios de transporte	0,059	8	06. Objetos colocados en la mesa ...	0,085
9	06. Objetos colocados en la mesa ...	0,056	9	05. Comidas y bebidas	0,084
10	10. La ciudad	0,047	10	07. La cocina	0,075
11	07. La cocina	0,040	11	08. La escuela	0,074
12	08. La escuela	0,038	12	10. La ciudad	0,067
13	16. Profesiones y oficios	0,038	13	09. Iluminación, calefacción ...	0,066
14	09. Iluminación, calefacción ...	0,037	14	11. El campo	0,056
15	11. El campo	0,037	15	18. Adjetivos	0,053
16	18. Adjetivos	0,035	16	19. Servicios, documentos ...	0,051
17	15. Juegos y distracciones	0,028	17	16. Profesiones y oficios	0,049
18	20. España	0,024	18	15. Juegos y distracciones	0,039
19	19. Servicios, documentos ...	0,023	19	13. Trabajos del campo ...	0,038
20	13. Trabajos del campo ...	0,020	20	20. España	0,035

Tabla 8. *Rango que alcanzan los centros de interés en función del índice de cohesión (ICOH) (comparativa entre las submuestras de estudiantes no inmigrantes y estudiantes inmigrantes)*

Por otro lado, en el Gráfico 5 se muestra la distribución que alcanzan los centros de interés en función del índice de cohesión (ICOH), tanto en la submuestra de estudiantes no inmigrantes como en la de estudiantes inmigrantes (INM). A pesar de que tanto el número de vocablos como el tamaño de los primeros son superiores a los de la submuestra INM, los índices de cohesión de este último grupo son superiores en todos los centros y posiciones.

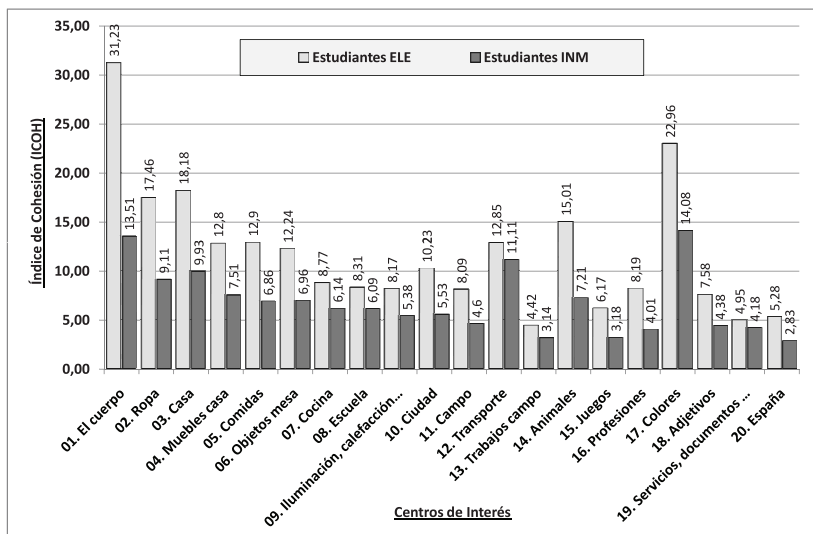


Gráfico 5. Distribución de los centros de interés respecto al índice de cohesión (ICOH) (comparativa entre ambas submuestras)

## 7. Conclusiones

De acuerdo a las variables sociológicas utilizadas para la estratificación de la muestra, el perfil del estudiante inmigrante de nuestra investigación es un hombre de entre 20 y 34 años, con una lengua eslava, que presenta un nivel intermedio de español y se encuentra en España estudiando español por razones principalmente de estudio. Igualmente, es un inmigrante que conoce otras lenguas extranjeras a demás del español y con un nivel sociocultural medio<sup>16</sup>.

Después de analizar los resultados obtenidos por los estudiantes inmigrantes de español en las pruebas de disponibilidad léxica, es interesante observar que los centros de interés donde mayor cantidad de palabras producen no son aquellos que tienen una relación directa con su condición de inmigrantes. Una de las razones para introducir los centros de interés 19 y 20 era precisamente poder conocer el léxico particular de este tipo de informante, donde asumíamos que iba a ser mayor su producción léxica. Sin embargo, contrario a nuestra hipótesis inicial, los centros donde los inmigrantes han actualizado un mayor número de palabras son ‘Partes del cuerpo’ (1.094 palabras) y ‘Comidas y bebidas’ (1.070), mientras que el centro ‘España’, ocupa la 9.ª posición con 694 palabras. El centro ‘Servicios, documentos, trámites o lugares necesarios



para permanecer en España' por su parte, se sitúa en la antepenúltima posición, con 522 palabras, solo superado por los centros 'Iluminación, calefacción y ventilación' y 'Trabajos del campo y del jardín' que por lo general ocupan las últimas posiciones en las investigaciones sobre disponibilidad léxica consultadas.

Si nos fijamos en el número de vocablos producidos por los estudiantes inmigrantes de español, observamos una situación bastante distinta. El centro de interés con mayor número de vocablos es precisamente 'España' (245 vocablos), mientras que 'Servicios, documentos, trámites o lugares necesarios para permanecer en España' ocupa la 7.<sup>a</sup> posición con 125 vocablos. A pesar de que este centro -que como hemos dicho anteriormente esperábamos que ocupara puestos significativos dentro de la producción léxica de los inmigrantes-, se sitúa en posiciones no demasiado relevantes, es importante señalar que se encuentra por encima de la media en producción de vocablos por centro de interés, que es de 116,35 vocablos.

Por otra parte, al comparar la frecuencia media de respuestas (FMR) de los estudiantes de ELE y de los estudiantes inmigrantes, encontramos que los primeros presentan un promedio mucho mayor respecto al alcanzado por los estudiantes inmigrantes de español (240,14 y 165,32 palabras por informante respectivamente). Esta diferencia, de más de 74 palabras, entre los promedios de palabras de ambos grupos de informantes confirma nuestra hipótesis según la cual los estudiantes de ELE producen un mayor número de palabras que los estudiantes INM, debido a que han seguido un proceso continuado de aprendizaje que favorece su competencia léxica. De igual manera observamos que los estudiantes de ELE presentan un mayor promedio de vocablos, lo cual significa que tienen una mayor riqueza léxica que los estudiantes INM.

Al observar el índice de cohesión (ICOH) alcanzado por ambos grupos de informantes, encontramos un comportamiento similar; mientras que los centros más compactos (o de mayor coincidencia en las respuestas de los informantes) para los estudiantes de ELE son el 'Partes del cuerpo' y 'Los colores', para los estudiantes inmigrantes son estos mismos centros pero alternando la posición. De igual manera encontramos que los centros más difusos o con menor coincidencia en las respuestas de los informantes son el 15, el 20, el 19 y el 13 para la submuestra de estudiantes de ELE, mientras que para los estudiantes inmigrantes de español son el 16, el 15, el 13 y el 20. Lo anterior nos permite concluir que la condición de inmigrante no parece ser determinante del carácter compacto o difuso de los centros de interés, medible a partir del cálculo del ICOH.

Para finalizar, después de realizar un análisis exhaustivo de los resultados obtenidos en las pruebas de disponibilidad léxica, hemos determinado que existe una relación entre el número de palabras producidas por

los estudiantes inmigrantes en cada uno de los centros de interés, con la naturaleza propia de cada área temática. Es decir, que los centros de interés que presentan una mayor producción léxica (como por ejemplo ‘Partes del cuerpo’, ‘Comidas y bebidas’ y ‘La ciudad’) se refieren a realidades próximas o conocidas para los informantes y por tanto les resultan más “familiares”, cuestión que se refleja en el alto número de palabras mencionadas en cada uno de ellos. Por el contrario, aquellos centros en los que hay una baja producción de palabras (‘Servicios, documentos, trámites o lugares necesarios para permanecer en España’, ‘Iluminación, calefacción y ventilación’ y ‘Trabajos del campo y del jardín’) constituyen áreas temáticas poco “familiares” que se refieren a realidades lejanas o de poco uso para los estudiantes extranjeros de español<sup>17</sup>.

Por otro lado, consideramos que existe una relación entre la mayor o menor producción de vocablos en cada centro de interés, y las posibilidades asociativas que presentan unos y otros centros. Aquellos centros que presentan un menor número de vocablos (menor riqueza léxica), como es el caso ‘La casa’, ‘Medios de transporte’ y ‘Los colores’, son áreas temáticas que presentan un número más reducido de posibilidades de actualización de vocablos que las que pueden ofrecer otras áreas temáticas (como por ejemplo el centro ‘España’, ‘Adjetivos’ o ‘Juegos y distracciones’), ya que son campos asociativos más restringidos. Según los resultados obtenidos a través de las pruebas de disponibilidad, observamos que las posibilidades de actualización de vocablos en estos centros, están restringidas a las posibilidades de asociación interna que presenta el área temática que sirve de estímulo para la actualización de vocablos.

Diego Javier Gallego Gallego  
Departamento de Filología, Comunicación y Documentación  
Facultad de Filosofía y Letras  
Universidad de Alcalá de Henares  
28801 – Alcalá de Henares (Madrid) – España  
dijaga@hotmail.com

Recepción: 08/05/2014; Aceptación: 27/08/2014

## Notas

- <sup>1</sup> Este trabajo se inscribe en el proyecto de investigación “Patrones sociolingüísticos y procesos de integración sociolingüística en Madrid” (Ref. FFI2011-29189-C05-02), financiado por el Ministerio de Ciencia e Innovación de España.
- <sup>2</sup> En el capítulo introductorio a la *Variación social del léxico disponible en la ciudad de Málaga*, Ávila Muñoz (2010) hace un recorrido sobre los estudios estadísticos del léxico, mostrando las principales diferencias entre las investigaciones basadas en el índice

de frecuencia y las que utilizan el índice de disponibilidad.

- <sup>3</sup> Formaban parte del *Centre d'Etude du Français élémentaire*, Georges Gougenheim (director), Paul Rivenc (subdirector adjunto), René Michéa y Aurélien Sauvegeot.
- <sup>4</sup> En el mismo artículo, Michéa (1953) establece no solo un nuevo concepto dentro de los estudios de léxico-estadística -el de la disponibilidad-, sino que hace una serie de reflexiones en cuanto a las ventajas que este presenta frente a los habituales estudios de léxico basados en la frecuencia.
- <sup>5</sup> Este mismo concepto es retomado más adelante por López Morales en el *Léxico disponible de Puerto Rico*, donde explica que “la disponibilidad léxica pasó a entenderse como el caudal léxico utilizable en una situación comunicativa dada” (1999: 11).
- <sup>6</sup> A diferencia de otros trabajos sobre disponibilidad léxica que trabajan con población inmigrante (por ejemplo la investigación desarrollada por Fernández-Merino Gutiérrez 2011), en nuestro estudio no hemos tenido en cuenta, además de los factores sociales explicados en el apartado 3 del presente artículo, otras cuestiones sociológicas que nos permitan establecer una clasificación del grupo de informantes inmigrantes. Según Andión (2006), los inmigrantes en España se pueden clasificar en cuatro grupos, atendiendo a diversos factores como el país de procedencia, el status económico, la condición legal o los motivos por los que se encuentran en territorio español. Sin embargo, en nuestro caso no es posible establecer una caracterización similar dado que los objetivos de nuestra investigación se centran en conocer el léxico disponible de los estudiantes extranjeros de español (incluyendo a los estudiantes inmigrantes), más que en la descripción sociológica de la población estudiada.
- <sup>7</sup> El propósito de conservar los 16 centros de interés “tradicionales” es permitir la comparación con las investigaciones hispánicas sobre disponibilidad léxica tanto en LM como en LE que utilizan dichos centros. Estos centros son: 01. ‘Las partes del cuerpo’, 02. ‘La ropa’, 03. ‘La casa’ (sin los muebles), 04. ‘Los muebles de la casa’, 05. ‘Comidas y bebidas’, 06. ‘Objetos colocados sobre la mesa para la comida’, 07. ‘La cocina (muebles y utensilios)’, 08. ‘La escuela (muebles y materiales)’, 09. ‘Iluminación, calefacción y ventilación’, 10. ‘La ciudad’, 11. ‘El campo’, 12. ‘Medios de transporte’, 13. ‘Trabajos del campo y del jardín’, 14. ‘Los animales’, 15. ‘Juegos y distracciones’ y 16. ‘Profesiones y oficios’.
- <sup>8</sup> En su investigación sobre la disponibilidad léxica de inmigrantes, Fernández-Merino Gutiérrez (2011), utiliza un centro de interés similar llamado ‘Administración pública: tramitar papeles y hacer gestiones’.
- <sup>9</sup> Consideramos que junto a la “edad”, la motivación es uno de los factores personales que pueden resultar más decisivos dentro del proceso de adquisición y aprendizaje de segundas lenguas y/o lenguas extranjeras. Es por esto que hemos decidido analizar si la motivación que tienen los informantes que estudian español únicamente por razones de estudio, y la motivación que tienen los informantes inmigrantes (que estudian español movidos por intereses distintos como conseguir trabajo, por requisitos administrativos o porque desean integrarse socialmente a través de la lengua), influye de manera significativa en su producción léxica.
- <sup>10</sup> Lengua sino-tibetana.
- <sup>11</sup> Es una lengua asiática de difícil clasificación pues comparte rasgos que la acercan a las lenguas drávidas, austroasiáticas o paleo-siberianas. Tiene además un gran parentesco no probado con el japonés, con el que comparte rasgos gramaticales muy similares. Según Del Moral Aguilera (2002: 151) esta lengua es considerada altaica.
- <sup>12</sup> Es una lengua asiática aislada que es considerada altaica, aunque no hay acuerdo en ello (Del Moral Aguilera 2002: 266). Comparte un léxico común de origen chino con el coreano, pero se alejan en los demás rasgos.
- <sup>13</sup> Según Del Moral Aguilera, el taiwanés es una “familia de lenguas extendidas por la isla de Taiwán que junto con la familia malayo-polinesia, mucho más extendida, forma la gran familia de lenguas austronésicas” (2002: 424).

- <sup>14</sup> La frecuencia media de respuestas (FMR) utilizada en el presente estudio es equivalente al promedio de respuestas por informante de otras investigaciones.
- <sup>15</sup> Llama la atención el poco conocimiento léxico que poseen los estudiantes inmigrantes en el centro ‘Servicios, documentos, trámites o lugares necesarios para permanecer en España’, donde se esperaba una mayor producción léxica de este colectivo de informantes. Una posible explicación del bajo número de respuestas en este centro podría deberse a la situación de legalidad o ilegalidad en la que se encontraban o en el tiempo de permanencia en España que estos inmigrantes llevaban en el momento de realización de la prueba. En el caso de estar ilegalmente en España o de llevar poco tiempo en el país, es posible que no tuvieran los conocimientos básicos necesarios sobre los documentos o vías de legalizar su situación; también es posible que estando legalmente en territorio español, hayan delegado en terceras personas el trámite de su situación legal, lo que explicaría el bajo nivel de respuestas en este centro. Sea como fuere, al no indagar sobre estos aspectos en las pruebas de disponibilidad, todas estas posibles explicaciones no dejan de ser conjeturas, por lo que nos limitamos a mostrar los resultados alcanzados por los informantes en cada centro. Por otra parte, es significativo que tanto el centro de interés ‘Administración pública: tramitar papeles y hacer gestiones’ utilizado en la investigación de Fernández-Merino Gutiérrez (2011), y el centro 19 de nuestra investigación con el que guarda gran similitud, ocupen posiciones inferiores en cuanto a producción de palabras (último y trasantepenúltimo lugar respectivamente), ya que ambos centros están relacionados específicamente con el tipo de informante utilizado en ambas investigaciones.
- <sup>16</sup> Este perfil, aunque se adapta en parte al perfil que presentan los inmigrantes en general, no es del todo equivalente. Según los últimos datos que se conocen del ENI (Encuesta Nacional de Inmigrantes 2007), el perfil del inmigrante en España es un varón con una edad media de 38,4 años, cuya motivación para venir a España es la búsqueda de un empleo mejor y que tiene una lengua materna indoeuropea (que ocupan el segundo lugar después de los inmigrantes hispanos).
- <sup>17</sup> A pesar de que en algunos casos la baja producción de palabras en los centros de interés está justificada por el poco conocimiento léxico que tienen los estudiantes inmigrantes, consideramos que la misma naturaleza que presentan los centros de más baja producción de palabras (‘Servicios, documentos, trámites o lugares necesarios para permanecer en España’, ‘Iluminación, calefacción y ventilación’, y ‘Trabajos del campo y del jardín’), parece influir de manera significativa en la poca productividad léxica. Es por esto que consideramos que dentro de los 20 centros de interés tenidos en cuenta en nuestra investigación, hay algunos que resultan más cercanos o conocidos para los informantes, lo que se refleja en una mayor producción léxica. Por el contrario, existen otras áreas temáticas o centros de interés que resultan menos accesibles o desconocidos para los informantes, lo cual incide en la baja producción léxica.

## Referencias bibliográficas

- Andión Herrero, María Antonieta. 2006. “La inmigración no hablante de español en España: situación y características”. *Revista de lingüística y lenguas aplicadas*, 1. 7-24.
- Ávila Muñoz, Antonio Manuel. 2006. *Léxico disponible de los estudiantes preuniversitarios de Málaga*. Málaga: Universidad de Málaga.
- Ávila Muñoz, Antonio Manuel. 2010. “Aproximación a los estudios estadísticos del léxico. Teoría y principios”. En *Variación social del léxico disponible en la ciudad de Málaga: diccionarios y análisis*, Ávila Muñoz y Villena Ponsoda (eds.), 15-34. Málaga: Editorial Sarriá, S. L.
- Bartol Hernández, José A. y Natividad Hernández. 2003. *Dispolex*. Disponible en <<http://www.dispolex.com/>>.

- Carcedo González, Alberto. 2000. *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Turun Yliopisto.
- Del Moral Aguilera, Rafael. 2002. *Diccionario Espasa lenguas del mundo*. Madrid: Espasa Calpe, S. A.
- Fernández-Merino Gutierrez, Pablo V. 2011. "Disponibilidad léxica de inmigrantes: propuesta para una necesidad". *Lengua y migración / Language and Migration*. 3: 2. 83-105.
- Gallego Gallego, Diego Javier. 2011. "Disponibilidad léxica en estudiantes de español como lengua extranjera en la Comunidad de Madrid. Aspectos metodológicos y primeros resultados". En *Documentos para el XVI Congreso Internacional de la ALFAL*, Ana María Cestero Mancera, Isabel Molina Martos y Florentino Paredes García (eds.). Alcalá de Henares: Universidad de Alcalá (Edición en CD).
- Gougenheim, Georges, René Michéa, Paul Rivenc y Aurélien Sauvageot. 1956. *L'elaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de bas*. Paris: Didier.
- Gougenheim, Georges, René Michéa, Paul Rivenc y Aurélien Sauvageot. 1964. *L'elaboration du français fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de bas*. Paris: Didier.
- López Chávez, Juan y Carlos Strassburguer Frías. 1987. "Otro cálculo del índice de disponibilidad léxica". En *Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada, Presente y perspectiva de la lingüística computacional en México*. México: Universidad Nacional Autónoma de México.
- López Chávez, Juan y Carlos Strassburguer Frías. 1991. "Un modelo para el cálculo del índice de disponibilidad léxica individual". En *La enseñanza del español como lengua materna: Actas del II Seminario sobre "Aportes de la lingüística a la enseñanza de la lengua materna"*, Humberto López Morales (ed.), 91-112. Río Piedras: Universidad de Puerto Rico.
- López Chávez, Juan. 1994. "Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica". *REALE*, 1. 67-84.
- López Morales, Humberto. 1999. *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.
- Michéa, René. 1950. "Vocabulaire et culture". *Les langues modernes*, 44. 188-189.
- Michéa, René. 1953. "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage". *Les langues modernes*, 47. 338-344.
- Moreno Fernández, Francisco, José Enrique Moreno Fernández y Antonio J. García De Las Heras. 1995. "Cálculo de disponibilidad léxica. El programa Lexidisp". *Lingüística*, 7. 243-249.
- Reher, David-Sven et al. 2008. *Informe Encuesta Nacional de Inmigrantes (ENI 2007)*. Madrid: Instituto Nacional de Estadística.
- Samper Hernández, Marta. 2002. *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: Asele.