

Disponibilidad léxica en alumnos de español lengua extranjera del Distrito de Oporto (Portugal)

Mirta Fernández dos Santos
Universidad de Oporto
mfernandez@letras.up.pt

CUESTIONES PRELIMINARES

Entendemos por léxico disponible el que un hablante puede activar inmediatamente, en función de las necesidades derivadas de la producción lingüística. Según Samper Hernández (2002), el léxico disponible “está compuesto por aquellas palabras que, sin ser particularmente frecuentes, son evocadas y empleadas en ciertas situaciones comunicativas o para hablar de ciertos asuntos.” Dicho de una forma más sencilla, podemos considerar que una palabra es disponible si acude rápida y fácilmente a la mente del hablante cuando se aborda un tema determinado.

Dubois y sus colaboradores (1998), citados por López Rivero (2008), señalan que el vocabulario disponible es un “conjunto de palabras con una frecuencia baja y poco estable, pero usuales y útiles, que están a disposición del locutor”.

En la misma línea, López Morales (1984) considera que el vocabulario disponible es el caudal léxico utilizable en una situación comunicativa dada, o, en sus propias palabras, el “conjunto de unidades léxicas de contenido semántico concreto que sólo se actualiza si lo permite el tema del discurso”.

López Rivero (2008) añade a las definiciones anteriores el hecho de que las palabras disponibles reflejan normalmente un vocabulario de carácter nocional.

Llegados a este punto, nos parece importante hacer referencia a la diferencia que existe entre las palabras frecuentes y las palabras disponibles, ya que a menudo se confunden.

Así, el léxico frecuente está constituido por el número de casos en los que se repite un vocablo en una medición o cuantificación. Es decir, las listas de frecuencia se organizan a partir del número de apariciones de los vocablos en el discurso. Por su parte, los diccionarios de frecuencia del español han manejado, desde siempre, la frecuencia como medida de selección de palabras.

A diferencia del léxico disponible, el léxico frecuente es el que posee mayor estabilidad estadística en la lengua y los vocablos ocupan los lugares más altos en los diccionarios de frecuencia dado que aparecen en cualquier contexto comunicativo, independientemente del tema de conversación.

Michéa (1953), citado por Samper Hernández (2002), denomina a los vocablos frecuentes “palabras atemáticas” y a los vocablos disponibles “palabras temáticas”. Las listas de palabras frecuentes están encabezadas por unidades gramaticales

(preposiciones, conjunciones, etc.), seguidas de lejos por verbos, adjetivos y adverbios. En cambio, los listados de palabras disponibles están formados en su mayor parte por sustantivos. La suma de ambas nóminas lexicales, disponible y frecuente, conforma el léxico fundamental de una lengua.

Los trabajos sobre la disponibilidad léxica se ubican dentro de la léxico-estadística, pero no son las únicas investigaciones que se sirven de sistemas estadísticos para estudiar el léxico. (Samper Hernández, 2002)

López Morales (1995) incluye dentro de los trabajos basados en la léxico-estadística los diccionarios generales, las listas de frecuencia y los léxicos básicos.

Los estudios sobre disponibilidad léxica en el ámbito hispánico son relativamente recientes y sólo podemos hablar propiamente de su existencia a partir de la década de los 70. Hoy en día, este tipo de investigaciones son bastante comunes no sólo en España sino también en Hispanoamérica.

En la actualidad, muchos de los investigadores que se dedican al estudio de la disponibilidad léxica en el ámbito hispánico trabajan de forma coordinada a ambos lados del Atlántico.

Existen varios proyectos, pero entre ellos sobresale el ambicioso *Proyecto Panhispánico*, coordinado por Humberto López Morales, que tiene como objetivo principal conocer el léxico disponible de estudiantes cuya lengua materna es el español y que todavía no han empezado a cursar estudios universitarios.

En el ámbito del estudio de la disponibilidad léxica en español de alumnos de español como lengua extranjera el pionero fue Carcedo González, profesor de la Universidad de Turku (Finlandia), quien, en un primer trabajo, se ocupó del análisis cuantitativo de los datos obtenidos de las encuestas aplicadas a estudiantes finlandeses con distintos grados de dominio del español y que posteriormente se centró en el estudio de distintos aspectos de la disponibilidad, como en los vocablos que predominaban en las listas, el análisis de errores, etc.

El propio Carcedo aludió a la importancia de las pruebas realizadas con extranjeros, pues nos permiten analizar las distintas fases de aprendizaje del léxico español, descubrir los errores más habituales, así como las tendencias de dichos estudiantes a lo largo de su aprendizaje, lo que nos aporta una valiosa información sobre su *interlengua*. Asimismo, se pueden aprovechar los resultados de estas investigaciones para planificar la enseñanza del léxico o establecer comparaciones dialectales o interculturales.

A las aportaciones de Carcedo hay que añadir el estudio llevado a cabo por Samper Hernández (2002), en el que analiza el léxico disponible de 45 informantes extranjeros estudiantes de español en Salamanca, que asistieron a las clases organizadas por los Cursos Internacionales de la Universidad durante el primer trimestre del curso escolar 1999-2000.

Según Samper Padilla, Bellón (1999) y la propia Samper Hernández (2002), respecto a los anteriores estudios llevados a cabo por Carcedo “la novedad metodológica principal consiste en que estos hablantes no nativos se hallaban inmersos en la vida y la cultura del idioma que estaban aprendiendo”.

Samper Hernández (2002) determinó cuatro variables para su estudio (sexo, lengua materna, grado de dominio del español y conocimiento de otras lenguas) y llegó a la conclusión de que únicamente las variables “lengua materna” y “grado de dominio del español” arrojaron resultados significativos, mientras que las otras dos variables no supusieron diferencias relevantes. En dicho estudio se compararon, asimismo, los resultados obtenidos con los recogidos por Carcedo en Finlandia (2000), y el cotejo de

ambas nóminas lexicales reveló una compatibilidad muy alta tanto desde el punto de vista cualitativo como cuantitativo.

Por último, no podemos dejar de citar la investigación llevada a cabo por Eva López Rivero en 2008, en la que analiza el léxico disponible de 43 estudiantes de español como lengua extranjera, residentes en la zona metropolitana de Madrid. Los resultados de esta investigación coinciden bastante con los obtenidos por Samper Hernández.

Como refiere esta autora (2002), es innegable el valor de las investigaciones sobre disponibilidad léxica para la enseñanza-aprendizaje de lenguas, tanto de la lengua materna como de lenguas extranjeras.

En concreto, en lo que respecta a las posibles aplicaciones de los estudios sobre disponibilidad léxica al ámbito de la enseñanza-aprendizaje de lenguas extranjeras, es importante conocer el vocabulario más útil del idioma para comunicarse. Benítez Pérez (1994) señala, asimismo, la importancia de contar con una “base científica sólida a la hora de seleccionar el vocabulario adecuado en la confección de los manuales de ELE”. Este autor considera que estos estudios pueden resultar muy provechosos para esta finalidad, “ya que aportan palabras que sólo aparecen en ciertos discursos y permiten discernir la utilidad de las mismas según su orden de aparición”.

Otros autores, como Mateo (1994) y Mateo y García Marcos (1997), destacan la importancia que pueden tener las investigaciones realizadas con nativos para su posterior aplicación a la enseñanza del español como lengua extranjera. En ese sentido, consideramos que conocer las palabras que utilizan los nativos en determinados contextos constituye una herramienta muy útil en la determinación de los términos más productivos a enseñar a nuestros alumnos extranjeros. Eso nos permitirá también graduar mejor los contenidos léxicos que queremos enseñar y proporcionará “un material fidedigno” a los autores de manuales de ELE¹.

NUESTRA INVESTIGACIÓN

Este artículo tiene como objetivo presentar los resultados de una investigación sobre disponibilidad léxica realizada a 30 alumnos de ELE del distrito de Oporto (Portugal). La muestra se divide en dos grupos constituidos por 15 alumnos cada uno, a fin de facilitar el objetivo comparativo.

En el momento de la recogida de los datos (marzo de 2012) todos los informantes estaban inscritos en el 9.º curso de la enseñanza básica portuguesa (equivalente a 3.º de la E.S.O.) y estudiaban español desde 7.º curso, es decir, llevaban dos años y medio estudiando la lengua, y se encontraban cursando el nivel B1 del MCER². Sus edades estaban comprendidas entre los 13 y los 14 años.

Quince de los informantes eran estudiantes del *Colégio do Rosário*, un centro educativo privado de élite situado en la ciudad de Oporto, y los otros quince cursaban sus estudios en la *Escola Secundária de Vilela*, ubicada en la localidad de Vilela, perteneciente al concejo de Paredes, que, a su vez, integra el distrito de Oporto. Esta localidad, predominantemente rural, dista aproximadamente 30 kilómetros de Oporto capital.

La distribución de los informantes por género y por edades es la siguiente:

Género		Edad	
Hombres	Mujeres	13 años	14 años

¹Español como Lengua Extranjera

² Marco Común Europeo de Referencia para las Lenguas

Escola Secundária de Vilela	7	8	4	11
Colégio do Rosário	11	4	1	14

Tabla 1: Distribución de los informantes por género y edad

Respecto a las variables que se han tenido en cuenta en esta investigación, no coinciden con las que suelen ser consideradas en los trabajos de disponibilidad léxica realizados con alumnos de español lengua extranjera, a saber, “conocimiento de otras lenguas”, “lengua materna” y “nivel de conocimientos de español”, sino que hemos optado por analizar las variables clásicas contempladas en la mayoría de los trabajos hispánicos, realizados a informantes nativos, puesto que nuestra intención es introducir los datos en *Dispolex*, lo que permitirá comparar nuestros resultados con otros datos alojados en dicha aplicación. Por lo tanto, las variables que se han tenido en cuenta son las siguientes:

- Edad de los informantes
- Sexo de los informantes
- Tipo de centro (público o privado) en el que estudian los informantes.

Para esta investigación hemos considerado solo 12 de los 16 centros de interés con los que se suele trabajar en el *Proyecto Panhispánico*. Esta decisión se debe fundamentalmente a dos motivos: en primer lugar, nos interesaba seleccionar aquellos centros de interés que suelen integrar los contenidos léxicos de los manuales de ELE hasta el nivel B1 y dejar de lado los centros de interés más específicos, es decir, los que se suelen enseñar en niveles de conocimiento de lengua superiores al que poseían nuestros informantes; en segundo lugar, al carecer de herramientas informáticas para el procesamiento y posterior análisis de los datos obtenidos, trabajar con 16 centros de interés habría supuesto un esfuerzo titánico, y se habrían multiplicado las posibilidades de comisión de errores en la contabilización de los datos.

Así, los 12 centros de interés seleccionados para esta investigación fueron:

01. “Partes del cuerpo”
02. “La ropa”
03. “Partes de la casa (sin muebles)”
04. “Los muebles de la casa”
05. “Alimentos y bebidas”
06. “Objetos colocados en la mesa para la comida”
07. “La escuela: muebles y materiales”
08. “La ciudad”
09. “Medios de transporte”
10. “Los animales”
11. “Juegos y distracciones”
12. “Profesiones y oficios”

En cuanto a las hipótesis de partida, estipulamos las que se exponen a continuación:

1. No tiene por qué existir coincidencia entre los centros de interés que presentan mayor número de palabras y los que presentan mayor número de vocablos.
2. Existe relación asociativa entre la variable “sexo” y la cantidad de palabras y vocablos aportados por los informantes.
3. Existe relación asociativa entre la variable “tipo de centro de enseñanza” (público o privado) y la cantidad de palabras y vocablos aportados por los informantes; en concreto, los informantes que estudian español en el centro privado aportarán más palabras y vocablos.
4. No existe relación asociativa significativa entre la variable “edad” y la cantidad de palabras y vocablos aportados por los informantes.
5. Existe relación asociativa entre la lengua materna de los informantes (portugués) y el tipo de errores ortográficos plasmados en las palabras y vocablos aportados por los mismos.

Tras la recogida de todo el material, el siguiente paso consistió en la informatización de los datos obtenidos. Como refiere Samper Hernández (2002), esta etapa de los trabajos de disponibilidad léxica es especialmente complicada, pues, antes de proceder al tratamiento estadístico de los datos es necesario “limpiar los materiales”, es decir, suprimir todas aquellas entradas que no sirven para nuestro propósito (términos ilegibles, vocablos que no tienen relación con un centro de interés determinado, palabras escritas en la lengua materna de los informantes, etc.).

Por ello, para afrontar esta etapa, se hace imprescindible el establecimiento de unos criterios de edición de materiales, fruto de un considerable proceso de reflexión por parte del estudioso, puesto que sus decisiones pueden repercutir no sólo en los resultados cuantitativos finales obtenidos sino también en la posible comparación de los datos de su investigación con los de otras investigaciones.

Si bien la toma de decisiones ya es difícil cuando los informantes son nativos y escriben en su lengua materna, la situación se complica en caso de trabajar con informantes extranjeros, debido a la interferencia de su lengua materna. Si, además, los informantes tienen como lengua materna el portugués, como es el caso, debemos añadir la dificultad que supone la existencia de términos cognados, es decir, de términos cuyo significado y significado coinciden en ambas lenguas, lo que torna imposible la tarea de determinar si el informante está recurriendo o no a su lengua materna cuando selecciona el vocablo.

En ese sentido, y teniendo en cuenta la multiplicidad de fenómenos detectados al transcribir los materiales, hemos establecido para esta investigación los siguientes criterios de edición de datos, basándonos, en parte, en los criterios seguidos por Samper Hernández (2002):

- *Eliminación de los términos repetidos*: hemos encontrado en algunos casos la duplicación del mismo lexema por informante y por centro de interés, por lo que hemos procedido a eliminar la entrada repetida.
- *Eliminación de las entradas ininteligibles*: todas aquellas entradas ilegibles se han eliminado del cómputo general de palabras y vocablos.
- *Eliminación de los términos que no se relacionan con los campos semánticos*: en ocasiones los informantes han incluido términos en la lista de un determinado

centro de interés que no se corresponden con dicho campo semántico. Por ejemplo: *chalé en partes de la casa (sin los muebles)*.

- *Eliminación de los términos no cognados totalmente escritos en portugués*: hemos eliminado del cómputo general de entradas y del número de vocablos aquellos términos para cuya selección los informantes han acudido sin ninguna duda a su lengua materna. Por ejemplo: *talleres por cubiertos* (en portugués los cubiertos se denominan “talheres”).
- *Inclusión de los términos cognados*: hemos incluido como entradas válidas para nuestra investigación las palabras cognadas, es decir, aquellas que se escriben igual y significan lo mismo en portugués y en español, porque no hay forma de determinar si los informantes están recurriendo o no a su lengua materna. Por ejemplo: *sala o pelos*.
- *Corrección de la ortografía*: aunque, teniendo en cuenta la especificidad de nuestros informantes, no se puede considerar la existencia en los materiales de “errores de ortografía” al uso, sino más bien oscilaciones gráficas causadas por deficiencias en el aprendizaje, hemos optado, siguiendo las pautas de anteriores trabajos sobre disponibilidad léxica, por ignorar dichas oscilaciones ortográficas y uniformizarlas, adaptándolas a las formas canónicas hispánicas, basándonos para ello en la 22.^a edición del *DRAE*³. Tampoco hemos tenido en cuenta los errores de acentuación. Por ejemplo *zebra/cebra* o *cucharra/cochara/cuchara/television/televisión*.
- *Unificación en un solo vocablo de las variantes flexivas*: hemos optado en este caso por la forma no marcada del paradigma (verbos en infinitivo y sustantivos y adjetivos en su forma masculina singular). Sin embargo, hemos considerado los femeninos irregulares como vocablos distintos. Por ejemplo: *actor/actriz* o *gallo/gallina*.
- *Sufijos diminutivos que implican distintos matices semánticos*: hemos considerado vocablos distintos aquellos términos a los que la adjunción de un sufijo diminutivo atribuye matices semánticos distintos al del término del cual derivan. Por ejemplo: *mesa/mesilla* o *pájaro/pajarillo*.
- *Formas plenas y acortamientos*: hemos considerado vocablos distintos en el cómputo general las formas plenas y sus acortamientos. Esto porque consideramos que el acortamiento es un procedimiento de creación de palabras nuevas muy productivo en nuestro idioma y que el uso de las formas acortadas añade un grado de coloquialismo al discurso de los informantes extranjeros que hay que tener en cuenta. Por ejemplo: *bici/bicicleta* o *metro/metropolitano*.
- *Juegos, distracciones y correspondientes acciones*: muchas veces los informantes aluden a los distintos juegos y distracciones solo con un sustantivo y otras veces dicho sustantivo va precedido del correspondiente verbo en infinitivo. En estos casos hemos considerado en el cómputo general el sustantivo y el infinitivo + sustantivo como vocablos distintos, aunque hagan referencia a la misma realidad. Por ejemplo: *cine/ir al cine* u *ordenador/jugar en el ordenador*.

³Diccionario de la Real Academia de la Lengua Española

- *Marcas comerciales*: con presencia especialmente significativa en el centro de interés 11 (*Juegos y distracciones*), hemos optado por incluir entre las entradas válidas para esta investigación las marcas comerciales, tanto las que están totalmente lexicalizadas (*coca-cola*) como las que no (*pes 2012*).
- *Supresión del artículo en aquellas entradas en las que los informantes lo han incluido*: algunos alumnos, curiosamente, han aportado sustantivos precedidos de artículo determinado, circunstancia debida probablemente a la forma como aprendieron dicho vocabulario. Nosotros hemos optado por suprimir todos los artículos a fin de uniformizar nuestros materiales. Por ejemplo: *cabeza/la cabeza* o *brazo/el brazo*.
- *Aceptación de los neologismos universalizados*: hemos considerado entradas válidas para nuestra investigación los casos presentes en los materiales de neologismos universalizados, es decir, aquellos que se usan con relativa frecuencia entre los hispanohablantes, pese a estar escritos en inglés (xenismos). Por ejemplo: *jeans*.
- *Amplitud de las relaciones asociativas*: en algunos centros de interés, especialmente en el número 2 (*la ropa*), algunos informantes han incluido entradas no relacionadas específicamente con el campo semántico, pero que se pueden relacionar por asociación. En los casos en los que nos ha parecido clara la asociación, hemos aceptado los términos propuestos como entradas válidas para nuestra investigación. Por ejemplo: *botas* en el centro de interés 2 (*la ropa*).

En lo que concierne a los principales resultados generales de nuestra investigación, podemos destacar los siguientes:

La *cantidad total de palabras obtenidas* en los dos grupos de la muestra se eleva a 2331, de las cuales 1197 han sido aportadas por los informantes de la *Escola Secundária de Vilela* y 1134 por los informantes del *Colégio do Rosário*. Es decir, los estudiantes del centro educativo público han superado a los del centro privado en lo que respecta al número total de entradas válidas aportadas.

En cuanto al número de palabras por centro de interés, en la tabla 1 se indica el *rango o posición relativa que ocupan los centros de interés según el número total de palabras aportadas* por ambos grupos:

Escola Secundária de Vilela			Colégio do Rosário		
Rango	Centro de interés	Cantidad de palabras	Rango	Centro de interés	Cantidad de palabras
1	11. JUE	137	1	05. ALI	143
2	05. ALI	134	2	01. CUE	120
3	07. ESC	130	3	07. ESC	117
4	01. CUE	122	4	09. TRA	114
5	08. CIU	117	5	11. JUE	110
6	12. PRO	108	6	08. CIU	100
7	02. ROP	105	7	10. ANI	97
8	09. TRA	100	8	12. PRO	97
9	10. ANI	91	9	04. MUE	73
10	04. MUE	74	10	03. CAS	59
11	03. CAS	67	11	56. ROP	56

Por otro lado, si atendemos al *número total de vocablos distintos* seleccionados por ambos grupos de la muestra, observamos que los informantes del centro educativo privado han ofrecido más variedad en sus respuestas que los del centro público. Así, vemos que los informantes de la *Escola Secundária de Vilela* aportaron 319 vocablos distintos, mientras que los informantes del *Colégio do Rosário*, por su parte, ofrecieron 340 vocablos.

Analicemos ahora la incidencia del *número total de vocablos aportados por centro de interés*:

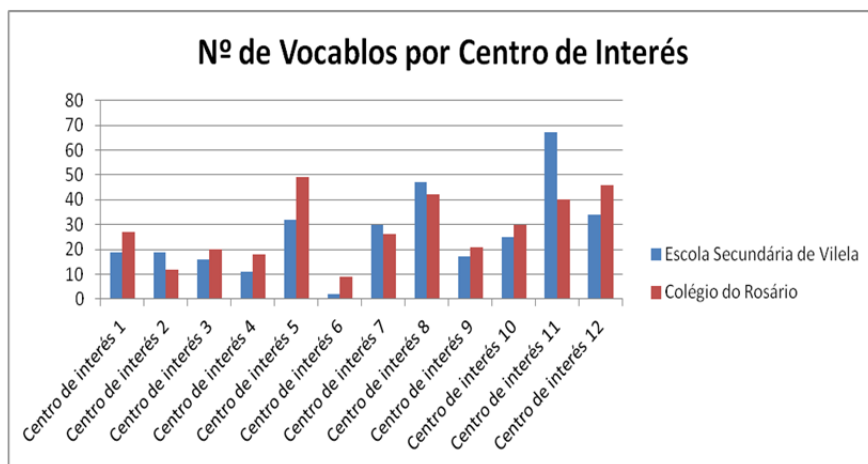


Figura 1: Número de vocablos por centro de interés

Observamos que el centro de interés en el que los informantes de la *Escola Secundária de Vilela* han seleccionado más vocablos es el centro de interés número 11 (*Juegos y distracciones*), mientras que el centro de interés en el que los informantes del *Colégio do Rosário* han ofrecido más vocablos es el número 5 (*Alimentos y bebidas*). En ambos casos existe coincidencia: han seleccionado más vocablos en aquellos centros de interés en los que también han aportado mayor número de palabras. Lo mismo se verifica en cuanto al centro de interés en el que ambos grupos han aportado menos vocablos. Se trata del centro de interés número 6 (*Objetos colocados en la mesa para la comida*), que coincide plenamente con aquel centro de interés en el que han ofrecido también menos palabras.

⁴Hemos utilizado las siguientes abreviaturas para referirnos a los nombres de los 12 centros de interés analizados:

- 01 Partes del cuerpo: CUE
- 02 La ropa: ROP
- 03 Partes de la casa (sin muebles): CAS
- 04 Los muebles de la casa: MUE
- 05 Alimentos y bebidas: ALI
- 06 Objetos colocados en la mesa para la comida: MES
- 07 La escuela: muebles y materiales: ESC
- 08 La ciudad: CIU
- 09 Medios de transporte: TRA
- 10 Los animales: ANI
- 11 Juegos y distracciones: JUE
- 12 Profesiones y oficios: PRO

⁵Vemos que los informantes del *Colégio do Rosário* han aportado el mismo número de palabras en los centros de interés 10 (*Los animales*) y 12 (*Profesiones y oficios*). En este caso hemos recurrido al orden alfabético para determinar cuál de los campos debería ocupar el rango más elevado.

A continuación presentamos los resultados cuantitativos más significativos relacionados con las variables independientes seleccionadas.

En ese sentido, relativamente a la *variable “tipo de centro educativo”*, la distribución de los informantes coincide exactamente con los dos grupos de la muestra, es decir, los 15 informantes que estudian en un colegio privado conforman en su totalidad el grupo de estudio *Colégio do Rosário*; por su parte, los 15 informantes que estudian en un centro educativo público son los que constituyen el grupo de estudio *Escola Secundária de Vilela*. Esto quiere decir que los resultados obtenidos al considerar la variable “tipo de centro educativo” coinciden con los resultados generales que ya hemos presentado. Por lo tanto, solo vamos a comentar los resultados de la investigación que se deben a la incidencia de las variables independientes “sexo” y “edad”.

En lo que respecta a la incidencia de la *variable “sexo” en cuanto al número total de palabras aportadas* por los informantes que conforman los dos grupos de la muestra, en la *Escola Secundária de Vilela* las mujeres aportaron 677 palabras y los hombres 520 (157 entradas menos), mientras que en el *Colégio do Rosário* las respuestas femeninas ascendieron a 373 y las masculinas a 761. No obstante, estas últimas cifras no son indicativas del comportamiento de la variable sexo, puesto que en este último grupo contamos con 11 hombres y solo 4 mujeres.

El análisis del número de vocablos aportados por los elementos masculinos y femeninos de ambos grupos de la muestra, nos indica que en la *Escola Secundária de Vilela* hombres y mujeres han coincidido en seleccionar más vocablos en los centros de interés 11 (*Juegos y distracciones*), 8 (*La ciudad*), 7 (*La escuela: muebles y materiales*) y 5 (*Alimentos y bebidas*), respectivamente. Por su parte, los centros de interés menos productivos para ambos sexos, desde el punto de vista de la aportación de vocablos, son el 6 (*Objetos colocados en la mesa para la comida*) y el 4 (*Muebles de la casa*). Por su parte, en la muestra del *Colégio do Rosário* observamos que los centros de interés en los que hombres y mujeres han seleccionado más vocablos son los campos 5 (*Alimentos y bebidas*), 8 (*La ciudad*) y 11 (*Juegos y distracciones*), respectivamente. En cambio, los centros de interés en los que tanto hombres como mujeres han aportado menos vocablos o unidades léxicas son los centros 6 (*Objetos colocados en la mesa para la comida*) y 2 (*La ropa*).

Deducimos, por lo tanto, que el centro de interés 6 (*Objetos colocados en la mesa para la comida*) es aquel en el que la muestra al completo ha demostrado conocer menos unidades léxicas.

Por otro lado, podemos colegir que el análisis de la incidencia de la *variable “edad”* tampoco arroja resultados significativos: en ambos grupos de la muestra el total y el promedio de las palabras y vocablos seleccionados por los estudiantes de 14 años es superior al seleccionado por los estudiantes de 13 años, pero hay que tener en cuenta que la muestra está constituida por 25 alumnos de 14 años y únicamente 5 alumnos de 13 años.

Respecto al desglose del número de vocablos por edad en ambos grupos de la muestra, llegamos a la conclusión de que los resultados son similares a los obtenidos en el análisis por sexo y por centro de interés. Es decir, para los informantes de la *Escola Secundária de Vilela*, respectivamente, los campos léxicos más productivos por edad siguen siendo el 11 (*Juegos y distracciones*) y el 8 (*La ciudad*), aunque ahora se verifica un empate en cuanto al tercer centro de interés con más vocablos entre los campos 5 (*Alimentos y bebidas*) y 12 (*Profesiones y oficios*). En lo concerniente a los centros menos prolíficos, siguen siendo los centros 6 (*Objetos colocados en la mesa para la comida*) y 4 (*Los muebles de la casa*). En el caso del *Colégio do Rosário* verificamos, sin sorpresa, que los centros de interés más productivos, haciendo el desglose por edad

de los informantes, son los mismos que en el desglose por sexo, es decir, los campos 5 (*Alimentos y bebidas*), 12 (*Profesiones y oficios*) y 8 (*La ciudad*), respectivamente. En cuanto a los centros de interés en los que los informantes han seleccionado menos vocablos, también hay coincidencia: se trata de los centros de interés 6 (*Objetos colocados en la mesa para la comida*) y 2 (*La ropa*).

Aunque el objetivo principal de nuestra investigación era la obtención de datos cuantitativos que nos permitieran analizar en términos numéricos el nivel de disponibilidad léxica de los estudiantes encuestados, nos parece importante hacer referencia también a los datos cualitativos más relevantes, a fin de ampliar el alcance de de nuestro trabajo.

Por ello, presentamos, por un lado, la lista de las unidades léxicas más utilizadas por los informantes de ambos grupos de la muestra, independientemente del centro de interés:

Unidad léxica	Ocurrencias
Silla/s	48
Mesa/s	43
Autobús/es⁶	30
Coche/s	30
Ojo/s	29
Perro/s	29
Cocina	27
Armario/s	26
Pantalones	26
Moto/s	26

Tabla 3: Unidades léxicas más utilizadas en las encuestas

Esto significa que las palabras reseñadas en la tabla son las más disponibles para los integrantes de los dos grupos de la muestra, es decir, son aquellos términos relacionados con los centros de interés propuestos que tienen más activos en su lexicón mental.

Indudablemente, la alta frecuencia de los vocablos “silla/s” y “mesa/s” tiene que ver con el hecho de que los informantes las han utilizado sistemáticamente en dos centros de interés: el número 4 (*Muebles de la casa*) y el número 7 (*La escuela: muebles y materiales*). Como estamos contabilizando el número de ocurrencias del término y no el número de vocablos, hemos procedido a sumar dichas ocurrencias, independientemente del hecho de que las mismas se hayan producido en el centro de interés número 4 o número 7.

Por otro lado, no podemos dejar de hacer referencia a dos datos que nos han llamado la atención al analizar desde el punto de vista cualitativo las encuestas que han servido de base a esta investigación y que están directamente relacionados con la lengua materna de nuestros informantes: el portugués.

En primer lugar, hemos detectado que muchos casos de vacilación ortográfica en las palabras escritas por los informantes tienen que ver con los fenómenos de “ceceo” y “seseo”. Concretamente, los informantes de la *Escola Secundária de Vilela* son seseantes y los del *Colégio do Rosário* son ceceantes. Aunque estos dos fenómenos no normativos son más propios de la oralidad que de la escritura, hemos comprobado que los informantes los reproducen también al escribir en español. Por ejemplo, uno de los alumnos del *Colégio do Rosário* escribió el término “jugador profesional” tal cual, mientras que un informante de la *Escola Secundária de Vilela* escribió “cosina” para

⁶Cabe recordar que, en cuanto a las unidades léxicas que presentan el mismo número de ocurrencias, se han organizado en la tabla por orden alfabético.

referirse a “cocina”. Lo más curioso es que tanto el “ceceo” como el “seseo” les han sido transmitidos por sus respectivas profesoras no nativas de español. Por eso consideramos que estos problemas deben atajarse proporcionando a los futuros docentes de español una formación sólida.

Otro fenómeno que nos ha sorprendido es la transposición que los alumnos luso-hablantes hacen del reciente *Acordo Ortográfico da Língua Portuguesa* al español en lo que respecta a la pérdida en la ortografía de determinadas consonantes sordas en grupos consonánticos. Por ejemplo, los términos portugueses “director” o “adoptar”, según las normas del nuevo *Acordo Ortográfico da Língua Portuguesa*, deben escribirse ahora así: “diretor” y “adotar”.

Aunque es una norma que sólo afecta a la lengua portuguesa, los alumnos luso-hablantes de español, desde la implantación de este acuerdo ortográfico, tienden a reproducir el fenómeno cuando escriben en castellano. En nuestras encuestas hemos recogido numerosos ejemplos de ello, en términos como “ator”, “diretor”, “calefacción”, etc.

Ambos fenómenos deben ser considerados como casos de interferencia o transferencia negativa de conocimientos de la lengua materna a la lengua extranjera, y es tarea de los profesores evitar que dichas interferencias, normales en estadios intermedios de aprendizaje, se fosilicen en la *interlengua* de los aprendientes, impidiéndoles alcanzar un nivel elevado de competencia comunicativa en la lengua extranjera.

CONCLUSIONES

El principal objetivo que nos planteamos al realizar este estudio era la comparación de la disponibilidad léxica de 30 estudiantes de Español Lengua Extranjera del distrito de Oporto (Portugal), divididos en dos grupos de 15 elementos: un grupo constituido por estudiantes de un centro educativo privado (*Colégio do Rosário*) y el otro por alumnos de un centro educativo público (*Escola Secundária de Vilela*). Todos los informantes tenían, en el momento de recoger los datos, entre 13 y 14 años, estudiaban español desde hacía dos años y medio y cursaban el nivel correspondiente a 3º de la E.S.O. en el sistema educativo portugués.

A ambos grupos de informantes se les administró una prueba de disponibilidad léxica con la que pretendíamos testar sus conocimientos respecto al vocabulario relacionado con 12 centros de interés.

Tras analizar y contabilizar manualmente el contenido de las encuestas, verificamos que el grupo de la *Escola Secundária de Vilela* aportó más palabras que el grupo del *Colégio do Rosário* (1197 frente a 1134), pero, en cambio, los informantes del *Colégio do Rosário* seleccionaron más vocablos o palabras distintas que los de la *Escola Secundária de Vilela* (340 frente a 319). Estos resultados contrarían parcialmente nuestra hipótesis número tres, en la que preveíamos que los estudiantes del centro educativo privado aportarían más palabras y también más vocablos que los estudiantes del centro educativo público.

Comprobamos además que, contrariamente a lo que habíamos vaticinado en nuestra hipótesis de partida número 1, los centros de interés en los que se registraron más palabras (*Juegos y distracciones* y *Los alimentos*), fueron también aquellos en los que fueron seleccionados más vocablos distintos. Del mismo modo, el centro de interés que registró menos respuestas (*Objetos colocados en la mesa para la comida*) coincidió con aquel en el que fueron seleccionados menos vocablos.

Respecto a la incidencia de las variables independientes (“sexo” y “edad”) en los resultados obtenidos, verificamos que los elementos femeninos del grupo de la *Escola Secundária de Vilela* escribieron más palabras en las encuestas que los elementos del sexo masculino (677 frente a 520), pero, sin embargo, aportaron la misma cantidad de vocablos (212). En cuanto a los datos del *Colégio do Rosário*, los hombres apuntaron más palabras que las mujeres (761 frente a 373) y también seleccionaron más vocablos (264 frente a 222). Esto significa que se ha cumplido parcialmente nuestra segunda hipótesis, en la que adelantábamos que existiría relación asociativa entre la variable “sexo” y la cantidad de palabras y vocablos aportados por los informantes.

Con relación a la incidencia de la variable “edad”, constatamos que los resultados obtenidos pueden ser engañosos teniendo en cuenta la desigual proporción que existe en la constitución de la muestra: hay muchos más alumnos de 14 años que de 13, en concreto, 25 alumnos de 14 años, frente a 5 alumnos de 13 años. Esta diferencia es más significativa, si cabe, en el grupo muestral del *Colégio do Rosário*, que integra tan solo un alumno con 13 años. Esta circunstancia torna imposible la confirmación o refutación de nuestra hipótesis número 4, en la que señalábamos que no existiría relación asociativa significativa entre la variable “edad” y la cantidad de palabras y vocablos aportados por los informantes.

En lo concerniente a los resultados arrojados por el análisis cualitativo de los materiales, cabe señalar que existe bastante coincidencia en lo que respecta a la selección de los cinco vocablos más frecuentes por centro de interés en ambos grupos de la muestra, si bien nos hemos dado cuenta, más por los términos divergentes que por los coincidentes, de que dicha selección léxica está lejos de ser arbitraria, ya que los vocablos aportados por los dos grupos de la muestra, tan heterogéneos entre sí, nos dan información sobre sus rutinas, hábitos culturales e incluso su nivel socioeconómico.

Por último, hemos verificado nuestra hipótesis número cinco, en la que preveíamos que existiría relación asociativa entre la lengua materna de nuestros informantes (el portugués) y el tipo de errores ortográficos vertidos en las encuestas. Dicha relación asociativa ha quedado patente en la escritura de términos “ceceantes” y “seseantes”, cuya causa hay que buscar en la forma como se les está enseñando la lengua española, y también en la transferencia negativa que estos alumnos hacen de normas del *Acordo Ortográfico da Língua Portuguesa* que ha entrado recientemente en vigor.

Aunque en este caso hemos enfocado nuestra investigación hacia la disponibilidad léxica en el ámbito del *Español como Lengua Extranjera*, nos parece fundamental mencionar también la importancia de este tipo de estudios en contexto de enseñanza del español como lengua materna o como segunda lengua: conociendo el caudal léxico del que disponen nuestros alumnos, podremos programar didácticamente el aprendizaje del vocabulario y poner en marcha estrategias destinadas a mejorar su competencia léxico-semántica, en caso de que sea necesario.

Referencias bibliográficas

- Bartol Hernández, José Antonio (2001): “Reflexiones sobre la disponibilidad léxica”, en José Antonio Bartol (ed.): *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*. Salamanca: Luso-Española Ediciones, 221-235.
- Benítez Pérez, Pedro y Jerzy Zebrowski, (1993): “El léxico español en los manuales polacos”, en *Actas*

- del III Congreso Nacional de la Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE), El español como lengua extranjera: de la teoría al aula.* Málaga: Servicio de Publicaciones de la Universidad de Málaga, 223-230.
- Benítez Pérez, Pedro (1994): "Léxico real/irreal en los manuales de español para extranjeros", en *Actas del II Congreso Nacional de ASELE, Español para extranjeros: didáctica e investigación.* Málaga: Servicio de Publicaciones de la Universidad de Málaga, 325-333.
- Carcedo González, Alberto (1998): "Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica". *Lingüística*, 10, 5-68.
- Carcedo González, Alberto (1999 a): "Análisis de errores léxicos del español en la interlengua de los finlandeses", en *Actas del IX Congreso Internacional de ASELE, Enfoque comunicativo y gramática.* Santiago de Compostela: Universidad de Santiago de Compostela, 465-472.
- Carcedo González, Alberto (1999 b): "Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas". *Documentos de Español Actual (DEA)*, 1. Turku: Publicaciones del Departamento de Lengua Española de la Universidad de Turku, 73-87.
- Carcedo González, Alberto (2001): *Léxico disponible de Asturias.* Colección *Monografías*, 1. Turku: Publicaciones del Departamento de Lengua Española de la Universidad de Turku.
- Dubois, Jean *et al.* (1998): *Diccionario de Lingüística.* Madrid : Alianza Editorial.
- Gougenheim, Georges *et al.* (1956): *L'élaboration du français élémentaire. Étude sur l'établissement d'un vocabulaire et d'une grammaire de base.* París: Didier.
- López Morales, Humberto (1973): *Disponibilidad léxica de los estudiantes de San Juan.* Madrid: MS.
- López Morales, Humberto (1984): *Enseñanza de la lengua materna. Lingüística para maestros de español.* Madrid: Playor.
- López Morales, Humberto (1995): "Los estudios de disponibilidad léxica: pasado y presente". *Boletín de Filología de la Universidad de Chile (Homenaje a Rodolfo Oroz)*, 35, 245-259.
- López Morales, Humberto (1999): *Léxico disponible de Puerto Rico.* Madrid: Arco Libros.
- López Rivero, Eva (2008): *Estudio de disponibilidad léxica en 43 estudiantes de ELE,* Memoria de Máster. Universidad Antonio Nebrija. Departamento de Lenguas Aplicadas.
- Mateo García, María Victoria (1998): *Disponibilidad léxica en el COU almeriense. Estudio de estratificación social.* Almería: Universidad de Almería.
- Mateo García, María Victoria y Francisco Joaquín GARCÍA MARCOS (1997): "Resultados de las encuestas sobre disponibilidad léxica realizadas en Almería". *REALE: Revista de Estudios de Adquisición de la Lengua Española*, 7, 57-68.
- Samper Hernández, Marta (2002): *Disponibilidad léxica en alumnos de español como lengua extranjera.* Monografías de ASELE, 4. Salamanca: Universidad de Salamanca.
- Samper Padilla, José Antonio (1998): "Criterios de edición del léxico disponible". *Lingüística*, 10, 311-333.
- Samper Padilla, José Antonio (1999): "Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria". *Estudios de Lingüística Hispánica. Homenaje a María Vaquero.* San Juan de Puerto Rico: Editorial de la Universidad de Puerto Rico, 551-573.