

## ASPECTOS SOCIOLINGÜÍSTICOS DEL LÉXICO DISPONIBLE CASTELLANO DE LOS PREUNIVERSITARIOS LERIDANOS\*

Serrano Zapata, Maribel

*Departamento de Filología Clásica, Francesa e Hispánica  
Universitat de Lleida  
Plaza Víctor Siurana, 1  
25003 Lleida (España)  
Tel: (0034) 973702077  
e-mail: mserrano@filcef.udl.es*

(Recibido julio 2004; aceptado septiembre 2004)

BIBLID [1133-682X (2004) 12: 147-165]

### Resumen

En este artículo presentamos los resultados que hemos obtenido al realizar una investigación de disponibilidad léxica en la ciudad de Lérida. Tras haber aplicado la prueba a un total de 48 estudiantes preuniversitarios de tres centros educativos de la ciudad, hemos analizado cuantitativamente los datos obtenidos en función de las distintas variables sociales contempladas. Las conclusiones a las que hemos llegado podrán corroborarse en futuros trabajos cuando incrementemos el volumen de la muestra manejada.

**Palabras clave:** lingüística aplicada, sociolingüística, disponibilidad léxica, léxico-estadística, variables sociales

### Abstract

In this article we present the results that we have obtained researching lexical availability in the city of Lérida. After applying the test to 48 preuniversity students from three educative centres of the city, we have analysed quantitatively the facts obtained according to the different variables contemplated. The conclusions that we have reached can be strengthened in future studies when we increase the volume of the sample managed.

**Key words:** applied linguistics, sociolinguistics, lexical availability, lexical statistics, social variables

### Resumé

Dans cette étude nous présentons les résultats obtenus en recherchant la disponibilité lexicale dans la ville de Lérida. Depuis d'avoir passer le test à 48 étudiants preuniversitaires de trois centres éducatifs de la ville, nous avons analysé quantitativement les données obtenues selon les différentes variables considérées. Les conclusions actuelles pourront être corroborées en futures études quand on augmente le chiffre de l'échantillon utilisé.

**Mots clé:** linguistique appliquée, sociolinguistique, disponibilité lexical, lexico-statistique, variables sociales.

### Sumario

0. Introducción. 1. Metodología. 2. Análisis cuantitativo. 2.1. Resultados generales. 2.2. Los centros de interés. 2.3. Comparación con otras sintopías del mundo hispánico. 2.4. Las variables sociales. 3. Conclusiones. 4. Referencias.

---

\* Este trabajo ha podido realizarse gracias al apoyo del Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Cataluña.

## 0. Introducción

A pesar de que la disponibilidad léxica nació en Francia a mediados de los años cincuenta<sup>1</sup>, esta línea de investigación tiene, en la actualidad, un gran auge en la lingüística hispánica. Concretamente, Humberto López Morales coordina y dirige desde principios de los años noventa un amplio proyecto que tiene como objetivo elaborar diccionarios de disponibilidad léxica para las distintas zonas del mundo hispanohablante<sup>2</sup>. Este tipo de estudios –que se enmarca dentro de la léxico-estadística–, contrariamente a otras clases de recuentos de palabras, pretende establecer el vocabulario que está presente en la mente de los hablantes, pero que solo se actualiza si la situación comunicativa así lo requiere<sup>3</sup>. Los resultados de las investigaciones sobre el léxico disponible de una determinada comunidad de habla pueden tener aplicaciones en campos muy diversos, como la dialectología, la psicolingüística o la etnolingüística<sup>4</sup>. Sin embargo, es el enfoque sociolingüístico el que ha imperado en la mayor parte de los estudios realizados hasta el momento<sup>5</sup>. El trabajo que aquí presentamos es una prueba más de ello por cuanto intentamos mostrar la relación existente entre el léxico disponible de una muestra de informantes de la ciudad de Lérida y una serie de variables sociales.

## 1. Metodología

La muestra que hemos utilizado en esta investigación de carácter exploratorio está compuesta por 48 estudiantes preuniversitarios procedentes de tres centros de enseñanza secundaria ubicados en el núcleo urbano de la ciudad de Lérida<sup>6</sup>. La edad de los informantes y el nivel educativo en cuestión –2º de bachillerato, el antiguo C.O.U.– se han ajustado a las pautas establecidas por los componentes del proyecto panhispánico, basándose en el convencimiento de que, de esta forma, se plasma fielmente la norma léxica culta no especializada de una comunidad de habla<sup>7</sup>.

---

<sup>1</sup> En 1956 se publicó la obra pionera de Gougenheim *et al.* En este trabajo citamos la edición revisada de 1964.

<sup>2</sup> Toda la información sobre las pautas del proyecto, así como un breve panorama de los estudios realizados hasta el momento, puede consultarse en [www3.usal/dispolex](http://www3.usal/dispolex) o [www.dispolex.com](http://www.dispolex.com).

<sup>3</sup> El creador de la noción de léxico disponible fue René Michéa, quien la expuso con detalle en su decisivo artículo de 1953: “Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé et se présente immédiatement et naturellement à l'esprit au moment où l'on en a besoin” (Michéa, 1953: 340).

<sup>4</sup> Véase un panorama de las distintas aplicaciones que ha tenido este tipo de estudios en López Morales (1995-1996, 1999: 20-26), Carcedo (1998: 53-63) y Samper, Bellón y Samper Hernández (2003: 105-123).

<sup>5</sup> De hecho, este enfoque ya estaba presente de algún modo, en los primeros estudios franceses (Gougenheim *et al.* 1964: 163-182).

<sup>6</sup> Este trabajo forma parte de una investigación más amplia que cuenta con los datos de 240 informantes y que tiene como objetivo describir y comparar el léxico disponible en las dos lenguas oficiales –castellano y catalán– de la provincia de Lérida. Las características metodológicas de esta investigación mayor pueden verse en Serrano Zapata (2003).

<sup>7</sup> Hay otra razón implícita: la facilidad metodológica que conlleva este criterio puesto que permite recabar datos de muchos informantes al mismo tiempo, como ya reconocieron Gougenheim *et al.*: “Nous avons choisi les écoliers parce qu'eux seuls fournissaient des témoins en nombre suffisant” (1964: 153).

En cuanto a las variables sociológicas que hemos tenido en cuenta, estas han sido algunas de las habitualmente examinadas en este tipo de estudios, como el *sexo*, el *tipo de centro educativo* o el *nivel sociocultural*, a las que hemos añadido otra, la *lengua materna*, que solo incluyen, lógicamente, los trabajos realizados en territorios bilingües. De acuerdo con estas variables, la muestra ha quedado estratificada sociológicamente de la siguiente forma:

- a) Sexo: varón, 20; mujer, 28.
- b) Tipo de centro: público, 24; privado, 24.
- c) Nivel sociocultural: bajo, 18; medio, 25; alto, 5<sup>8</sup>.
- d) Lengua materna: castellano, 19; catalán, 29.

Debemos señalar que el nivel sociocultural ha sido determinado teniendo en cuenta la profesión y los estudios del padre y de la madre así como el nivel cultural de la familia. Este último parámetro ha sido establecido mediante las respuestas del informante a una serie de preguntas que incorporamos en la encuesta encaminadas a conocer el interés general de la familia por la cultura. A pesar de que los estudios de disponibilidad léxica han advertido de forma categórica sobre la importancia del factor sociocultural, aún no se ha llegado a un acuerdo unánime sobre la forma de establecerlo. Algunas de las investigaciones llevadas a cabo en Hispanoamérica (López Morales 1999 o Alba 1995) tuvieron en cuenta tanto los estudios y la profesión de los padres como los ingresos familiares. En general, la mayoría de las investigaciones españolas ha seguido muy de cerca la clasificación establecida por Samper y Hernández (1997: 230) y ha considerado solo dos parámetros: la profesión y el nivel educativo alcanzado por los progenitores. En otras ocasiones (Benítez Pérez 1992 o Blas Arroyo y Casanova 2003), el nivel sociocultural se ha establecido basándose únicamente en un parámetro, a saber, el nivel educativo de los padres.

En nuestro caso, consideramos los dos factores principalmente estudiados en las investigaciones hispánicas (las profesiones y los estudios de los progenitores) junto con este parámetro nuevo al que denominamos “nivel cultural”, habida cuenta de que, en nuestra comunidad de habla, no siempre un nivel de estudios elevado y una profesión prestigiosa equivalen a un nivel cultural alto. Otorgamos a los tres parámetros el mismo valor y de este modo dividimos a los informantes en los tres estratos sociales.

El mecanismo de obtención de los datos fue el siguiente. En un primer momento, nos pusimos en contacto con los centros educativos seleccionados<sup>9</sup> y se nos concedió una hora lectiva para poder aplicar la encuesta a un grupo de 2º de bachillerato elegido al azar. Una vez en el aula, entregamos a cada estudiante un grupo de fotocopias grapadas donde se halla-

---

<sup>8</sup> Hemos procurado contar con informantes de todas las variantes sociales escogidas. Sin embargo, esto no ha sido posible en todos los casos y solo 5 informantes pertenecen al estrato alto del espectro social, lo que, como veremos más adelante, tendrá repercusiones en el análisis.

<sup>9</sup> Estos fueron los centros públicos I.E.S. Joan Oró, I.E.S. Escola del Treball y el colegio privado Lestonnac. Agradecemos desde aquí al profesorado y al alumnado su amable participación en nuestro estudio.

ban dos páginas referentes a las variables sociológicas y los distintos *centros de interés*<sup>10</sup>, distribuidos a razón de uno por página<sup>11</sup>. Recabadas las listas de palabras, estas se sometieron a un riguroso y laborioso proceso de edición<sup>12</sup>. Finalmente, los materiales fueron introducidos en el banco de datos de la página web [www.dispolex.com](http://www.dispolex.com) para su cuantificación matemática.

## 2. Análisis cuantitativo

### 2.1. Resultados generales

En la tabla I podemos apreciar el total de palabras y de vocablos<sup>13</sup> que se han obtenido en los 16 centros de interés y el promedio respectivo de palabras por informante:

Total de palabras	Total de vocablos	Promedio de palabras por informante
14.415	3.244	300,3

Tabla 1. *Resultados globales*

### 2.2. Los centros de interés

Al iniciar la investigación partimos de la hipótesis de que podrían apreciarse diferencias cuantitativas en función del enunciado que se ofreciera como estímulo verbal. Además, los estudios de disponibilidad léxica realizados con anterioridad en otras regiones del ámbito hispánico ya han corroborado esta hipótesis mostrando que la producción léxica difiere según el ámbito de la realidad que se sondee. Seguidamente mostramos los resultados obtenidos atendiendo al centro de interés en función de distintos parámetros.

Centros de interés	PALABRAS	PROMEDIOS
01. El cuerpo humano	1302	27,1
14. Animales	1288	26,8
05. Alimentos y bebidas	1237	25,8

<sup>10</sup> Estos son, con algunas modificaciones mínimas en la formulación, los mismos que utilizaron los lingüistas franceses (Gougenheim *et al.* 1964: 152-153) y los que se estudian en el marco del proyecto panhispánico.

<sup>11</sup> Esta es otra novedad metodológica de nuestra investigación puesto que, habitualmente, se distribuyen tres centros de interés por página. Consideramos que, de esta manera, es muy difícil que se dediquen exactamente dos minutos a cada *centro* puesto que, al terminar el primero, el estudiante ya tiene el segundo al lado para ir escribiendo las palabras o simplemente para ir pensando en él. En cambio, como nuestro sistema exige pasar la hoja para cambiar de centro, de esta forma, se controla mejor la tarea del alumnado. Véase también la opinión de Galloso (2002: 22).

<sup>12</sup> En cuanto a los criterios de edición de las listas, hemos seguido principalmente las decisiones adoptadas por Samper (1998) y Carcedo (2001: 36-51) con pequeñas alteraciones.

<sup>13</sup> La distinción entre palabras y vocablos que utilizamos a lo largo de estas páginas es ya clásica en el ámbito de la disponibilidad. Entendemos por *palabra* toda aquella unidad léxica actualizada por los informantes en la encuesta, mientras que con el término *vocablo* nos referimos únicamente a aquellas palabras que no se repiten.

02. Ropa y complementos	1137	23,7
10. La ciudad	1093	22,8
08. La escuela: muebles y materiales	1082	22,5
16. Profesiones y oficios	970	20,2
11. El campo	925	19,3
12. Medios de transporte	841	17,5
15. Juegos y distracciones	808	16,8
07. La cocina y sus utensilios	784	16,3
06. Objetos colocados en la mesa para la comida	691	14,4
03. Partes de la casa (sin los muebles)	666	13,9
04. Los muebles de la casa	593	12,4
09. Calefacción, iluminación y medios de airear un recinto	558	11,6
13. Trabajos del campo y del jardín	440	9,2

Tabla II. Centros de interés ordenados según el número decreciente de respuestas

Tal y como se observa en la tabla II, las diferencias entre el número total de palabras que cada centro genera son más que notables. Si nos fijamos en los promedios, las 9 palabras que produce de media cada informante ante el estímulo 13. *Trabajos del campo y del jardín* contrastan con las 27 –el triple– que se obtienen ante el primero de los centros de interés. En principio, es de esperar que los encuestados produzcan más palabras en aquellas áreas temáticas que sientan más cercanas, es decir, que guarden relación con sus actividades más habituales y cotidianas. Así podríamos explicar que alcancen una posición tan elevada el centro 01. *El cuerpo humano*, que ostenta el primer rango, o los centros 05. *Alimentos y bebidas* y 02. *Ropa y complementos*, que ocupan la tercera y cuarta posiciones de la tabla, respectivamente. Contrariamente a lo que podría esperarse, no influye para nada el orden en que se rellenan los centros en la productividad de los mismos.

Por otro lado, aquellas áreas léxicas en las que el alumnado ha actualizado menos léxicas son 13. *Trabajos del campo y del jardín* y 09. *Calefacción, iluminación y medios de airear un recinto*. En estos casos creemos que influye la poca familiaridad de un informante urbano de 17 ó 18 años de edad con estas parcelas léxicas<sup>14</sup>. Además, según parece, la mayor o menor concreción de un centro puede influir en el número de palabras. Así, por ejemplo, nótese que los centros 06. *Objetos colocados en la mesa para la comida* o 09. *Calefacción, iluminación y medios de airear un recinto*, cuentan con enunciados que restringen mucho una determinada área semántica.

<sup>14</sup> Idea que compartimos con Samper, Bellón y Samper Hernández (2003: 59).

Centros de interés	VOCABLOS
11. El campo	297
16. Profesiones y oficios	295
10. La ciudad	285
15. Juegos y distracciones	273
05. Alimentos y bebidas	262
14. Animales	223
08. La escuela: muebles y materiales	204
01. El cuerpo humano	195
07. La cocina y sus utensilios	181
02. Ropa y complementos	175
13. Trabajos del campo y del jardín	167
06. Objetos colocados en la mesa para la comida	151
09. Calefacción, iluminación y medios de airear un recinto	140
12. Medios de transporte	138
03. Partes de la casa (sin los muebles)	134
04. Los muebles de la casa	124

Tabla III. *Centros de interés ordenados según el número decreciente de vocablos*

Como hemos constatado anteriormente, la cantidad de vocablos hace referencia al número de palabras diferentes que se han logrado en los diversos centros. Es decir, este índice mide las posibilidades asociativas de cada uno de los enunciados. Lo primero que observamos al comparar esta tabla con la anterior es que los resultados no guardan necesariamente relación entre sí. Si bien es cierto que algunos centros, como el 3. *Partes de la casa (sin los muebles)* y el 4. *Los muebles de la casa*, se sitúan en los últimos puestos de ambas clasificaciones, la mayoría ostenta rangos diferentes. Por ejemplo, el centro 01. *El cuerpo humano*, que tenía el mayor rango concerniente a la productividad, solo alcanza una posición intermedia con respecto al número de vocablos o voces distintas. En cambio, el caso contrario es el del centro 11. *El campo*, que ocupaba una posición media con respecto al número de palabras y, sin embargo, es el que más vocablos aporta<sup>15</sup>. Lo que en realidad ocurre es que el alumnado produce más lexías en el centro 01 que en el centro 11 pero, en el primer caso, hay una gran coincidencia entre las respuestas, y en el segundo, no. Dicho de otro modo, en el centro 11 las palabras que se han aportado son más variadas, en cambio, en el centro 01 el número de vocablos es menor porque las posibilidades asociativas son asimismo muy inferiores o porque la mayoría, si no todos, ha pensado en los mismos conceptos. Otro caso llamativo es el del centro 13. *Trabajos del campo y del jardín*, que ocupa el último lugar en relación con el número de palabras y, no obstante, su posición no es tan baja

<sup>15</sup> Otros investigadores han constatado, con anterioridad, las mismas divergencias (vg. Carcedo 2000: 213).

en la tabla de vocablos. Esto es debido igualmente a que las pocas palabras que se aportan son muy distintas entre sí.

Centros de interés	ÍNDICE DE COHESIÓN
01. El cuerpo humano	0,139
02. Ropa y complementos	0,135
12. Los medios de transporte	0,127
14. Los animales	0,120
08. La escuela: muebles y materiales	0,110
03. Partes de la casa (sin los muebles)	0,104
04. Los muebles de la casa	0,100
05. Alimentos y bebidas	0,098
06. Objetos colocados en la mesa para la comida	0,095
07. La cocina y sus utensilios	0,090
09. Calefacción, iluminación y medios de airear un recinto	0,083
10. La ciudad	0,080
16. Profesiones y oficios	0,068
11. El campo	0,065
15. Juegos y distracciones	0,062
13. Trabajos del campo y del jardín	0,055

Tabla IV. *Centros de interés ordenados según el valor decreciente del índice de cohesión*

Para el estudio de las posibilidades asociativas de los centros de interés utilizamos otro parámetro que tiene en cuenta los promedios: el *índice de cohesión*. Este índice es una fórmula matemática que relaciona el número de palabras y de vocablos con el fin de caracterizar la mayor o menor dispersión del léxico consignado en torno a los centros de interés<sup>16</sup>.

Para obtener el índice de un determinado centro de interés se divide el promedio de respuestas por centro entre el número de vocablos de dicho centro. Si se diera el caso de que todos los sujetos proporcionasen las mismas respuestas, al calcular el índice de cohesión, este nos daría 1. Por tal razón, los centros que se acercan más a este valor y que, por consiguiente, alcanzan las primeras posiciones en la tabla IV son aquellos en los que la coincidencia entre los informantes es mayor, mientras que los centros que producen respuestas más variadas ocupan los últimos lugares de la clasificación.

<sup>16</sup> Echeverría (1991: 62), creador de la fórmula de cálculo índice, afirma al respecto: "nuestro índice de cohesión informa si el centro es cerrado o compacto, o más bien abierto o difuso". Este autor ya utiliza el índice en su trabajo de 1987.

Como ya vimos en la tabla III, cada centro de interés produce un número de vocablos, y por tanto de asociaciones, muy distinto. El índice de cohesión pondera este parámetro con la cantidad de respuestas y nos da una visión de la densidad interna de cada campo semántico. En nuestro caso, el centro 01. *El cuerpo humano* ocupa la primera posición, lo que indica que es el centro más compacto o cerrado de todos, ya que es en el que ha habido un mayor número de coincidencias en relación con las palabras aportadas. Si nos fijamos en la tabla IV, los centros más cohesionados son 01. *El cuerpo humano*, 02. *Ropa y complementos* y 12. *Los medios de transporte*; por el contrario, la mayor difusión la obtienen 13. *Trabajos del campo y del jardín*, 15. *Juegos y distracciones* y 11. *El campo*. Parece que los tres primeros centros ofrecen menos posibilidades asociativas o, como mínimo, los hablantes tienden a realizar siempre las mismas asociaciones. Incluso un centro como el 14. *Animales*, que cuenta con un número alto de palabras, tiene un índice de cohesión bastante elevado con respecto al total. Por el contrario, un enunciado como 10. *La ciudad*, que ocupa la quinta posición empezando por el final, el mencionado 11. *El campo* o el 15. *Juegos y distracciones*, pueden provocar un amplio abanico de asociaciones en el que cabrían infinidad de evocaciones. El caso del centro 13. *Trabajos del campo y del jardín* es especialmente llamativo puesto que ocupa la última posición en las tablas II y IV, lo que indica que los informantes han actualizado pocas palabras pero además muy distintas entre sí. Consideramos que, a pesar de que la provincia de Lérida se caracteriza por una amplia presencia del sector agrícola, el escaso conocimiento del informante sobre el tema del que se pregunta hace que las respuestas sean escasas y muy diversas<sup>17</sup>.

### 3.3. Comparación con otras sintopías del mundo hispánico

Al comenzar este estudio nos preguntamos si los estudiantes leridanos producirían una cantidad de respuestas similar a la del resto de comunidades hispánicas. Pretendíamos comprobar si el hecho de pertenecer a una comunidad en la que conviven dos lenguas, no siendo la habitual el castellano, influía en el conocimiento del léxico castellano. El número de palabras y de vocablos totales son fácilmente alterables y para poder compararlos tendríamos que contar con muestras de igual tamaño. Por esta razón, utilizamos para el cotejo los promedios de respuestas a cada centro de interés. En el gráfico número 1 podemos observar cómo se distribuye el promedio obtenido en Lérida en relación con el general<sup>18</sup>.

---

<sup>17</sup> Confróntese con la opinión de Galloso (2002: 61).

<sup>18</sup> Los datos relativos al promedio general hispánico del gráfico 1 y los de los distintos lugares consignados en el gráfico 2 han sido extraídos del exhaustivo artículo de Samper, Bellón y Samper Hernández (2003: 59-60 y 57). Con respecto al gráfico 2, Samper, Bellón y Samper Hernández (2003: 57) ofrecen los promedios por centro de interés de los que nosotros hemos calculado la media aritmética.

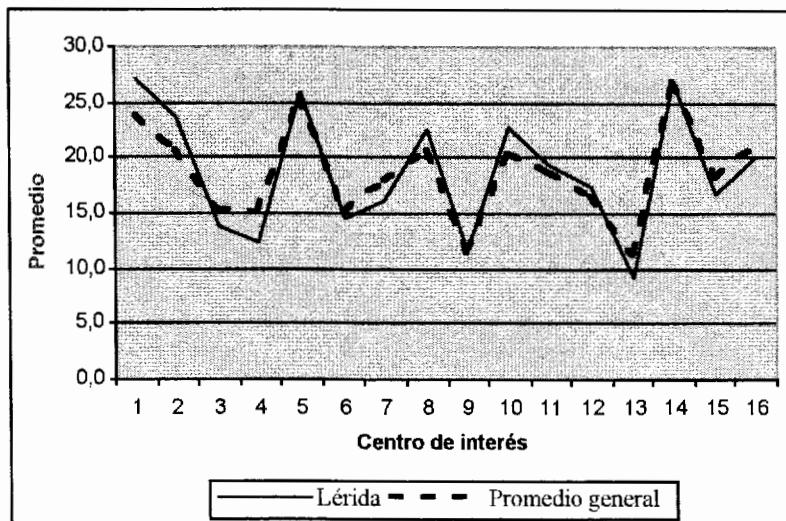


Gráfico 1. *Promedio de respuestas por informante en las investigaciones del proyecto panhispánico y en Lérida*

Tal y como se observa en el gráfico 1, hay un gran paralelismo entre la cantidad de palabras que se logra en nuestra investigación y la correspondiente general del mundo hispánico. Nótese además que, como ya dijimos al analizar el número de palabras, las correspondencias entre los centros más y menos productivos son también más que notables. Nuestros resultados son similares en este punto a los generales, puesto que los centros que en este caso producen un mayor número de respuestas son el 14, el 05 y el 01 y los menos productivos son los centros 04, 09 y 13, lo que coincide, con alguna ligera alteración en el orden, con los resultados obtenidos en Lérida. Si observamos ahora el promedio general de cada sintopía hispánica (gráfico 2), podemos constatar que, en general, nuestros informantes se sitúan más cerca de las últimas posiciones que de las primeras. Sin embargo, el promedio de Lérida (18,76) es muy cercano al de otras regiones monolingües peninsulares como Cádiz (18,57) e incluso supera a otro territorio bilingüe como Bilbao (17,51), y está bastante próximo al de Asturias (19,04), donde se da una situación lingüística peculiar.

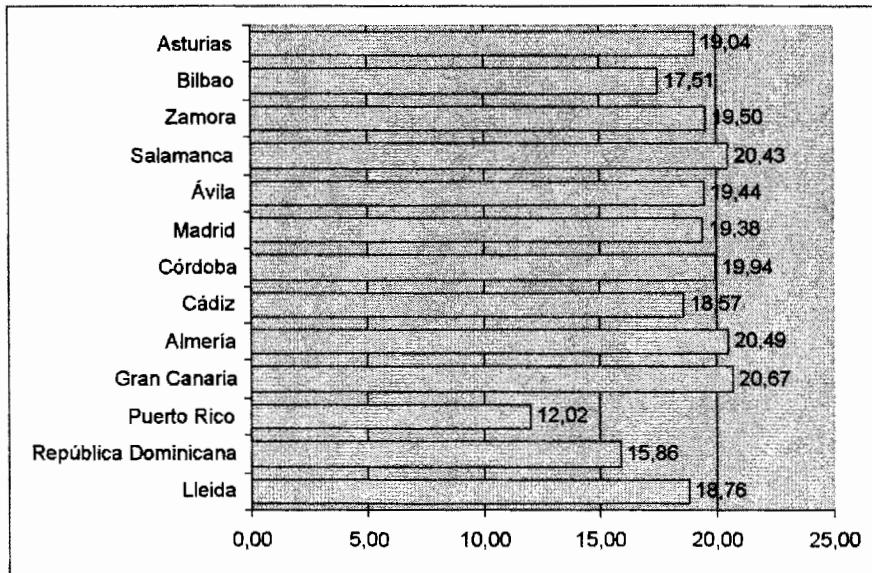


Gráfico 2: Promedios generales de palabras en distintas sintopías del mundo hispánico

## 2.4. Las variables sociales<sup>19</sup>

### Sexo

En la tabla V podemos apreciar la distribución de los promedios en función del sexo de los informantes en cada una de las áreas temáticas.

Centros de interés	VARONES	MUJERES
01. El cuerpo humano	25,5	28,5
02. Ropa y complementos	21,0	25,6
03. Partes de la casa (sin los muebles)	15,0	13,1
04. Los muebles de la casa	10,8	13,5
05. Alimentos y bebidas	23,7	27,5
06. Objetos colocados en la mesa para la comida	13,1	15,4
07. La cocina y sus utensilios	14,8	17,4
08. La escuela: muebles y materiales	20,5	24,0

<sup>19</sup> Los promedios son los únicos índices matemáticos que utilizamos para efectuar la comparación entre las distintas variables. Para poder comparar los totales (por ejemplo, de palabras y de vocablos) deberíamos contar con el mismo número de informantes en cada variable social.

09. Calefacción, iluminación y medios de airear un recinto	11,3	11,9
10. La ciudad	21,2	23,9
11. El campo	19,2	19,3
12. Medios de transporte	17,7	17,4
13. Trabajos del campo y el jardín	8,6	9,6
14. Los animales	26,8	26,9
15. Juegos y distracciones	17,5	16,4
16. Profesiones y oficios	19,8	20,5

Tabla V. *Promedio de respuestas en función del sexo*

En principio, podemos afirmar que las mujeres aventajan a los varones en la producción de palabras. Sin embargo, hay tres centros de interés (3. *Partes de la casa (sin los muebles)*, 15. *Juegos y distracciones* y 12. *Medios de transporte*, por este orden de preferencia) en los que los informantes masculinos consiguen resultados más altos. Con todo, en términos relativos, las estudiantes producen de media 2,1 palabras más que sus respectivos compañeros; por el contrario, en los tres únicos campos léxicos en los que los varones muestran su preponderancia, la media de diferencia es menor, solo de 1,1 puntos. Además, en términos absolutos, la mayor distancia que presentan los varones es de 1,9 puntos por encima de las mujeres en el centro 03. *Partes de la casa*, mientras que el sexo femenino llega a superar hasta en 4,6 puntos el promedio masculino, como se ve en el centro 02. *Ropa y complementos*<sup>20</sup>. Los centros de interés en los que se producen resultados más próximos entre varones y mujeres y que, por consiguiente, no favorecen la diferenciación sexolectal son 11. *El campo*, 14. *Los animales* y 12. *Medios de transporte*, con tan solo unas décimas de distancia.

Estos resultados están en consonancia con los obtenidos en otros estudios. En general, las investigaciones realizadas con anterioridad han rendido cuenta de las ligeras diferencias cuantitativas existentes en favor de uno u otro sexo, aunque no parece haber homogeneidad sobre cuál de los dos es el que ofrece más términos. Así, López Morales (1979), Benítez (1994) y Galloso (2002)<sup>21</sup> hallaron que los informantes masculinos producían más lexías que los femeninos. En cambio, Echeverría *et al.* (1987 y 1991), Alba (1995), Samper y Hernández (1997), González Martínez (2002), Paredes (2001) y Gómez Molina y Gómez

<sup>20</sup> No parece casual que la diferencia mayor se dé justo en esta área temática puesto que, como algunos autores ya han advertido (v.g. Echeverría *et al.* 1987 o Valencia y Echeverría 1999), las discrepancias halladas en torno a algunos centros de interés pueden relacionarse con los esteriotipos de género tradicionales, que todavía persisten en la sociedad española.

<sup>21</sup> Para ser más exactos, esta autora comprueba que los varones superan a las mujeres en dos de las tres provincias objeto de su estudio, concretamente en Zamora y en Salamanca (Galloso 2002: 227-228).

Devís (2004) constataron unos índices más elevados en el colectivo femenino<sup>22</sup>. A pesar de este hecho, la mayoría de los autores reconoce que estas divergencias suelen ser poco notables y revelan antes una homogeneidad que una diferencia verdaderamente significativa<sup>23</sup>.

### Tipo de centro educativo

Centros de interés	PÚBLICOS	PRIVADOS
1. El cuerpo humano	27,0	27,2
2. Ropa y complementos	23,1	24,3
3. Partes de la casa (sin los muebles)	12,0	15,3
4. Los muebles de la casa	12,0	13,3
5. Alimentos y bebidas	25,5	26,0
6. Objetos colocados en la mesa para la comida	14,1	14,7
7. La cocina y sus utensilios	15,7	17,0
8. La escuela: muebles y materiales	21,8	23,3
9. Calefacción, iluminación y medios de airear un recinto	10,8	12,4
10. La ciudad	22,4	23,1
11. El campo	18,2	20,3
12. Medios de transporte	17,1	18,0
13. Trabajos del campo y el jardín	8,6	9,7
14. Los animales	27,0	26,8
15. Juegos y distracciones	17,1	16,6
16. Profesiones y oficios	18,9	21,5

Tabla VI. *Promedio de respuestas en función del tipo de centro educativo*

Tras cotejar los datos de los dos grupos educativos, expuestos en la tabla VI, podemos concluir que el tipo de educación recibida sí parece influir en los resultados puesto que el alumnado procedente del colegio privado logra en general mejores promedios. Como en el caso anterior, la distancia cuantitativa entre ambos grupos es muy pequeña y la media de

<sup>22</sup> Obsérvese que los estudios sociolingüísticos han evidenciado una situación similar en el sentido de que, aun siendo mucho mayores las diferencias detectadas entre varones y mujeres que las que consignan los estudios de disponibilidad léxica, el mismo fenómeno lingüístico (sea debilitamiento de /-s/, sea rehilamiento, sea arrinconamiento del futuro sintético, etc.) en unas comunidades de habla es patrocinado por un sexo y en otras por otro, de modo que no puede delinearse un patrón sexual homogéneo evidente. No obstante, López Morales (1989) llegó a la conclusión de que, por debajo de este comportamiento lingüístico tan dispar, subyacía la mayor conciencia sociolingüística de las mujeres, que las lleva a producir, en cada colectivo, el rasgo que recibe una mejor consideración social.

<sup>23</sup> Mateo (1998: 247) en su investigación almeriense reconoce que el factor sexo ha roto sus expectativas iniciales y que las diferencias encontradas no son sustanciales. También Carcedo (2001: 67) y Bellón (2003: 372) afirman que este factor social no ejerce un poder discriminador sobre la producción del léxico disponible de sus respectivas comunidades.

palabras que actualizan de más los informantes del colegio privado es solo de 1,5 respuestas. Sin embargo, esta ligera ventaja es más constante que en el caso de la variable *sexo*. Aquí, los informantes de los institutos aventajan solo en dos ocasiones a los del colegio, en los centros 14. *Animales* y 15. *Juegos y distracciones*; pero lo hacen con una diferencia casi inapreciable, de 0,3 décimas. Por eso podemos concluir que los informantes procedentes de los colegios privados muestran una mayor disponibilidad léxica que los de los públicos, que los siguen a corta distancia.

También en esta ocasión nuestros resultados van a la par de los obtenidos por el resto de investigadores que trabajan en disponibilidad. A pesar de que las diferencias recabadas son en unos casos mayores y en otros no tan apreciables, los estudios realizados en zonas de Hispanoamérica (Alba 1995; Valencia y Echeverría 1999) y de España (García Domínguez *et al.* 1994; Mateo 1998; Carcedo 2001; González Martínez 2002; Galloso 2002; Bellón 2003; Gómez Molina y Gómez Devís 2004) han mostrado la relevancia de esta variable siempre en favor de los centros educativos privados. En alguna ocasión aislada, también se ha registrado la nula incidencia de este factor (Blas y Casanova 2003).

Además, ninguna investigación ha constatado hasta la fecha la preponderancia de los informantes procedentes de centros educativos públicos. Este hecho podría indicar que la educación impartida en los centros privados es mejor que la de los centros públicos; sin embargo, cabe recordar que, el tipo de educación guarda relación con el nivel sociocultural de la familia, cuya relevancia analizamos a continuación.

### Nivel sociocultural

Los datos relativos a los promedios obtenidos por cada nivel sociocultural son los que se observan en la tabla VII.

<b>Centros de interés</b>	<b>BAJO</b>	<b>MEDIO</b>	<b>ALTO</b>
1. El cuerpo humano	26,3	27,6	28,6
2. Ropa y complementos	22,4	27,6	8,6
3. Partes de la casa (sin los muebles)	11,2	17,6	4,6
4. Los muebles de la casa	11,7	14,5	4,2
5. Alimentos y bebidas	23,4	27,5	25,4
6. Objetos colocados en la mesa para la comida	12,9	17,4	4,8
7. La cocina y sus utensilios	16,0	18,6	6,0
8. La escuela: muebles y materiales	21,2	26,6	7,2
9. Calefacción, iluminación y medios de airear un recinto	10,9	13,7	4,0
10. La ciudad	20,9	27,0	8,2
11. El campo	17,3	23,4	5,4
12. Medios de transporte	15,1	21,5	6,4
13. Trabajos del campo y el jardín	8,4	10,9	3,2

14. Los animales	24,6	31,8	10,2
15. Juegos y distracciones	15,5	19,9	6,4
16. Profesiones y oficios	17,5	24,8	7,2

Tabla VII. *Promedio de respuestas en función del nivel sociocultural*

Lo más sorprendente de estos resultados es, sin duda alguna, los bajos promedios registrados por el nivel sociocultural alto. En general, el estrato más alto del espectro social ha obtenido resultados inferiores incluso a los logrados por el nivel sociocultural bajo. La explicación de este comportamiento inesperado la hallamos en la proporción numérica de la muestra. Como ya vimos en el apartado metodológico, solo 5 de los 48 informantes pertenecen al nivel alto. Por lo tanto, podemos afirmar que esta cantidad no parece ser representativa de este estrato social y que tendríamos que tener una muestra mayor para poder contar con más individuos pertenecientes a este estrato. Además, pueden haberse dado otros condicionantes no controlables, como el rendimiento académico o el cansancio, que hayan tergiversado los resultados de estos informantes. Por todas estas razones, decidimos prescindir en este trabajo de estos datos y, seguidamente, comentamos únicamente las diferencias halladas entre los estratos medio y bajo.

En primer lugar, tenemos que recordar que para la estratificación del nivel sociocultural utilizamos un nuevo índice creado por nosotros. Después de ver los resultados, podemos constatar que la fórmula de los tres indicadores sociales parece ser adecuada para establecer diferencias sociolectales y muestra una vez más la relevancia del condicionante sociocultural en el caudal léxico disponible. De este modo, la superioridad del nivel medio frente al bajo queda evidenciada en los resultados. Además, la distancia que separa los dos estratos sociales, de 4,3 palabras de media, es la más acusada de todas las variables sociales analizadas en este estudio. Además, la gran importancia de este condicionante social ha sido confirmada, de forma unánime, por todas las investigaciones ya realizadas (Echeverría *et al.* 1987, 1991 y 1999; Alba 1995; Etxebarria 1999; Samper y Hernández 1997; Mateo 1998; Carcedo 2001; Galoso 2002; González Martínez 2002; Blas y Casanova 2003; Bellón 2003; Gómez Molina y Gómez Devís 2004)<sup>24</sup>.

### Lengua materna

El análisis de la incidencia de la lengua materna arroja resultados, cuando menos, llamativos. Después de cotejar los promedios concernientes a cada centro de interés observamos que los informantes cuya lengua materna es el catalán sobrepasan en todos los campos léxicos a los castellanohablantes con una diferencia media de 2,43 palabras por centro de

<sup>24</sup> Ya López Morales en 1979 se percató de la influencia de este factor en su estudio sobre el léxico disponible de escolares puertorriqueños y lo puso en relación con la teoría del déficit de Bernstein (1971/1989).

interés. Consideramos que lo que puede suceder es que se sume la influencia de las distintas variables y, por esta razón, tengamos estos resultados tan inesperados<sup>25</sup>.

Centros de interés	CASTELLANO	CATALÁN
1. Partes del cuerpo	26,1	28,0
2. Ropa y complementos	22,4	24,6
3. Partes de la casa (sin los muebles)	11,4	15,5
4. Los muebles de la casa	11,7	12,8
5. Alimentos y bebidas	24,6	26,4
6. Objetos colocados en la mesa para la comida	13,0	15,3
7. La cocina y sus utensilios	16,1	16,5
8. La escuela: muebles y materiales	21,2	23,4
9. Calefacción, iluminación y medios de airear un recinto	10,8	12,2
10. La ciudad	20,8	24,1
11. El campo	17,4	20,5
12. Medios de transporte	15,2	19,0
13. Trabajos del campo y el jardín	8,6	9,6
14. Los animales	24,4	28,4
15. Juegos y distracciones	15,7	17,6
16. Profesiones y oficios	17,5	22,0

Tabla VII. *Promedio de respuestas en función de la lengua materna*

Sin embargo, al estudiar la incidencia de esta variable en la producción de interferencias<sup>26</sup>, la situación cambia ostensiblemente. En este caso, como es lógico, son los informantes de lengua materna catalana los que superan con creces a los castellanohablantes. En general, sobre 100 vocablos, los catalanohablantes producen 5,7 interferencias mientras que los castellanohablantes solo actualizan 2,7. En el gráfico siguiente podemos ver el porcentaje de interferencias de cada grupo social en cada uno de los centros de interés.

A partir de estos datos podemos sospechar que la mayor producción léxica de los catalanohablantes se debe, en cierta medida, a que echan mano de ambas lenguas al contestar el

<sup>25</sup> Esta posibilidad parece bastante probable aun más cuando sabemos que el 62% de los informantes cuya lengua materna es el catalán son del nivel social medio y el 58,6% acuden al colegio privado. De cualquier forma esperamos poder confirmar esta hipótesis al ampliar el tamaño de la muestra.

<sup>26</sup> Nos referimos con este término a todos los préstamos, transferencias semánticas o calcos sintácticos que aparecen en las listas como consecuencia de la influencia de la lengua catalana.

cuestionario, por consiguiente, puede ser que posean un lexicón en el que mezclen o sumen el vocabulario castellano y el catalán.

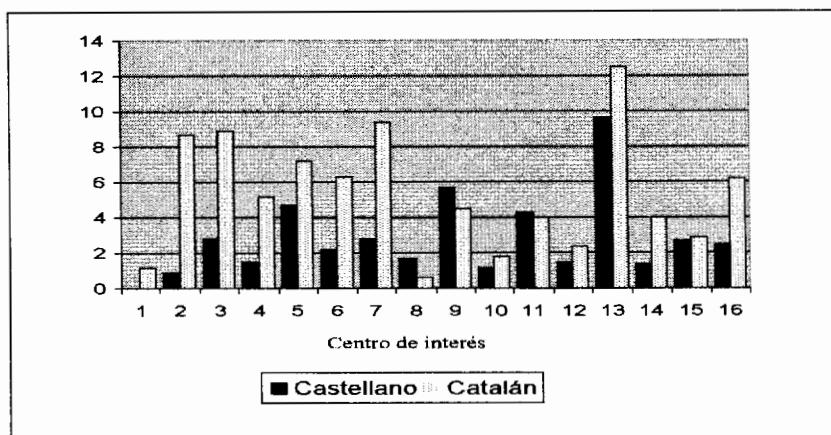


Gráfico 3: Porcentaje de interferencias según la lengua materna

### 3. Conclusiones

Las diferentes áreas temáticas, denominadas tradicionalmente *centros de interés*, han mostrado una incidencia más que notable con respecto a la producción de palabras y de vocablos. En este sentido, los centros que han generado un mayor número de respuestas han sido el 01. *El cuerpo humano*, el 14. *Los animales* y el 05. *Alimentos y bebidas*. Los centros en torno a los que el alumnado ha aportado más vocablos o palabras diferentes han sido el 11. *El campo*, el 16. *Profesiones y oficios* y el 10. *La ciudad*. No hay, por tanto, una correspondencia entre los centros que producen más y menos palabras y vocablos. Por otro lado, el promedio de productividad general de nuestros informantes se acerca bastante al obtenido en otras sintopías del mundo hispánico. También hay una considerable coincidencia entre los centros de interés que resultan más y menos productivos con respecto al número de palabras.

En cuanto a la distribución social del léxico entre el alumnado preuniversitario de la ciudad de Lérida, nuestros resultados finales muestran que el prototipo de individuo que evoca más palabras, por consiguiente, el que tiene una mayor riqueza léxica, es mujer catalanohablante de colegio privado y de nivel medio. La diferencia más acusada es la referente al nivel sociocultural, que establece una distancia de hasta 7 puntos en determinados centros de interés entre el grupo medio y el bajo del espectro social. Además, cuantitativamente, se observa que el alumnado cuya lengua materna es el catalán produce más interferencias que los castellanohablantes.

Esperamos poder corroborar en futuros trabajos con una muestra de mayor tamaño las tendencias aquí constatadas donde, además, analizaremos otras cuestiones no abordadas en esta ocasión, como es la disponibilidad léxica del catalán en la comunidad de habla leridana, para comprobar si existen o no diferencias entre una lengua y otra con respecto a los *centros de interés* y a las variables sociales tenidas en cuenta.

#### 4. Referencias

- ALBA, O. (1995), *Léxico disponible de la República Dominicana*, Pontificia Universidad Católica Madre y Maestra, Santiago de los Caballeros.
- BELLÓN FERNÁNDEZ, J.J. (2003), *Léxico disponible de la provincia de Córdoba*. Tesis doctoral inédita. Universidad de las Palmas de Gran Canaria, Las Palmas de Gran Canaria.
- BENÍTEZ PÉREZ, P. (1992), "Disponibilidad léxica en la zona metropolitana de Madrid", *Boletín de la Academia Puertorriqueña de la Lengua Española (BAPLE)* 1/1, 71-102.
- BENÍTEZ PÉREZ, P. (1994), "Convergencia y divergencias en el léxico de alumnos de COU", *Revista de Estudios de Adquisición de la Lengua Española (REALE)* 2, 39-45.
- BERNSTEIN, B. (1971/1989), *Clases, códigos y control*. II vols, Akal, Madrid.
- BLAS ARROYO, J. L. y M. CASANOVA ÁVALOS (2003), "La influencia del modelo educativo y del entorno sociocultural en la disponibilidad léxica. Estudio de las comunidades de habla castellonenses", en F. Sánchez Miret (ed.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, Salamanca, (24-30 de septiembre de 2001), Niemeyer, Tübingen.
- CARCEDO GONZÁLEZ, A. (1998), "Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica", *Lingüística* 10, 5-68.
- CARCEDO GONZÁLEZ, A. (2000), *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*, Turun Yliopiston, Turku.
- CARCEDO GONZÁLEZ, A. (2001), Léxico disponible de Asturias. Universidad de Turku, Publicaciones del Dpto. de Lengua Española, Turku.
- ECHEVERRÍA, M. S., M. O. HERRERA, P. MORENO y F. PRADENAS (1987), "Disponibilidad léxica en Educación Media", *Revista de Lingüística Teórica y Aplicada* 25, 55-115.
- ECHEVERRÍA, M. S. (1991), "Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio", en H. López Morales (ed.), *La enseñanza del español como lengua materna*. Universidad de Puerto Rico, Río Piedras, 61-78.
- ETXEBARRIA AROSTEGUI, M. (1999), "Disponibilidad léxica y enseñanza de la lengua materna en el área metropolitana del Gran Bilbao", en Samper Padilla, J. A. et al. (eds.) *Actas del XI Congreso Internacional de la ALFAL*, Universidad de las Palmas de Gran Canaria, Las Palmas de Gran Canaria, tomo II, 1479-1494.

- GALLOSO CAMACHO, M<sup>a</sup> V. (2002), *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Tesis doctoral inédita. Universidad de Salamanca, Salamanca.
- GARCÍA DOMÍNGUEZ, M<sup>a</sup> J., V. MARRERO PULIDO, J. A. PÉREZ MARTÍN y G. PIÑERO PIÑERO (1994), "Estudio de la disponibilidad léxica en Gran Canaria. La variante geográfica y el tipo de educación", *Revista de Estudios de Adquisición de la Lengua Española (REALE)* 2, 65-72
- GÓMEZ MOLINA, J.R. y B. GÓMEZ DEVÍS (2004), *La disponibilidad léxica de los estudiantes preuniversitarios valencianos. Estudio de estratificación sociolingüística*, Universitat de València, Valencia.
- GONZÁLEZ MARTÍNEZ, A. (2002), *La disponibilidad léxica de los alumnos preuniversitarios de la provincia de Cádiz*, Universidad de Cádiz, Cádiz.
- GOUGENHEIM G., R. MICHÉA, P. RIVENC y A. SAUVEGOT (1964), *L'élaboration du Français Fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Didier, Paris.
- LÓPEZ MORALES, H. (1979), "Disponibilidad léxica y estratificación socioeconómica", en *Dialectología y Sociolingüística. Temas puertorriqueños*, Hispanova de Ediciones, Madrid, 173-181.
- LÓPEZ MORALES, H. (1989), *Sociolingüística*, Gredos, Madrid.
- LÓPEZ MORALES, H. (1995-1996), "Los estudios de disponibilidad léxica: pasado y presente", *Boletín de Filología de la Universidad de Chile* 35, 245-259.
- LÓPEZ MORALES, H. (1999), *Léxico disponible de Puerto Rico*, Arco/Libros, Madrid.
- MICHÉA, R. (1953), "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage", *Les langues modernes* 47, 338-344.
- MATEO, M<sup>a</sup> V. (1998), *Disponibilidad léxica en el C.O.U. almeriense. Estudio de estratificación social*, Universidad de Almería/Grusta, Almería.
- PAREDES, F. (2001), "Disponibilidad léxica de alumnos de enseñanza secundaria de Alcalá y su comarca: resultados generales" en Cruz, I. et al. (eds.), *La Lingüística Aplicada a finales del siglo XX. Ensayos y propuestas. Actas del XVII Congreso de AESLA*, Universidad de Alcalá, tomo 2, 721-728.
- SAMPER PADILLA, J. A. y C. E. HERNÁNDEZ CABRERA (1997), "Estudio de la disponibilidad léxica en Gran Canaria. Datos iniciales y variación sociolingüística", en *Contribuciones al estudio de la lingüística hispánica. Homenaje al profesor Ramón Trujillo*, Universidad de la Laguna, La Laguna, vol. II, 229-239.
- SAMPER PADILLA, J. A. (1998), "Criterios de edición del léxico disponible: sugerencias", *Lingüística* 10, 311-333.
- SAMPER PADILLA, J. A., J J. BELLÓN FERNÁNDEZ y M. SAMPER HERNÁNDEZ (2003), "El proyecto de estudio de la disponibilidad léxica en español" en Ávila, R., J.A. Samper, H. Ueda et al., *Pautas y pistas en el análisis del léxico hispano(ameriano)*. Iberoamericana-Verbuert, Madrid-Frankfurt, 27-139.

SERRANO ZAPATA, M. (2003), "Disponibilidad léxica en la provincia de Lleida: aspectos metodológicos", *Interlingüística* 14, 929-937.

VALENCIA, A. y ECHEVARRÍA, M. (1999), *Disponibilidad léxica en estudiantes chilenos*, Ediciones Universidad de Chile-Universidad de Concepción, Santiago de Chile.