

La variable 'nivel de español' en el léxico disponible de los estudiantes de español como lengua extranjera
MARTA SÁNCHEZ-SAUS LASERNA

Universidad de Cádiz
Departamento de Filología
Facultad de Filosofía y Letras
Avda. Gómez Ulla 1 — 11003 Cádiz
Tel.: +34 956015590
Fax: +34 956015501
<marta.sanchezsaus@uca.es>

RESUMEN Son ya muy numerosos los estudios de disponibilidad léxica realizados sobre la lengua materna de una comunidad. Sin embargo, su aplicación a la enseñanza de lenguas extranjeras apenas se ha investigado. Este artículo, fruto de una investigación en la que se ha estudiado el léxico disponible de los estudiantes de español como lengua extranjera de la Universidad de Cádiz y los factores que le influyen, analiza el efecto del nivel de español sobre el número de unidades léxicas que se han actualizado en las encuestas de disponibilidad. Además, hace especial hincapié en la importancia de que se lleve a cabo, no solo un análisis estadístico descriptivo de los resultados, sino también inductivo, el cual permite extrapolarlos a la totalidad del grupo de estudiantes que es objeto del estudio.

PALABRAS CLAVE Lingüística aplicada. Variación. Léxico. Estadística. Disponibilidad léxica. Enseñanza del español como lengua extranjera.

SUMARIO 1 Introducción. – 2 La disponibilidad léxica. – 3 Estudio del léxico disponible de los alumnos de español como lengua extranjera de la Universidad de Cádiz. 3.1 La muestra y la recogida de datos. 3.2 Variables seleccionadas. 3.3 Los centros de interés y la encuesta. 3.4 Edición y tratamiento estadístico de los datos. – 4 Resultados cuantitativos. 4.1 Análisis descriptivo. 4.2 Análisis inductivo. – Conclusiones. – Referencias.

La variable "niveau d'espagnol" dans le lexique disponible des étudiants d'espagnol comme langue étrangère

RÉSUMÉ Les études sur la disponibilité lexicale réalisées sur la langue maternelle d'une communauté sont déjà nombreuses. Pourtant, celles qui ont appliqué cette méthodologie sur l'enseignement des langues étrangères sont toujours rares. Cet article, qui est sorti de la recherche sur le lexique disponible des étudiants d'espagnol langue étrangère à la Université de Cadix et des facteurs qui ont de l'influence en lui, analyse l'effet du niveau d'espagnol sur le nombre d'unités lexicales que les personnes interrogées ont actualisé dans les enquêtes de disponibilité. D'un autre côté, il met l'accent sur l'importance de réaliser une étude statistique des résultats non seulement descriptive, mais aussi inductive, qui permet de les extrapoler à la totalité du groupe d'étudiants qui fait l'objet de l'étude.

MOTS CLÉ Linguistique appliquée. Variation. Statistique lexicale. Disponibilité lexicale. Enseignement de l'espagnol langue étrangère.

SOMMAIRE 1 introduction. – 2 La disponibilité lexicale. – 3 Étude du lexique disponible des élèves d'espagnol comme langue étrangère de l'Université de Cadix. 3.1 L'échantillon y la récolte des données. 3.2 Variables sélectionnées. 3.3 Les centres d'intérêts et le sondage. 3.4 Édition et traitement statistique des données. – 4 Résultats quantitatifs. 4.1. Analyse descriptive. 4.2 Analyse inductive. – Conclusion. – Références

The "level of Spanish" variable in the context of the lexicon available to students of Spanish as a foreign language

ABSTRACT There are already plenty of lexical availability studies for the mother tongue of a community. For the teaching of foreign languages, on the other hand, there are very few. This article is based on a study of the lexical availability of students of Spanish as a foreign language at the University of Cadiz and of the factors which affect this availability. It analyzes the effect of the students' level of Spanish on the number of lexical units found in the availability surveys. Further, this paper emphasizes the importance of not only a descriptive statistical analysis but also an inductive one, since it allows extrapolating them to the whole group of students which is the focus of the study.

KEY WORDS Applied linguistics. Variation. Lexical statistics. Lexical availability. Spanish as a foreign language.

SUMMARY 1 Introduction. – 2 Lexical availability. – 3 Study of lexicon available to students of Spanish as a foreign language at UCA. 3.1 Samples and data collection. 3.2 Selected variables. 3.3 Focus of interest and survey. 3.4 Statistical analysis. – 4 Quantitative results. 4.1 Descriptive analysis. 4.2 Inductive analysis. – Conclusions. – References.

La variable “nivel de español” en el léxico disponible de los estudiantes de español como lengua extranjera

Marta Sánchez-Saus Laserna

1 Introducción

Los estudios de disponibilidad léxica han experimentado un gran auge en los últimos años; sin embargo, pocas han sido las investigaciones que han aprovechado los métodos de análisis del léxico que ofrecen estos estudios para aplicarlos a la enseñanza y aprendizaje de lenguas extranjeras, a pesar de que este es un campo de interés de una relevancia cada vez mayor. El creciente atractivo que tiene el español a nivel mundial ha traído consigo la aparición de una ingente cantidad de materiales didácticos en los que se detecta con frecuencia una deficiente o, en cualquier caso, inadecuada selección del vocabulario que se ofrece al estudiante de español. Este artículo es fruto de una investigación que se fundamenta, pues, en la necesidad que existe de este tipo de análisis para que la enseñanza del léxico en español como lengua extranjera tenga unas bases sólidas y se corrijan las inadecuaciones en la selección del vocabulario que presentan en ocasiones los libros y manuales que se emplean para su enseñanza.

Desde que surgieron los estudios de disponibilidad, a mediados del siglo pasado, se les ha definido como el método de análisis del léxico complementario a las tradicionales investigaciones basadas en la frecuencia, ya que desvela no qué palabras son las que más se usan, las más frecuentes, sino aquellas que vienen primero a la mente del hablante en una situación comunicativa dada, de tal manera que la suma de ambos léxicos, el disponible y el frecuente, conforma el léxico fundamental de una lengua. Si tenemos esto en cuenta, así como lo avanzado de tales estudios sobre el español como lengua materna, veremos claramente la necesidad que existe de aplicar el mismo método de investigación sobre el español que hablan los estudiantes extranjeros de nuestra lengua, pues es preciso sentar las bases que definen qué diferencias existen entre ambos léxicos, qué características propias presenta el vocabulario de estos estudiantes y qué variables le afectan, para así fijar el léxico que deben aprender y la metodología que debe aplicarse para su enseñanza.

2 La disponibilidad léxica

Los estudios de disponibilidad léxica son, como señalan todas las revisiones teóricas sobre el tema, relativamente recientes. El primer trabajo publicado que hace referencia a este procedimiento estadístico se publicó en la década de 1950 (Michéa 1953), aunque no es hasta los años 80 cuando se empiezan a realizar investigaciones con un criterio común y con la fórmula estadística definitiva. Sin embargo, la preocupación por determinar el léxico fundamental de una lengua, por definir qué palabras son las que más se utilizan y, sobre todo, por establecer qué vocabulario se debe enseñar es más antigua. Y esto se debe a la absoluta conciencia de que, al igual que los hablantes adultos de una lengua no manejan todos los vocablos que su sistema les ofrece, una restricción, y por tanto una selección, debe llevarse a cabo también sobre el vocabulario que se enseña. Como comentan Grève y Van Passel (1971: 105), “a la vista de la limitación lexicológica [...] tanto en el terreno literario como en el de la lengua cotidiana, se impone una limitación semejante en la enseñanza lingüística, tanto de la lengua materna como de las lenguas extranjeras”.

En este sentido, desde finales del siglo XIX, y también con anterioridad, pero fundamentalmente a partir de esa época, se vienen publicando obras lexicográficas destinadas a recoger el vocabulario básico de una comunidad. No siempre estuvieron pensadas para la enseñanza de lenguas y a muchas se les ha criticado que únicamente tuvieran en cuenta textos escritos, pero deben considerarse pioneras en este campo¹.

Esta creencia general de que los elementos léxicos más frecuentes, en función de su número de apariciones en un corpus, son los más útiles en el manejo de cualquier lengua y, por tanto, conforman su vocabulario básico, tuvo que ser revisada a partir de los trabajos que hacia 1950 llevó a cabo un grupo de investigadores franceses con el fin de proporcionar a los estudiantes de francés como lengua extranjera una primera etapa sólida en la adquisición de dicha lengua, los cuales dieron lugar a la elaboración del *Français Fondamental* (G. Gougenheim, R. Michéa, P. Rivenc & A. Sauvageot 1964). El principio en el que se basaba este estudio para la selección de vocablos era el de la frecuencia de aparición, que se ordenaba según rangos, es decir, estableciendo una jerarquía en función de la cantidad de veces que un vocablo se repetía en un texto. Sin embargo, pronto advirtieron que este método revelaba una frecuencia muy baja e incluso la ausencia de palabras muy conocidas y muy usadas por los hablantes de francés, las cuales no aparecían en las listas debido a la vinculación que tenían con determinadas circunstancias y temas de conversación (Michéa 1953).

¹ Carcedo (2000: 11 y ss.) lleva a cabo un detenido repaso historiográfico sobre el tema, basándose fundamentalmente en Gougenheim *et alii* (1956). Remitimos a estas dos obras para una mayor información sobre el tema.

La ausencia en las listas de vocablos tan cotidianos como *tenedor*, *metro* o *diente* hizo sospechar de la insuficiencia de esta metodología, por lo que se decidió completar los datos de frecuencia con la investigación de la disponibilidad del vocabulario: se necesitaba una fórmula que no se basara únicamente en las listas de frecuencia, sino también en su dispersión simple en un primer momento para pasar a la dispersión compleja y a calcular así el índice de uso. El rendimiento de estos léxicos básicos era mejor que el de los vocabularios frecuentes, ya que la dispersión evitaba el peligro que supone obviar la distribución de las frecuencias en el texto.

La asociación que se descubre entre este tipo de léxico muy básico, o léxico temático, y las situaciones comunicativas concretas, así como la revisión crítica consiguiente sobre los métodos que se utilizaban para medir el léxico, llevan a la creación de una nueva metodología de trabajo y de un nuevo término: la disponibilidad léxica y el concepto de léxico disponible. Michéa (1953: 342) fue el primero en utilizar el término, definiéndolo como sigue:

En présence d'une situation donnée, les mots qui viennent les premiers à l'esprit sont ceux qui sont liés tout spécialement à cette situation et la caractérisent. [...] Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé, et se présente immédiatement à l'esprit au moment où l'on en a besoin.

Muchas de estas palabras disponibles resultaban ser poco estables en las listas de frecuencia, pero de gran estabilidad en los recuentos realizados a partir de determinados estímulos temáticos o centros de interés, en torno a los cuales el encuestado produce una serie de unidades léxicas. Se separan, por tanto, los conceptos de palabra frecuente y palabra disponible y la disponibilidad léxica se convierte en un tipo de análisis propio, complementario del análisis de frecuencia, ya que, mientras que en las listas de frecuencia la mitad de las palabras que aparecen son palabras gramaticales –artículos, preposiciones y conjunciones– y las restantes, en orden decreciente, verbos, adjetivos y, por último, sustantivos, en el léxico disponible la gran mayoría de los vocablos son sustantivos, salvo algunos verbos, adverbios y adjetivos –debido, fundamentalmente, a la metodología empleada y al tipo de centros de interés seleccionados–. Dado que el léxico fundamental de una lengua lo constituye la suma de ambos conjuntos de vocabulario usual, el frecuente y el disponible, estos dos tipos de estudio se complementan, es decir, el léxico fundamental de una lengua está compuesto por el frecuente más el disponible.

3 Estudio del léxico disponible de los alumnos de español como lengua extranjera de la Universidad de Cádiz

3.1 La muestra y la recogida de datos

La disponibilidad léxica, por tanto, se puede definir como el caudal de vocabulario que puede aparecer cuando la situación comunicativa así lo requiere. Para recoger los datos sobre este léxico son necesarias unas pruebas de tipo asociativo en las que el informante actualiza esas unidades léxicas a partir de unos estímulos o centros de interés y anota todas las palabras que se relacionan con ellos, dando lugar a listados de vocablos que permiten conocer el léxico real con el que cuenta un hablante cuando el tema de conversación gira en torno a uno de estos centros.

Nuestro estudio se ha llevado a cabo sobre un total de 81 alumnos universitarios que estudian español como lengua extranjera en el Centro Superior de Lenguas Modernas (CSLM) de la Universidad de Cádiz. La recolección de datos se produjo en el mes de abril de 2008, aunque los estudiantes habían permanecido en España al menos desde enero-febrero de 2008, si bien algunos habían llegado a nuestro país a principios de curso, en septiembre-octubre de 2007. Es decir, se trata de un estudio realizado sobre alumnos universitarios extranjeros, estudiantes de español como lengua extranjera y residentes, al menos temporalmente, en España.

3.2 Variables seleccionadas

Las variables que se han tenido en cuenta difieren significativamente de las que se han utilizado para los estudios de disponibilidad léxica sobre hablantes nativos de español. Estos suelen considerar el sexo, el tipo de centro escolar (público o privado), la localización del centro (rural o urbano) y el estrato sociocultural; en ocasiones, cuando la prueba se realiza sobre comunidades bilingües, se recogen también los datos relativos a la lengua materna. Nosotros, sin embargo, hemos trabajado con las variables sexo, lengua materna, nivel de español, conocimiento de otras lenguas románicas como lenguas extranjeras y frecuencia con que los encuestados utilizan el español.

Las variantes de cada variable, así como el número de encuestados con el que cada una cuenta, son los siguientes:

- sexo: hombre (28) y mujer (53);
- nivel de español: A (18), B (42) y C (21);
- lengua materna: inglés (21), francés (7), alemán (28), italiano (8), lenguas eslavas (10), lenguas finougrias (7);

- conocimiento de otras lenguas: conoce otra u otras lenguas extranjeras, de las cuales ninguna es románica (39), conoce otra u otras lenguas extranjeras, de las cuales al menos una es románica (39);
- uso del español: alto o frecuente (40) y bajo o poco frecuente (35).

Respecto a las variantes que examinamos relacionadas con el nivel de español, que es el factor que ahora nos interesa, el Centro Superior de Lenguas Modernas de la Universidad de Cádiz divide a sus alumnos según los niveles del Marco Común Europeo de Referencia (mcer) -de menor a mayor nivel: A1, A2, B1, B2, C1 y C2-, pero consideramos que seis variantes diferentes son excesivas y no se diferenciarían lo suficiente entre ellas, por lo que las hemos reducido a tres: A, B y C, cada una de las cuales aglutina a sus dos subgrupos 1 y 2. Como se puede comprobar en el apartado dedicado al análisis de los resultados, la división de la muestra en tres grupos según su nivel es adecuada y proporciona resultados significativos.

3.3 Los centros de interés y la encuesta

Los centros de interés elegidos suponen una de las grandes diferencias de esta investigación con respecto a la mayor parte de las ya publicadas sobre disponibilidad léxica. Se ha partido, generalmente, de los dieciséis campos utilizados por los investigadores pioneros, G. Gougenheim, R. Michéa, P. Rivenc y A. Sauvageot (1964), los cuales ha utilizado la mayoría de los estudios publicados para el español: 1) *partes del cuerpo humano*; 2) *la ropa*; 3) *partes de la casa (sin muebles)*; 4) *los muebles de la casa*; 5) *alimentos y bebidas*; 6) *objetos colocados en la mesa para la comida*; 7) *la cocina y sus utensilios*; 8) *la escuela: muebles y materiales*; 9) *iluminación, calefacción y medios de airear un recinto*; 10) *la ciudad*; 11) *el campo*; 12) *medios de transporte*; 13) *trabajos del campo y del jardín*; 14) *los animales*; 15) *juegos y distracciones*; 16) *profesiones y oficios*.

Desde el principio, nos dimos cuenta de que algunos de estos centros podían no resultar de mucho interés cuando se trabaja con informantes extranjeros que, por definición, no dominan el español. En algunos casos, se debe a la poca productividad del centro, como ocurre con *Iluminación, calefacción y medios de airear un recinto* y *Trabajos del campo y el jardín*; otros parecían excesivamente específicos: *Partes de la casa (sin muebles)*, *Los muebles de la casa*, *Objetos colocados en la mesa para la comida* y *La escuela: muebles y materiales*; y, por último, se dejaban de lado temas básicos que aparecen en cualquier listado de contenidos léxicos para los primeros niveles de aprendizaje del español como lengua extranjera. Por ello, nos hemos basado en la lista de temas de vocabulario que ha de conocer un alumno que alcance el nivel A1 del mcer (robes ávila 2006: 83) y, comparándola con las empleadas en otros estudios,

hemos creado una propia. Esta, además, incluye centros que tienen el objetivo de registrar elementos léxicos pertenecientes a categorías lingüísticas diferentes del sustantivo, ya que la mayoría de los trabajos de disponibilidad se limita casi exclusivamente a la recogida de estos.

Así pues, los centros que hemos utilizado son los siguientes: 1) *el cuerpo humano*; 2) *la ropa*; 3) *la casa*; 4) *alimentos y bebidas*; 5) *la cocina y sus utensilios*; 6) *escuela y universidad*; 7) *la ciudad*; 8) *el campo*; 9) *medios de transporte*; 10) *los animales*; 11) *ocio y tiempo libre*; 12) *profesiones y trabajos*; 13) *tiempo meteorológico y clima*; 14) *acciones y actividades habituales*; 15) *aspecto físico y carácter*; 16) *la familia*; 17) *viajes y vacaciones*; y 18) *el dinero*.

Como hemos dicho anteriormente, consideramos que los centros elegidos nos permiten, como objetivo principal, abarcar la mayor parte del léxico que los alumnos conocen o deberían conocer, según el mcer, manteniendo al mismo tiempo los suficientes puntos en común con otras investigaciones como para poder llevar a cabo, más adelante, una comparación de los resultados.

La aplicación de la encuesta se llevó a cabo de acuerdo con las directrices habituales en los trabajos de disponibilidad léxica².

3.4 Edición y tratamiento estadístico de los datos

La edición de los datos obtenidos en las encuestas se ajusta a las recomendaciones que han seguido los trabajos de disponibilidad léxica sobre hablantes nativos de español³.

Para el tratamiento de los datos se ha utilizado el programa *dispolex*, que supone una revisión del programa pionero que se creó para los estudios de disponibilidad, *lexidisp*. *dispolex* extrae los datos relativos a los índices cuantitativos por centro de interés (número de palabras totales, número de palabras diferentes, número de palabras por persona e índices de cohesión) y a la disponibilidad léxica de las palabras de cada centro. Todo ello, además, puede cruzarse con las distintas variables que se tienen en cuenta para la investigación.

2 El cuestionario se cumplimentó de forma escrita. Antes de que la prueba comenzara dimos las instrucciones necesarias sobre cómo ha de responderse a las preguntas relacionadas con los datos de los informantes, explicamos las reglas que se debían seguir para realizar el test –las palabras deben escribirse en español, los temas se escriben en la pizarra y a partir de ese momento comienza el tiempo para escribir todas las palabras que se conozcan relacionadas con el tema, pueden escribirse todas las palabras para las que se tenga tiempo, la ortografía no influye en el resultado y no deben repetirse palabras en un mismo centro– y se explicó que para cada tema se contaba con dos minutos.

3 Remitimos a Samper Padilla (1998) para su consulta.

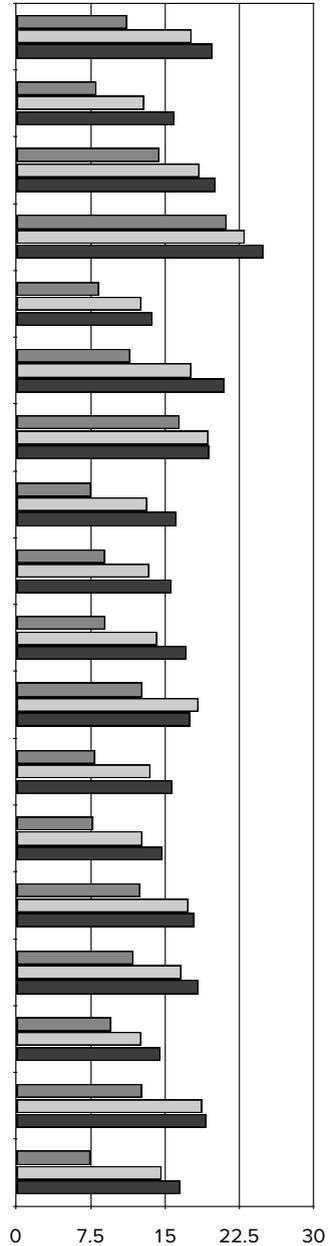
4 Resultados cuantitativos

4.1 Análisis descriptivo

López Morales (1983: 209-210) señala, en su introducción al capítulo sobre lingüística estadística de su obra *Introducción a la lingüística actual*, que la estadística ya no es solo un recurso técnico o una disciplina de apoyo para la lingüística actual, sino que constituye un soporte metodológico imprescindible y que la razón de tal cambio ha sido el auge de la léxico-estadística. El trabajo que hemos llevado a cabo –y del que este artículo describe solo una parte– es fundamentalmente un estudio estadístico. Desde esta perspectiva, vamos a mostrar los resultados derivados del uso de la variable “nivel de español” como posible factor relevante para establecer diferencias en la cantidad de palabras que conocen los informantes. A continuación, compararemos estos resultados estrictamente descriptivos con los procedentes del análisis de regresión lineal, que nos permite extrapolar los resultados obtenidos sobre la muestra al total de la población estudiada, de tal manera que podamos concluir si la variable que se ha empleado explica realmente la variación que se da en los individuos, superando los problemas que presenta el análisis exclusivamente descriptivo, como son la desproporción en el número de informantes de los distintos grupos y el posible solapamiento de unas variables sobre otras.

De los factores que se han tenido en cuenta en el estudio, los cuales, como hemos dicho, han sido sexo, lengua materna, conocimiento de otras lenguas románicas – aparte del español– como lengua extranjera, uso del español y nivel de español, este último se ha revelado como el causante de una mayor diferencia entre los alumnos encuestados. Explicamos anteriormente que hemos seleccionado tres variantes para esta variable: nivel A, nivel B y nivel C, según el mcer. La muestra cuenta con 18 informantes en el nivel A, 42 en el B y 21 en el C, los cuales escribieron un total de 3.568 palabras en el grupo inicial, 12.024 en el medio y 6.671 en el superior. La tabla de la página siguiente reproduce las medias de palabras por persona en cada centro de interés y en función de los tres niveles observados. A su derecha, la gráfica pone de relieve las diferencias:

CENTROS DE INTERÉS	NIVEL A	NIVEL B	NIVEL C
1 el cuerpo humano	11.11	17.64	19.76
2 la ropa	8.06	12.83	15.90
3 la casa	14.39	18.43	20.05
4 alimentos y bebidas	21.17	23.00	24.86
5 la cocina y sus utensilios	8.28	12.57	13.71
6 escuela y universidad	11.44	17.60	20.95
7 la ciudad	16.44	19.33	19.38
8 el campo	7.50	13.14	16.10
9 medios de transporte	8.89	13.36	15.62
10 los animales	8.94	14.14	17.14
11 ocio y tiempo libre	12.67	18.33	17.48
12 profesiones y trabajos	7.89	13.50	15.67
13 tiempo meteorológico y clima	7.67	12.67	14.71
14 acciones y actividades habituales	12.44	17.31	17.90
15 aspecto físico y carácter	11.78	16.60	18.33
16 la familia	9.50	12.55	14.52
17 viajes y vacaciones	12.61	18.67	19.10
18 el dinero	7.44	14.62	16.48
MEDIA	11,01	15,91	17,95



Tanto en la tabla como en la gráfica se observa que las diferencias entre niveles de español son importantes en algunos aspectos. La media de respuestas por centro en el nivel A es de 11,01 palabras, en B es de 15,91 y en C de 17,95. Esto implica un 44,43% de respuestas más en B que en A y un aumento del 10,96% en las palabras escritas por el nivel C con respecto al nivel B. Entre los niveles A y C la subida es del 63,03%.

Entre el nivel B y el A la diferencia media entre palabras escritas por los estudiantes es de 4,89 palabras por centro. Si nos fijamos en los resultados por campo, se dan las mayores diferencias en el 18) *El dinero* (7,18 palabras), el 1) *El cuerpo humano* (6,53) y el 6) *Escuela y universidad* (6,16). En el lado opuesto, los centros 4) *Alimentos y bebidas* y 7) *La ciudad* son, con una diferencia de 1,83 y 2,89 palabras, respectivamente, los que registran un menor contraste.

Entre los niveles B y C las diferencias son mucho menores, con una media de 1,74 palabras más por informante del nivel C con respecto a los del B. Entre los campos con mayor desigualdad en las respuestas se observa de nuevo el 6) *Escuela y universidad* (3,35 palabras más en C), al que le acompañan el 2) *La ropa* (3,07 palabras) y el 8) *El campo* (2,98 palabras). En esta comparación hay un centro que recoge más palabras en el nivel B que en el C: el 11) *Ocio y tiempo libre*, con una diferencia de 0,85 palabras. Sin llegar a ser negativo el resultado, los centros 7) *La ciudad* y 17) *Viajes y vacaciones* son los que menos contraste presentan en las respuestas de los niveles B y C: 0,05 y 0,43 respuestas más en C, respectivamente.

Al observar las diferencias entre los extremos, entre A y C el nivel superior registra una media de 6,63 palabras más que el inferior. Las divergencias más significativas se obtienen, de nuevo, en 6) *Escuela y universidad* (9,51 palabras más), en 18) *El dinero* (9,04) y en 1) *El cuerpo humano* (8,65). Las menores se dan, y volvemos a ver centros repetidos, en 7) *La ciudad* (2,94 palabras más en C que en A), en 4) *Alimentos y bebidas* (3,69) y en 11) *Ocio y tiempo libre* (4,81).

Así pues, la variable “nivel de español”, al menos desde un punto de vista descriptivo, guarda una relación muy estrecha con el número de palabras que los estudiantes actualizan en las encuestas. Conforme se aumenta el nivel, la media de palabras asciende. Sin embargo, el crecimiento no es homogéneo: se produce un aumento muy significativo en la cantidad de unidades léxicas registradas entre los niveles inicial y medio, cercano al cincuenta por ciento; no obstante, entre el medio y el superior la tendencia cesa y el crecimiento es menor. De aquí podemos deducir, pues, que en el grupo de estudiantes que se ha tomado como muestra el número de palabras conocidas y actualizables en el uso resulta ser de mayor importancia para diferenciar al nivel medio del inferior que para separar al superior del medio. Además, aquellos alumnos situados en los niveles más bajos están presumiblemente en pleno proceso de aprendizaje léxico, mientras que, para los que pertenecen a los grupos avanzados, el proceso se ha ralentizado.

Fijémonos ahora en las diferencias entre centros. Aquellos en los que menos difieren los resultados son los más relacionados con la vida diaria, con los aspectos más prácticos, con los temas más cercanos a los estudiantes desde que viven en España: la comida y la bebida, la ciudad, el ocio y las vacaciones. Sin embargo, las palabras vinculadas a temas como las partes del cuerpo, el dinero o la universidad parecen ser consecuencia de un mayor tiempo de estudio, como revela el gran aumento que experimenta el número de palabras relacionadas con ellos conforme aumenta el nivel de los encuestados.

4.2 Análisis inductivo

Cualquier estudio estadístico tiene su razón de ser no en lo que pueda decir sobre el grupo estudiado, sino en la información que los resultados de una muestra de una población determinada puedan ofrecer sobre el conjunto de esa población.

Hasta ahora el análisis cuantitativo que hemos llevado a cabo ha sido tan solo descriptivo, en el sentido de que los datos comentados se refieren exclusivamente a las ochenta y una encuestas realizadas. Ahora bien, este tipo de descripción solo se refiere a los datos que se tienen, dado que no se puede inducir a partir de ellos ninguna información referente a la población que se quiere estudiar, en este caso, los alumnos extranjeros que estudian español en la Universidad de Cádiz.

Un análisis cuantitativo basado en regresiones lineales, sin embargo, permite extrapolar los datos que se han obtenido a la población que se estudia, esto es, analizarlos como representativos del grupo sobre el que se lleva a cabo la investigación y del que los informantes son tan solo una muestra.

El análisis se ha llevado a cabo mediante el programa *stata*, paquete estadístico diseñado para el análisis de datos, con capacidades semejantes a otros programas como *spss*, aunque más versátil. Los datos que se le han introducido son el número de palabras que ha escrito cada persona en cada centro de interés, el total de palabras por encuesta y las características de cada informante relacionadas con las variables y variantes que se han tenido en cuenta.

La regresión lineal realizada sobre la variable “nivel de español” es la siguiente:

TABLA 2
REGRESIÓN LINEAL
SOBRE EL NIVEL
DE ESPAÑOL

Número de observaciones **81**
F (2, 78) **36.25**
Prob > F **0**
R² **0.4477**

	COEFICIENTE TOTAL	DESVIACIÓN ESTÁNDAR ROBUSTA	t	P > t	[95% INTERVALO DE CONFIANZA]	
n2	88.06349	12.72639	6.92	0	62.72719	113.3998
n3	119.4444	15.10957	7.91	0	89.3636	149.5253
_cons	198.2222	10.17272	19.49	0	177.9699	218.4746

$$\text{total} = 198.22 + 88.06n_2 + 119.44n_3$$

$$(10.17) (12.73) (15.11)$$

$$R^2 = 0.4477$$

$$F (2, 78) = 36.25 \text{ con valor } p \ 0$$

En ella n₂ y n₃ son variables ficticias, de las cuales n₂ toma el valor 1 si la persona encuestada está en el nivel B y 0 si no lo está, y n₃ toma el valor 1 si se sitúa en el C y 0 si no, de tal manera que, si ninguna toma el valor 1, el informante pertenece al nivel A.

La media de palabras escritas por los informantes del nivel A nos la proporciona la constante (*cons*), 198,22, con una desviación estándar de 10,17 palabras –es decir, con un margen medio de variación de algo más de 10 palabras-. En el nivel B se han registrado, de media, 88,06 palabras más que en A, con una desviación estándar de 12,73 palabras, y en el nivel C se han escrito 119,44 palabras más que en el nivel A, con una desviación estándar de 15,10 palabras, y, por tanto, 31,38 palabras más que en el nivel B. Contrastando la hipótesis nula (*H*₀) –hipótesis que apunta a que la diferencia entre las variantes sea en realidad 0– de que los coeficientes de n₂ y n₃ son iguales a 0 conjuntamente, es decir, contrastando que el nivel no tiene efecto, encontramos que la estadística *F* tiene un valor *p* igual a 0, lo cual rechaza la hipótesis nula. La probabilidad de que la diferencia entre las variantes no sea 0 es, por tanto, casi del 100%, es decir, esta variable es significativa para explicar la variación del número de palabras. Por otro lado, el coeficiente de determinación (*R*²) indica que este modelo explica el 44,77% de la variación, por lo que la variable “nivel de español”, además de ser significativa, es, por sí sola, capaz de explicar gran parte de la variación.

El análisis cuantitativo inductivo corrobora, pues, la intuición del descriptivo y permite aplicar los resultados a la población objeto de estudio, de tal manera que, a partir de esta observación, podemos afirmar que el nivel en el que los estudiantes se encuentran influye de manera determinante en la cantidad de palabras que conocen

y que son capaces de actualizar, pero no siempre de la misma manera, pues también en este análisis ha quedado de manifiesto que el crecimiento se ralentiza significativamente en los niveles más altos.

Conclusiones

Los trabajos sobre disponibilidad léxica ofrecen unas posibilidades de observación y análisis muy fecundas, como lo están demostrando los numerosos estudios que han aparecido en los últimos años que, aunque de diferente índole, utilizan esta metodología. La investigación que aquí hemos presentado entra a formar parte, en la medida de sus posibilidades, del colectivo de documentos versados sobre cuestiones de léxico disponible, aunque se diferencia claramente de la mayoría de ellos⁴ en ciertos elementos que hacen a nuestro trabajo novedoso, ya que no se realiza sobre una lengua materna, sino sobre una lengua extranjera: el español que aprenden los alumnos que vienen a estudiar a una universidad española. El estudio del que este artículo es fruto, pues, ha procurado contribuir a su desarrollo y perfeccionamiento mediante un mejor conocimiento del léxico que manejan los estudiantes extranjeros de español y de las variables sociolingüísticas que afectan a la comunidad estudiada, las cuales, por las características propias que muestra, son muy diferentes de las que influyen a los hablantes nativos.

Tras exponer brevemente las bases teóricas y metodológicas de este tipo de estudios, nos hemos centrado en el análisis de la incidencia que el nivel de español de los estudiantes extranjeros de nuestra lengua tiene sobre la media de palabras que conocen y son capaces de actualizar cuando se les propone un tema determinado. Los resultados obtenidos son, en un principio, relevantes: los informantes del nivel C -el más elevado- registraron un promedio de 2,04 palabras más que los del nivel B y 6,94 más que los del nivel A, esto es, un 10,96% más en el primer caso y un 63,03% más en el segundo; entre B y A la diferencia es de 4,9 palabras, lo que supone un crecimiento del 44,43%. Estos resultados, por tanto, confirman en principio la hipótesis de que el nivel de español de los alumnos influye en la cantidad de palabras que conocen, de tal manera que cuanto más alto sea el nivel, mayor es el número de palabras que manejan, con la particularidad de que este crecimiento no es homogéneo entre los tres estadios, sino mucho más pronunciado entre el nivel inferior y el medio que

⁴ Aunque escasos, existen algunos trabajos publicados sobre el léxico disponible de estudiantes de español. Carcedo (2000) lo lleva a cabo sobre alumnos finlandeses de enseñanza media y Samper Hernández (2002) sobre un grupo de estudiantes extranjeros de varias nacionalidades en la Universidad de Salamanca.

entre el medio y el superior.

Sin embargo, estos datos son producto de una descripción que no ha tenido en cuenta ni la desproporción en el número de informantes de los distintos grupos observados ni el posible solapamiento que se pueda producir entre los factores examinados. Esto quiere decir que los resultados del análisis descriptivo podrían no ser válidos para la población en general, pues podrían verse influidos por otros factores o por lo aleatorio de la muestra. El análisis de tipo inductivo que hemos llevado a cabo a continuación nos ha revelado, sin embargo, que el nivel de español sí explica por sí mismo gran parte de la variación que la muestra presenta respecto al número de palabras conocidas, es decir, que se trata de una variable significativa a la hora de determinar cuántos vocablos conocen los estudiantes de español como lengua extranjera.

Referencias

- 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)*. Turku, Turun Yliopisto.
- Gougenheim,, G.; Michéa, R.; Rivenc, P. & A. Sauvageot (1964), *L'élaboration du Français Fondamental (I Degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Paris, Didier.
- Grève, M. y Van Passel, F. (1971), *Lingüística y enseñanza de lenguas extranjeras*, Madrid, Fragua.
- López Chávez, J. & C. Strassburger (1991), "Un modelo para el cálculo del índice de disponibilidad léxica individual", López Morales, H. (ed.), *La enseñanza del español como lengua materna*, Río Piedras, Universidad de Puerto Rico, 91-112.
- López Morales, H. (1983), "Lingüística estadística", López Morales, H. (coord.), *Introducción a la lingüística actual*, Madrid, Playor, 209-225.
- Michéa, R. (1953), "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage", *Les Langues Modernes*, 47, 338-344.
- Raso Nadal, J. M.; Martín Vide, J. y Clavero Aparicio, P. (1987), *Estadística básica para ciencias sociales*, Barcelona, Ariel.
- Robles Ávila, S. (coord.) (2006), *La enseñanza del español como lengua extranjera a la luz del Marco Común Europeo de Referencia. Diseño curricular de los cursos para extranjeros de la Universidad de Málaga*, Málaga, Universidad de Málaga.
- Samper Hernández, M. (2002), *Disponibilidad léxica en alumnos de español como lengua extranjera*, Málaga, ASELE.
- Samper Padilla, J. A. (1998), "Criterios de edición del léxico disponible: sugerencias", *Lingüística*, 10, 311-333.