

## La variable lengua materna en la disponibilidad léxica de una comunidad bilingüe

MARIBEL SERRANO ZAPATA

Universitat de Lleida

maribel.serrano@udl.cat

<https://orcid.org/0000-0003-4378-4389>

### LA VARIABLE LENGUA MATERNA EN LA DISPONIBILIDAD LÉXICA DE UNA COMUNIDAD BILINGÜE

**RESUMEN:** En este trabajo se presenta la incidencia de la variable lengua materna sobre el léxico disponible en castellano y en catalán en la provincia de Lleida. Se analiza mediante procedimientos descriptivos y estadísticos el peso de este factor en la producción léxica en las dos lenguas de la comunidad, basándose en una muestra de 240 informantes bilingües y 480 encuestas. Los resultados muestran de forma contundente que los índices cuantitativos son superiores entre el colectivo que completa la encuesta en su lengua materna.

**PALABRAS CLAVE:** léxico disponible; lengua materna; bilingüismo; castellano; catalán.

**SUMARIO:** 1. Introducción. 2. Metodología. 3. Análisis. 4. Conclusiones. 5. Referencias.

### THE MOTHER TONGUE VARIABLE IN THE LEXICAL AVAILABILITY OF A BILINGUAL COMMUNITY

**ABSTRACT:** This paper introduces the incidence of the mother tongue factor on the lexicon available in Spanish and Catalan in the province of Lleida. The descriptive and statistical procedures of the following analysis measure the effect of the mother tongue factor on the lexical production in the two languages of the community, based on a sample of 240 bilingual informants and 480 surveys. The results clearly show that the quantitative indexes are higher among the group that completes the survey in their mother tongue.

**KEYWORDS:** lexical availability; mother tongue; bilingualism; Spanish; Catalan.

**SUMMARY:** 1. Introduction. 2. Methodology. 3. Analysis. 4. Conclusions. 5. References.

### LA VARIABLE LANGUE MATERNELLE DANS LA DISPONIBILITÉ LÉXICALE D'UNE COMMUNAUTÉ BILINGUE

**RÉSUMÉ:** Dans ce travail, on présente l'incidence de la variable langue maternelle sur le lexique disponible en espagnol et en catalan dans la province de Lleida. À l'aide de procédures descriptives et statistiques, on analyse le poids de ce facteur dans la production lexicale dans les deux langues de la communauté, sur la base d'un échantillon de 240 informateurs bilingues et de 480 enquêtes. Les résultats montrent de manière concluante que les indices quantitatifs sont plus élevés dans le groupe qui termine l'enquête dans sa langue maternelle.

**MOTS-CLÉS:** lexique disponible; langue maternelle; bilinguisme; espagnol; catalan.

**SOMMAIRE:** 1. Introduction. 2. Méthodologie. 3. Analyse. 4. Conclusions. 5. Références.

**Fecha de recepción:** 15/03/2021

**Fecha de revisión:** 25/02/2022

**Fecha de aceptación:** 18/05/2022

**Fecha de publicación:** 01/12/2022

## 1. Introducción

La disponibilidad léxica es una línea de investigación que cuenta con un notable desarrollo en la lingüística del mundo hispánico<sup>1</sup>. Los estudios realizados en España, donde el castellano convive con otras lenguas en varias comunidades, se han basado en un primer momento en el castellano (Etxebarria, 1996; Gómez Molina y Gómez Devís, 2004), y solo más tarde se ha prestado atención a las distintas lenguas autóctonas (Llopis y Gómez Devís, 2010; López Meirama y Álvarez de la Granja, 2014; Serrano Zapata, 2014). Sin embargo, los repertorios léxicos obtenidos en comunidades donde coexisten diversas lenguas nos ofrecen una información inestimable sobre la vitalidad y el estado de cada una de ellas y permiten, a su vez, estudiar otros aspectos, como los trasvases lingüísticos que se producen (Serrano Zapata, 2020).

Mediante una prueba asociativa, este tipo de estudios recaba el vocabulario que los hablantes tienen en la mente y del que disponen cuando se menciona un tema en concreto. En todos los trabajos que se enmarcan, como el presente, en el Proyecto Panhispánico de Estudio de la Disponibilidad Léxica se utilizan 16 áreas temáticas, a las que en nuestro caso hemos sumado *Colores*, tal y como se realiza en numerosas investigaciones. Este hecho —junto con las restantes condiciones de realización de la encuesta (dos minutos, estudiantes preuniversitarios...)— posibilita la comparación posterior de los resultados entre las distintas investigaciones. La mayoría de ellas incorpora un enfoque sociolingüístico y hoy en día ha quedado sobradamente demostrado que puede establecerse una estratificación social a partir de los datos que aporta el análisis conjunto de los campos léxicos y las distintas variables sociales. Tradicionalmente, se estudia el sexo, el tipo de centro, la zona geográfica y el nivel sociocultural; además, en los territorios donde conviven dos o más lenguas, se suelen incluir otras variables como la lengua materna, la lengua habitual, la comunidad de habla o el modelo de enseñanza.

El objetivo de este trabajo es conocer la incidencia de la variable lengua materna en el léxico disponible en castellano y en catalán en una situación de bilingüismo social como la que se da en la provincia de Lleida. Hasta la fecha, este factor ha mostrado que los informantes que completan la encuesta en su propia lengua materna obtienen mejores resultados, ya sea en castellano (Blas y Casanova, 2003; Ayora, 2006; Fernández Smith *et al.*, 2008 o Areta, 2009) o en catalán (Gómez Devís, 1995; Llopis, 2009). Sin embargo, no hay ningún estudio que analice la influencia de esta variable sobre la disponibi-

---

<sup>1</sup> Para un panorama general, puede verse Samper, Bellón y Samper Hernández (2003) y Samper Padilla y Samper Hernández (2006).

lidad léxica de las dos lenguas que están en contacto —basándose para ello en los mismos informantes— razón por la que nuestro estudio presenta cierta originalidad<sup>2</sup>. Además, desde la investigación de Hernández Muñoz (2006), conocemos que las variables cognitivas implicadas en la disponibilidad son la tipicidad, la familiaridad y la edad de adquisición; de forma que las palabras más típicas de una categoría, las más familiares y las más tempranamente aprendidas son más disponibles. Por todos estos motivos, partimos de las siguientes hipótesis:

1. En lengua castellana presentarán mejores promedios aquellos informantes cuya lengua materna sea el castellano mientras que en lengua catalana sucederá lo mismo con aquellos que tengan el catalán como primera lengua;
2. la estratificación social tenderá a ser similar en las dos lenguas, esto es, la variable lengua materna mostrará significación estadística en las dos lenguas o en ninguna de ellas; y
3. las diferencias en función del colectivo lingüístico seguirán mostrando preponderancia de aquel que complete la encuesta en su lengua materna; esta superioridad se evidenciará tanto en el promedio de palabras como en el número de vocablos.

## 2. Metodología

Nuestra investigación se basa en una sola muestra compuesta por 240 informantes<sup>3</sup> preuniversitarios que han completado en dos ocasiones la misma encuesta, tras un intervalo mínimo de dos semanas. El instrumento de investigación constaba de 19 páginas. En las dos primeras se hallaba una serie de preguntas encaminadas a determinar las variables sociológicas del informante —entre las cuales figuraba la de lengua materna— y, a partir de la tercera página, se distribuían los centros de interés, a razón de uno por página, en torno a los cuales el alumnado tenía que producir, en el tiempo máximo de dos minutos, las listas de unidades léxicas.

En la primera ocasión, se recabaron los datos en castellano; y en la segunda, los del catalán<sup>4</sup>. Los materiales obtenidos se procesaron mediante el programa informático *Lexidisp* (versión 1.02) y, más tarde

<sup>2</sup> Otra línea de investigación muy fecunda es la que compara la producción léxica en lengua materna y en lenguas extranjeras (Ferreira y Echevarría, 2010-2011; Santos Díaz, 2017, entre otros muchos estudios).

<sup>3</sup> Estos suponen el 9.71 % del universo estudiado. Para más información sobre la metodología de la investigación, consúltase Serrano Zapata (2014: 225-312).

<sup>4</sup> Fue relativamente fácil mantener el anonimato y, al mismo tiempo, saber qué persona había completado las dos encuestas, debido a que en las dos primeras páginas se preguntaban numerosas cuestiones relativas al individuo y sus progenitores; además de que conocíamos qué encuestas se habían pasado en cada instituto.

se volcaron en *Dispolex*<sup>5</sup>, con lo que pudimos obtener los índices cuantitativos (promedio de palabras y número de vocablos) en los que se basa el presente estudio. Con la finalidad de conocer la aplicación inferencial de nuestros resultados, también los procesamos mediante el paquete estadístico SPSS (versión 21.0.0). De este modo comprobamos si los datos recogidos podían extrapolarse al conjunto de la población estudiada.

El contacto entre el castellano y el catalán es extenso y estable en la provincia de Lleida, dado que la mayoría de la población es bilingüe<sup>6</sup> y no hay una distribución rígida de los ámbitos de uso de cada lengua, con lo que es la propia experiencia vital de cada individuo la que determina la distribución de las dos lenguas en su esfera social. Todos los informantes que participaron en nuestro estudio han sido escolarizados en lengua catalana<sup>7</sup>, son bilingües y dominan las dos lenguas desde la infancia, dada la situación de bilingüismo social en la que se hallan inmersos<sup>8</sup>. En la tabla 1 puede verse la estratificación de la muestra según la variable lengua materna.

Lengua materna	Castellano	Catalán	Ambas	Aranés
Número de informantes	66	157	10	7
Frecuencia relativa	27.50 %	65.42 %	4.17 %	2.92 %

Tabla 1: Distribución de los informantes según la variable lengua materna

Tal como puede apreciarse, el colectivo cuya lengua materna es el castellano constituye casi la tercera parte de la muestra y son mayoría los informantes cuya lengua materna es el catalán, hecho que refleja la realidad del territorio estudiado. De otra parte, son realmente pocos los informantes que declaran tener ambas lenguas como maternas o el aranés<sup>9</sup>.

<sup>5</sup> <http://www.dispolex.com/>

<sup>6</sup> Según la última encuesta sobre usos lingüísticos de la población (Generalitat de Catalunya, 2019) más del 94 % de la población de Cataluña entiende el catalán. El porcentaje aumenta hasta el 97 % en el caso del castellano. En la distribución por ámbitos territoriales, en Ponent —donde se sitúa Lleida— los porcentajes son del 94.4 % para el catalán y del 99.6 % para el castellano (datos del Institut d'Estadística de Catalunya del año 2018).

<sup>7</sup> Todos los centros educativos de Cataluña utilizan el catalán como lengua vehicular. La *Llei de Normalització Lingüística* del año 83 y el actual *Estatut d'Autonomia* determinan que el catalán es la lengua propia de la enseñanza en todos los niveles educativos.

<sup>8</sup> Con la finalidad de asegurar esta circunstancia, desechamos los datos de los informantes que no llevaran 10 años viviendo en Lleida. Además, cabe recordar que al finalizar el curso en el que se encuentran matriculados —2º de bachillerato— llevan a cabo un examen de las dos lenguas, dentro de las pruebas de selectividad, cuyo nivel es idéntico.

<sup>9</sup> El aranés es una variante del occitano propia del Valle de Arán, un pequeño territorio de 620 km<sup>2</sup>. Su presencia se debe a que uno de los puntos de la provincia donde aplicamos la encuesta fue Viella. La mitad de las encuestas han sido recogidas en la ciudad de Lleida y la otra mitad en distintos puntos, como La Seu d'Urgell, Tàrraga, Almacelles, Les Borges Blanques o Solsona.

Por este motivo, decidimos prescindir de estos datos<sup>10</sup> y nuestro análisis se basa en los dos grupos mayoritarios, a los que denominaremos castellanohablantes y catalanohablantes, respectivamente.

### 3. Análisis

El análisis se articula en tres secciones. En primer lugar, estudiamos el peso de la variable lengua materna en cada uno de los corpus, basándonos en el promedio de palabras obtenido en global y por centros de interés (3.1). A continuación, describimos las principales tendencias de los dos grupos lingüísticos —castellanohablantes y catalanohablantes— con relación a la producción del léxico disponible en cada una de las dos lenguas (3.2). Finalmente, completamos esta información con el número de vocablos producido por cada grupo de informantes (3.3).

#### 3.1. LA VARIABLE LENGUA MATERNA EN EL LÉXICO DISPONIBLE CASTELLANO Y CATALÁN

En la tabla 2 pueden observarse los promedios generales y por centro de interés de cada grupo lingüístico en cada uno de los corpus.

Nº	Centros de interés	CORPUS CASTELLANO		CORPUS CATALÁN	
		LM: castellano	LM: catalán	LM: castellano	LM: catalán
1	El cuerpo humano	26.32	25.17	27.65	29.53
2	Ropa y complementos	22.91	20.79	22.12	23.46
3	Partes de la casa (sin los muebles)	15.02	13.67	13.14	15.22
4	Muebles de la casa	13.39	11.93	12.89	12.85
5	Comidas y bebidas	24.95	23.54	25.56	26.83
6	Objetos colocados en la mesa para la comida	14.68	12.29	12.48	14.36
7	La cocina y sus utensilios	16.50	14.59	14.64	17.17
8	La escuela (muebles y materiales)	22.00	20.72	21.70	22.39

<sup>10</sup> Varias razones nos han llevado a tomar esta decisión, de un lado, el exiguo número de representantes de estos dos colectivos, y de otro, el hecho de que las características sociológicas concretas de estos informantes (por ejemplo, todos los informantes con lengua materna aranesa son mujeres) podían interferir en los resultados.

9	Calefacción, iluminación y formas de airear un recinto	11.30	10.92	9.89	11.55
10	La ciudad	21.15	21.36	19.97	21.69
11	El campo	19.08	17.50	16.86	19.01
12	Medios de transporte	16.45	16.99	16.24	17.45
13	Trabajos del campo y del jardín	8.92	8.73	8.73	10.06
14	Animales	27.23	25.38	24.09	27.20
15	Juegos y distracciones	16.14	16.10	14.92	15.93
16	Profesiones y oficios	19.71	19.27	19.11	20.77
17	Colores	19.17	17.87	17.98	18.95
Global		18.52	17.46	17.53	19.08

Tabla 2: Promedio de palabras por centro de interés según la variable lengua materna en los dos corpus

Los datos relativos a la encuesta en castellano se plasman en el gráfico 1. Al comparar la actuación de estos dos grupos de informantes, se comprueba claramente cómo los castellanohablantes obtienen mejores resultados. Este hecho se da de forma constante en todos los centros de interés, con la excepción de *La ciudad* y *Medios de transporte*, donde las cifras son, sin una gran distancia —0.21 en el primer caso y 0.54 en el segundo—, favorables a los catalanohablantes. En las restantes áreas temáticas se observa, de modo constante, la superioridad productiva de los castellanohablantes. Las diferencias más acusadas se dan en *Ropa y complementos* y en *Objetos colocados en la mesa para la comida*, con 2.12 y 2.39 palabras de más, respectivamente, siempre a favor del grupo cuya lengua materna es el español. Cabe señalar que la distancia en *Juegos y distracciones* es realmente pequeña, con 4 centésimas a favor del colectivo castellanohablante<sup>11</sup>. También en *Trabajos del campo y del jardín* la diferencia a favor del castellano es muy pequeña, de 0.19 décimas.

<sup>11</sup> Al analizar las diferencias del total de la muestra según la lengua de la encuesta, el resultado ya era superior en este campo en lengua castellana (Serrano Zapata, 2014: 316-345). En este centro de interés se observa la influencia del castellano en el ocio juvenil a través de los mercados de consumo.

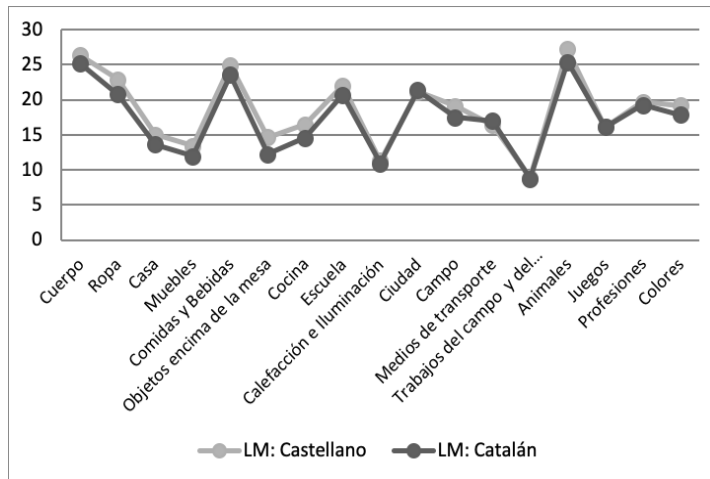


Gráfico 1: Promedio de palabras por centro de interés según la variable lengua materna en lengua castellana

Con respecto al corpus catalán, también aquí se observa, con mayor contundencia si cabe, la superioridad léxica del grupo que completa la encuesta en su lengua materna, tal como se ilustra en el gráfico 2. En todos los centros de interés el resultado de los catalanohablantes es mejor salvo en *Muebles de la casa*, donde por pocas centésimas el promedio es más elevado en el colectivo castellanohablante. La diferencia más acusada es la del centro 14, *Animales*, de más de tres palabras de diferencia (3.11). Las disimilitudes superan las dos unidades en tres áreas más: *Partes de la casa*, *La cocina y sus utensilios* y *El campo*.

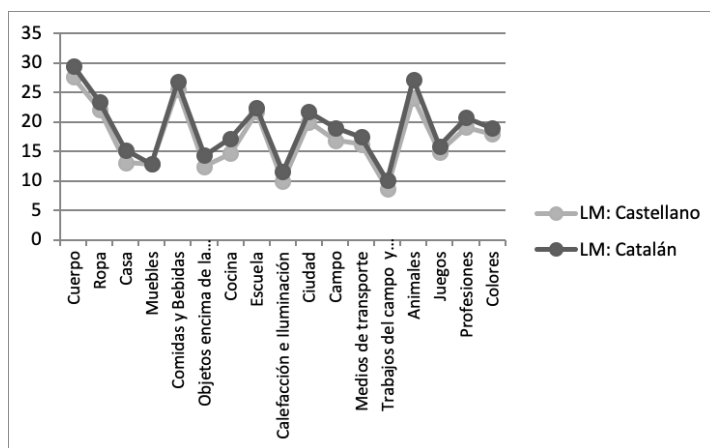


Gráfico 2: Promedio de palabras por centro de interés según la variable lengua materna en lengua catalana

A continuación, comprobamos si las diferencias detectadas son estadísticamente significativas mediante la comparación de los promedios de cada grupo lingüístico con ANOVA. Tras realizar el análisis de la varianza pertinente, se obtiene un nivel de significación del 0.012 en lengua castellana y de 0.001 en catalana, del que se colige que las diferencias globales provocadas por la variable lengua materna, en ambas lenguas, son predecibles y extrapolables al universo estudiado, puesto que el nivel de probabilidad se sitúa entre 0.000 y 0.05. Además, tal como se observa en la tabla 3, dichas diferencias pueden también extenderse a siete centros de interés en el caso de los castellanohablantes (*Ropa y complementos, Partes de la casa, Muebles de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Animales y Colores*<sup>12</sup>) y a diez áreas en el caso de los catalanohablantes (*El cuerpo humano, Partes de la casa, Objetos colocados en la mesa para la comida, La cocina y sus utensilios, Calefacción, iluminación y formas de airear un recinto, La ciudad, El campo, Trabajos del campo y del jardín, Animales y Profesiones y oficios*).

Nº	Centros de interés	CORPUS CASTELLANO	CORPUS CATALÁN
1	El cuerpo humano	0.185	0.006
2	Ropa y complementos	0.002	0.062
3	Partes de la casa (sin los muebles)	0.015	0.000
4	Muebles de la casa	0.006	0.942
5	Comidas y bebidas	0.067	0.158
6	Objetos colocados en la mesa para la comida	0.000	0.002
7	La cocina y sus utensilios	0.007	0.001
8	La escuela (muebles y materiales)	0.050	0.353
9	Calefacción, iluminación y formas de airear un recinto	0.469	0.001
10	La ciudad	0.941	0.038
11	El campo	0.066	0.012
12	Medios de transporte	0.351	0.063
13	Trabajos del campo y del jardín	0.893	0.008
14	Animales	0.014	0.000
15	Juegos y distracciones	0.951	0.146

<sup>12</sup> El centro 8, *La escuela*, se sitúa justo en el límite de la significación estadística (0.050).



16	Profesiones y oficios	0.515	0.022
17	Colores	0.018	0.097

Tabla 3: Análisis estadístico de la diferencia según la lengua materna.  
Valor de la 'p' en las dos lenguas

Es interesante analizar la diferencia que existe entre los promedios globales de los dos subgrupos. Como hemos visto, los resultados que tienen significación estadística siempre favorecen al grupo que completa la encuesta en su propia lengua materna. No obstante, la distancia que existe en el corpus castellano entre la media global de los dos grupos lingüísticos, 1.06, es inferior a la que se detecta en el corpus catalán, situada en 1.55 respuestas. Mientras que el límite inferior es prácticamente similar en los dos corpus (17.46 los catalanohablantes en el corpus castellano y 17.53 los castellano hablantes en el corpus catalán), la cifra que alcanzan los catalanohablantes en su propia lengua materna es superior a la de los castellano hablantes en la suya (19.08 frente a 18.52). En nuestra opinión, esto sucede porque los informantes catalanohablantes refuerzan el léxico de su lengua materna a través de la escuela, donde se trabaja el vocabulario de varias de las áreas de la encuesta. De otra parte, la media que alcanzan los catalanohablantes en el corpus castellano (17.46) es muy similar a la de los castellano hablantes en el corpus catalán (17.53), lo que nos da una idea de la coexistencia de los dos sistemas lingüísticos en la sociedad catalana.

Por otro lado, el hecho de que cada colectivo obtenga los mejores resultados en su propia lengua materna también nos indica que el léxico que se evoca a través de los diecisiete centros de interés está ligado al entorno básico y a la familia antes que al currículum escolar, dado que, de modo contrario, no se observarían diferencias en función de la lengua materna, puesto que todos se escolarizan en la misma lengua.

Vale la pena comentar que en los únicos centros de interés donde los catalanohablantes alcanzan mejores promedios en castellano (*La ciudad* y *Medios de transporte*) y, de forma inversa, en los que los castellano hablantes obtienen mejores cifras en catalán (en *Muebles de la casa*), la probabilidad es superior a 0.05, dato que indica que estas diferencias carecen de significación estadística.

Desde la investigación de Hernández Muñoz (2006) sabemos que la disponibilidad se asocia a las palabras más típicas de una categoría, las más familiares y las más tempranamente aprendidas. La lengua materna influye, de forma decisiva, en estos dos últimos factores, dado que se suele entrar en contacto, pensar o nombrar una palabra antes en lengua materna que en una segunda lengua y siempre se aprenden antes las palabras en la primera lengua que en la segunda. Creemos que este último factor, la edad de adquisición, influye en la

diferencia que se observa tanto en castellano como en catalán en *Animales*, o en *El cuerpo humano* en el corpus catalán, toda vez que son áreas cuyo léxico fundamental se aprende en la más tierna infancia. De modo inverso, esta misma característica puede estar incidiendo en la mayor productividad de *Ropa y complementos* y de *Colores*, en el corpus castellano<sup>13</sup>.

Cabe destacar que tres centros de interés —*Partes de la casa*, *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios*— presentan una diferencia significativa, siempre a favor del grupo que completa la encuesta en su propia lengua materna, en los dos corpus. La estrecha relación que guardan estas áreas con el hogar provoca que su vocabulario casi siempre se actualice en la lengua usada en este ámbito. Por tanto, la segunda de las características halladas por Hernández Muñoz (2006), esto es, la familiaridad, influye aquí en el sentido de que el informante entra en contacto con las palabras (o las nombra o piensa en ellas) en la lengua usada en el hogar, ya que el vocabulario de estas áreas raramente se enuncia fuera del ámbito estricto de la familia. Esta es la causa por la que, tal vez, las disimilitudes se extienden, en castellano, a otras áreas como *Muebles de la casa* y *Ropa y complementos*.

Llama la atención que en un área como *Animales* se sigan produciendo diferencias a favor de la lengua materna de cada colectivo en los dos corpus. Esto sucede a pesar de que el vocabulario que se aprende en esta categoría tiene una importante presencia en el currículum escolar. Al respecto, Hernández Muñoz (2006) demuestra que *Animales*, en castellano como L1, comprende palabras aprendidas antes que las pertenecientes a otras categorías (como *Partes del cuerpo humano*, *Ropa*, *Muebles* o *Inteligencia*), lo que puede explicar, en nuestro caso, que los resultados siempre favorezcan al colectivo que completa la encuesta en su lengua materna en este centro de interés.

### 3.2. DIFERENCIAS EN LA PRODUCCIÓN LÉXICA DE CASTELLANOHABLANTES Y CATALANOHABLANTES

A continuación, analizamos las diferencias de cada colectivo lingüístico según la lengua en que completa la encuesta.

---

<sup>13</sup> Con respecto a los resultados del catalán, sin embargo, hay que tener también en cuenta que tanto el vocabulario de *El cuerpo humano* como el de *Animales* se trabaja en la escuela, al igual que el de muchas otras categorías donde la diferencia observada es significativa (*La ciudad*, *El campo*, *Trabajos del campo y del jardín*, *Profesiones y oficios* y, parcialmente, *Calefacción, iluminación y formas de airear un recinto*). El léxico de estos centros de interés se refuerza, fundamentalmente, en la asignatura de *Coneixement del medi natural, social i cultural*, presente en todos los cursos de la educación primaria obligatoria.

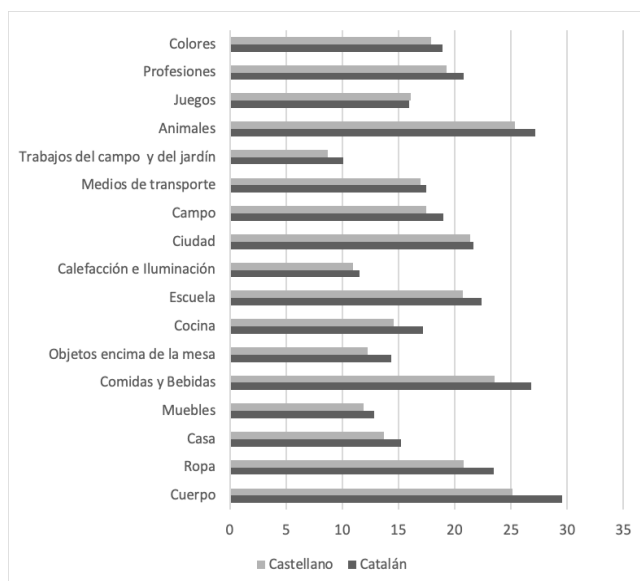


Gráfico 3: Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el castellano

En la comparación de la producción de los castellanohablantes en las dos lenguas (gráfico 3), se detecta una mayor productividad léxica en la lengua materna de los informantes. Concretamente, los resultados de este colectivo son superiores en castellano en todos los centros, excepto en *El cuerpo humano* y en *Comidas y bebidas*, donde los informantes alcanzan mejores promedios en catalán; en el primer centro, con una diferencia de 1.33 y, en el quinto, de 0.61. Salvo en estos dos casos, en las restantes áreas los promedios son siempre superiores en castellano. En *Animales*, los informantes actualizan de media 3.14 palabras más; también en *El campo* la divergencia es destacable, de más de dos unidades (2.22), al igual que en *Objetos colocados en la mesa para la comida* (2.2), un centro que, como hemos ido repitiendo, se relaciona estrechamente con la lengua utilizada en el hogar.

La diferencia entre el promedio global que los castellanohablantes tienen en castellano (18.52) y en catalán (17.53) es, prácticamente, de una respuesta. Debido a la importancia de este último dato, analizamos esta diferencia con el test T de Student y comparamos la media del grupo cuya lengua materna es el castellano en las dos lenguas. La divergencia hallada es estadísticamente significativa, a favor de la lengua castellana, puesto que el valor de 'p' se sitúa en 0.001, es decir, por debajo de 0.05<sup>14</sup>.

<sup>14</sup> Los estadísticos, las correlaciones y las pruebas de muestras relacionadas en las dos lenguas pueden consultarse en Serrano Zapata (2014: 466-469).

Finalmente, analizamos los resultados del colectivo mayoritario en la población estudiada, esto es, el que tiene el catalán como lengua inicial. Su actuación lingüística es la que presenta una mayor diferencia en función de la lengua de la encuesta (gráfico 4). Los promedios en todos los centros de interés son favorables al catalán salvo en *Juegos y distracciones*, donde por 0.17 décimas, la media es superior en castellano. La diferencia más acusada es la de *El cuerpo humano*, con 4.36 respuestas. A esta cifra, sigue la que se detecta en *Comidas y bebidas*, de 3.29 respuestas, y las de los centros *Ropa y complementos*, *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios* —áreas relacionadas con el entorno inmediato y la lengua del hogar— donde se superan siempre las dos respuestas.

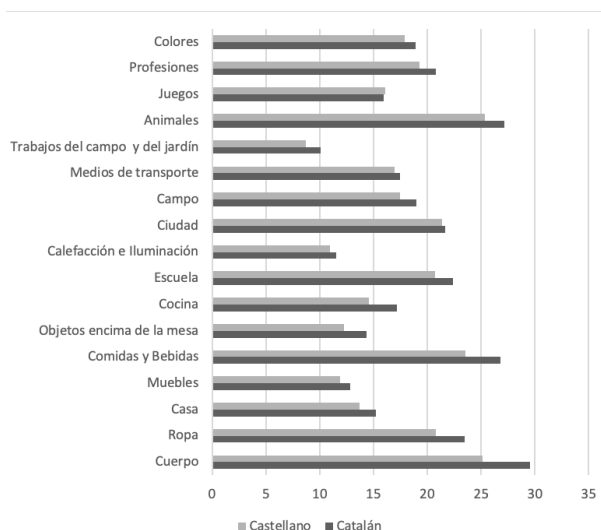


Gráfico 4: Comparación de la producción en castellano y en catalán de los informantes cuya lengua materna es el catalán

Al igual que con el colectivo castellanohablante, comparamos las medias de los catalanohablantes en los dos idiomas (17.46, en castellano y 19.08, en catalán). Al estudiar el comportamiento concreto de este colectivo en las dos encuestas, se obtiene un valor de  $p = 0.000$ , esto es, por debajo de 0.05, el límite de la significación estadística. Este valor nos indica que la lengua en que se completa la encuesta es un factor que determina el número de respuestas que actualizan los informantes.

A la luz de estos resultados y conociendo que la diferencia entre los colectivos tiene significación estadística, vale la pena detenernos, de nuevo, en las divergencias que se aprecian según el grupo lingüístico analizado. De un modo similar a como ocurría al analizar la influencia de la lengua materna en cada corpus por separado, se aprecia que la

diferencia entre el caudal léxico de los catalanohablantes en las dos lenguas es mayor (1.62) a la que se observa entre el colectivo castellanohablante (0.99), de la misma forma que el promedio del grupo catalán en su propia lengua materna es superior (19.08) al que alcanzan los castellanohablantes en la suya (18.52), quienes llegan a un nivel de dominio del léxico disponible de los dos idiomas más similar.

### 3.3. TOTAL DE VOCABLOS (PALABRAS DISTINTAS) DE CADA GRUPO LINGÜÍSTICO

Con respecto a la riqueza léxica, esto es, al número de vocablos recogido por cada grupo lingüístico en las dos lenguas, los datos se plasman en la tabla 4. Cabe señalar la relación tan evidente que existe entre el número de informantes y el total de vocablos recabado en cada área. Por este motivo, siempre el grupo que ostenta un mayor número de vocablos es el catalanohablante, puesto que es mayoría en la muestra.

Nº	Centro de interés	LM: Castellano		LM: Catalán	
		CAST	CAT	CAST	CAT
1	CUERPO	239	212	304	279
2	ROPA	197	186	233	259
3	CASA	143	136	190	215
4	MUEBLES	157	159	208	219
5	COMIDAS	323	324	377	441
6	OBJETOS	152	155	192	223
7	COCINA	220	207	259	297
8	ESCUELA	234	228	330	376
9	CALEFACCIÓN	154	135	220	240
10	CIUDAD	356	334	512	505
11	CAMPO	406	355	546	607
12	TRANSPORTE	152	164	223	248
13	TRABAJO	221	177	311	321
14	ANIMALES	232	262	284	364
15	JUEGOS	309	270	455	452
16	PROFESIONES	342	334	477	466
17	COLORES	143	139	193	194
Global		3980	3777	5314	5706

Tabla 4: Total de vocablos por centro de interés según la variable social lengua materna

Si comparamos la actuación de cada colectivo en las dos lenguas, observamos cómo, al igual que sucedía con los promedios de palabras, los castellanohablantes tienen mayor riqueza léxica en su lengua materna, tal como se manifiesta tanto en el total (donde se da una diferencia de 203 vocablos con respecto al catalán) como en la mayoría de los centros de interés. Concretamente en doce, la cifra de vocablos es superior en castellano, mientras que, en los cinco restantes (*Muebles de la casa*, *Comidas y bebidas*, *Objetos colocados en la mesa para la comida*, *Medios de transporte* y *Animales*), se obtiene una cantidad mayor en catalán<sup>15</sup>. Las mayores disimilitudes, a favor siempre del castellano, se dan en *El campo* (de 51 vocablos), en *Trabajos del campo y del jardín* (de 44) y en *Juegos y distracciones* (de 39).

En cuanto a los catalanohablantes, estos también actualizan, con una diferencia de 392 unidades, más vocablos en su lengua materna, lo que se plasma en todas las áreas salvo en *El cuerpo humano*, *La ciudad*, *Juegos y distracciones* y *Profesiones y oficios*, donde por muy pocos vocablos —excepto en el primero de los centros de interés curiosamente<sup>16</sup>— tienen mejores resultados en castellano. La superioridad de vocabulario en su lengua materna se aprecia, en mayor grado, en *Animales*, con 80 palabras distintas más, en *Comidas y bebidas*, con 64 y en *El campo*, con 61.

#### 4. Conclusiones

En relación con la primera de las hipótesis planteadas, hemos comprobado cómo el grupo que completa la encuesta en su propia lengua materna obtiene, tanto en el global como en la mayoría de los centros de interés, resultados superiores a los que obtiene el colectivo que realiza la encuesta en su segunda lengua. En concreto, en el corpus castellano, los castellanohablantes tienen mejores promedios en todas las áreas, salvo en dos, *La ciudad* y *Medios de transporte*, mientras que, en el corpus catalán, la preponderancia de los catalanohablantes es constante, excepto en *Muebles de la casa*. Con respecto a la segunda hipótesis, se infiere del análisis estadístico que la variable lengua materna provoca diferencias estadísticamente significativas, dado que el valor de la probabilidad es inferior a 0.05 tanto en castellano como en catalán. Además, es posible extender esta diferencia, en el caso del castellano, a siete áreas y, en el caso del catalán, a diez.

---

<sup>15</sup> No obstante, es justo indicar que salvo en *Medios de transporte* y en *Animales*, en las restantes áreas se observan diferencias insignificantes de 1, 2 o 3 vocablos. En *Animales* se obtienen más vocablos en catalán, independientemente del colectivo que estudiemos. La abundante presencia de transferencias lingüísticas en torno a esta área puede explicar este hecho (Serrano Zapata, 2016).

<sup>16</sup> Nótese que tanto castellanohablantes como catalanohablantes tienen más vocablos en castellano, pero mejores promedios de palabras en catalán en *El cuerpo humano*.

Mientras que la actuación lingüística en el corpus castellano es más similar en función de la lengua materna (los castellanohablantes tienen un promedio de 18.52 y los catalanohablantes de 17.46), en el corpus catalán se detectan más diferencias a favor de la lengua materna de los informantes (17.53 y 19.08 respectivamente), por tanto, la distancia que existe en el corpus castellano entre la media global de los dos grupos lingüísticos (1.06) es inferior a la que se detecta en el corpus catalán, situada en 1.55 respuestas. Por el contrario, el límite inferior es prácticamente idéntico en los dos corpus (17.46 los catalanohablantes en el corpus castellano y 17.53 los castellanohablantes en el corpus catalán).

Con respecto a la tercera hipótesis, tras analizar las diferencias en la producción considerando la lengua de la encuesta y cada grupo lingüístico, se concluye que el grupo castellanohablante muestra una preferencia clara por el castellano, con un promedio de 18.52 en su lengua materna y de 17.53 en su segunda lengua. Tras aplicar la prueba T de Student se constata cómo esta diferencia es significativa estadísticamente, lo que nos permite extrapolar este resultado al conjunto de población estudiada. De modo similar, el colectivo integrado por los informantes cuya lengua materna es la catalana presenta un promedio superior en esta lengua (19.08 frente a 17.46); también en el colectivo catalanohablante la diferencia entre las medias globales es significativa. La misma tendencia constatada con respecto a la productividad se evidencia en la riqueza léxica de cada grupo lingüístico. De este modo, los castellanohablantes tienen más vocablos en castellano, con una diferencia de 203 palabras diferentes y los catalanohablantes, por su lado, también obtienen una cifra más alta de palabras distintas en su lengua materna, concretamente, 392 más<sup>17</sup>.

En último lugar, es destacable el hecho de que en algunas áreas donde el factor lengua materna tiene relevancia estadística, como *Partes de la casa*, *Objetos colocados en la mesa para la comida* y *La cocina y sus utensilios*, se aprecie con total claridad la relación del léxico evocado en ellas con la lengua usada en el hogar. En otras, como en *Animales*, se siguen produciendo diferencias en función de la lengua materna en los dos corpus. En este caso, tal como demuestra Hernández Muñoz (2006), las divergencias se explican porque esta categoría comprende palabras aprendidas antes que las pertenecientes a otras categorías y, como sabemos, la edad de adquisición, al igual que la familiaridad y la tipicidad, son los factores cognitivos que explican la producción del léxico disponible como L1 (Hernández Muñoz, 2006). Sería deseable contar con más estudios que profundizaran en la importancia de este factor, la edad de adquisición, sobre el léxico dispo-

<sup>17</sup> Recuérdese que el número de informantes influye sobre el total de vocablos recogido en cada área y que el grupo catalanohablante se compone de 157 informantes, mientras que el castellanohablante tiene 66.

nible y, en general, sobre los factores cognitivos que intervienen en la adquisición tanto de una primera lengua como de una segunda, en un caso como el nuestro de bilingüismo social extenso con hablantes bilingües desde la infancia.

## 5. Referencias

- ARETA, M. (2009): *El léxico disponible de los estudiantes preuniversitarios navarros. Estudio sociolingüístico*. Tesis doctoral. Universidad de Navarra.
- AYORA, C. (2006): *Disponibilidad léxica en Ceuta: aspectos sociolingüísticos*, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- BLAS, J. L. y CASANOVA, M. (2003): "La influencia de la lengua materna en algunos valores de la disponibilidad léxica en el español de una comunidad bilingüe", Doval, I. y Pérez, M.<sup>a</sup> R. (eds.), *Adquisición, enseñanza y contraste de lenguas, bilingüismo y traducción*, Vigo: Universidad de Vigo, pp. 49-61.
- ETXEBARRÍA, M. (1996): "Variación e interferencia en una comunidad bilingüe: vasco y español en contacto", Arjona, M. et al. (eds.), *Actas del X Congreso Internacional de ALFAL*, México: UNAM, pp. 908-912.
- FERREIRA, R. y ECHEVARRÍA, M. S. (2010-2011): "Redes semánticas en el léxico disponible de inglés L1 e inglés LE", *Onomázein*, 21, pp.133-153.
- FERNÁNDEZ SMITH, G., RICO, A. M.<sup>a</sup>, MOLINA, M.<sup>a</sup> J. y JIMÉNEZ, M.<sup>a</sup> Á. (2008): *El léxico disponible de Melilla: estudio sociolingüístico y repertorios léxicos*, Madrid: Arco/Libros.
- GENERALITAT DE CATALUNYA (2019): *Els usos lingüístics de la població de Catalunya*, Barcelona: Generalitat de Catalunya. Disponible en: <https://llengua.gencat.cat/web/.content/documents/dadesestudis/altres/arxius/dossier-eulp-2018.pdf> (Fecha de consulta: 1/11/2020)
- GÓMEZ DEVÍS, B. (1995): *Estudi descriptiu sobre disponibilitat léxica del valencià en la ciutat de València i àrea metropolitana*. Tesis doctoral. Universitat de València.
- GÓMEZ MOLINA, J. R. y GÓMEZ DEVÍS, B. (2004): *La disponibilidad léxica de los estudiantes preuniversitarios valencianos*, Valencia: Universitat de València.
- HERNÁNDEZ MUÑOZ, N. (2006): *Hacia una teoría integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellanos-manchegos*. Salamanca: Ediciones de la Universidad de Salamanca.
- LLOPIS, F. (2009): *El lèxic del valencià central: un estudi de disponibilitat léxica*, Valencia: Servei de Publicacions de la Universitat de València.
- LLOPIS, F. y GÓMEZ DEVÍS, B. (2010): *El lèxic disponible de València*, Valencia: Denes.
- LÓPEZ MEIRAMA, B. y ÁLVAREZ DE LA GRANJA, M. (2014): *Léxico disponible do galego*, Santiago de Compostela: Servizo de Publicacións e Intercambio Científico.
- SAMPER PADILLA, J. A., BELLÓN FERNÁNDEZ, J. J. y SAMPER HERNÁNDEZ, M. (2003): "El proyecto de estudio de la disponibilidad léxica en español", Ávila, R. et al. (eds.), *Pautas y pistas en el análisis del léxico hispano(americano)*, Madrid-Frankfurt: Iberoamericana-Vervuert, pp. 27-139.



- SAMPER PADILLA, J. A. y SAMPER HERNÁNDEZ, M. (2006): "Aportaciones recientes de los estudios de disponibilidad léxica", *Lynx. Panorámica de Estudios Lingüísticos*, 5, pp. 5-95.
- SANTOS DÍAZ, I. C. (2017): "Organización de las palabras en la mente en lengua materna y lengua extranjera (inglés y francés)", *Pragmalingüística*, 25, pp. 603-617. Disponible en: <https://revistas.uca.es/index.php/pragma/article/view/2406>
- SERRANO ZAPATA, M. (2014): *Disponibilidad léxica en la provincia de Lleida: estudio comparado de dos lenguas en contacto*. Lleida: Universitat de Lleida. Disponible en <http://hdl.handle.net/10803/285008>
- SERRANO ZAPATA, M. (2016): "Fenòmens de contacte lingüístic en el lèxic disponible català de Lleida", *Studia Romanica Posnaniensia*, 43(2), pp. 51-64.
- SERRANO ZAPATA, M. (2020): "Disponibilidad léxica en las comunidades bilingües de España" en Gimeno Menéndez, F. (coord.), *Lengua, sociedad y cultura*, Publicaciones de la Universidad de Alicante, pp. 313-334.

