

DISPONIBILIDAD LÉXICA Y ESCRITURA DEL ESPAÑOL COMO LENGUA EXTRANJERA: PROPUESTA DE COMPARACIÓN DE DOS CORPUS

MARÍA LUISA HELEN FREY PEREYRA
UNIVERSIDAD DE ALCALÁ
luisahelenfrey@gmail.com

Resumen: *La disponibilidad léxica del español como lengua materna ha sido ampliamente estudiada. Si bien el léxico disponible del ELE ya ha sido sujeto también de algunos estudios, pocas veces se ha comparado con corpora que recojan otros aspectos de la interlengua de los alumnos de ELE. La primera parte de esta comunicación dará a conocer una línea de investigación en la que se están manejando, comparativamente, dos corpora: uno de disponibilidad léxica y otro de textos; ambos de alumnos de ELE. La segunda parte, explicará algunos problemas para el almacenamiento y manejo de los datos, y después describirá el diseño y funcionamiento de una base de datos que resuelve estos problemas facilitando el manejo de la información. Finalmente, se mencionarán algunas otras posibilidades de investigación con este tipo de corpora que aún no han sido abordadas.*

Palabras clave: léxico disponible, escritura, adquisición de léxico del ELE, bases de datos.

Abstract: *Lexical availability in native Spanish speakers has been widely studied, lexical availability of students of Spanish as a foreign language has also received some attention, however it has not yet been compared with corpus of other aspects of interlanguage. The first part of this paper will describe one research which uses two kinds of corpus: one of lexical availability and another of students writings, both of Spanish as a foreign language. The second part will explain some of the problems to store and manage de data. Then, I will describe the design and development of a database that solves these problems and makes data far more reachable. Finally, some other possibilities of research with this kind of corpus will be suggested.*

Key words: lexical availability, writing, vocabulary acquisition of Spanish as a foreign language, data base.

Esta comunicación tiene como objetivo principal dar a conocer una investigación en curso sobre léxico disponible y textos escritos en ELE que parte de las siguientes preguntas: ¿Qué porcentaje del léxico disponible usan los alumnos de ELE en sus escritos? ¿Cometen los mismos errores al escribir palabras en la prueba de léxico disponible que al escribirlas en contexto? Al responder a una segunda prueba de léxico disponible, ¿sus respuestas serán mismas o qué tipo de diferencias presentarán? Para explicar cómo esta investigación responderá a estas preguntas, resulta útil iniciar con una breve exposición sobre la evolución del estudio de disponibilidad léxica.

1. Introducción al estudio del léxico disponible

El estudio del léxico disponible se inicia en Francia en los años 50, cuando el grupo encargado de seleccionar vocabulario para un manual de enseñanza de lengua francesa dirigido a las excolonias, se dio cuenta de que en las listas de léxico frecuente, obtenidas a partir de la recolección de palabras que aparecían en textos escritos, no aparecían ciertas palabras muy conocidas y usadas por los hablantes nativos. Las investigaciones consecuentes de R. Michéa

(1953) llevaron a la distinción entre el léxico frecuente y el disponible. El léxico frecuente se definió como aquél que recogía "las palabras de mayor estabilidad estadística" (HUMBERTO LÓPEZ MORALES 1999: 8), mientras que el léxico disponible como el que se usa en una situación comunicativa específica, actualizándose sólo ante la necesidad de comunicar algo referente a esa situación concreta (H. LÓPEZ MORALES 1999: 9).

A partir de entonces empezó a recogerse el léxico disponible de comunidades determinadas aplicando pruebas de asociación que consisten, a grandes rasgos, en mencionar un centro de interés –por ejemplo: "las partes del cuerpo"– para que los informantes escriban en una lista las palabras que este estímulo les trae a la memoria durante un tiempo determinado. En un principio las listas fueron ordenadas por rangos de frecuencia, pero actualmente se procesan con un programa llamado *Lexidisp*, que pone en práctica una fórmula matemática diseñada por Juan López Chávez y Carlos Strassburger (2000) para considerar en el cálculo no sólo los grados de frecuencia sino también la disponibilidad. Es decir, además de tomar en cuenta el número de apariciones de cada vocablo, considera también el lugar que éste ocupa en la lista de cada informante. El resultado es una nómina de disponibilidad léxica de la comunidad de los sujetos que han respondido la encuesta. En ella aparece la disponibilidad de cada vocablo, el porcentaje de frecuencia, el porcentaje de aparición y la frecuencia acumulada. Con *Lexidisp* también es posible realizar comparaciones entre las nóminas léxicas de grupos de hablantes distintos, o bien dentro una misma nómina, pero comparando las variables que se hayan registrado, tales como sexo, nivel sociocultural o edad de los informantes.

Desde la publicación de *Français fondamental* (Gougenheim *et al.*, 1964) –el manual en cuya preparación iniciaron los trabajos sobre la disponibilidad léxica– se han realizado numerosas investigaciones con hablantes nativos de diversas lenguas y, en algunos casos, comparando la disponibilidad léxica de dos comunidades que comparten la misma lengua pero que se encuentran en dos regiones geográficas distintas. En la lingüística hispánica las investigaciones de disponibilidad léxica inician en los años 70 con los trabajos del puertorriqueño Humberto López Morales, los cuales posteriormente han dado lugar a una serie de trabajos en lugares de habla hispana como el Juan López Chávez en México, Alba Valencia y Max S. Echeverría en Chile, y un gran número de investigadores en España. La mayoría de ellos forman parte del Proyecto Panhispánico, coordinado por Humberto López Morales. Éste proyecto se ha propuesto la recolección de nóminas léxicas de varias regiones de habla hispana para la posterior realización de diccionarios dialectológicos.

También se han realizado ya, aunque en menor medida, estudios de disponibilidad léxica del ELE. Algunos de ellos comparan el vocabulario de manuales de enseñanza con ciertas nóminas léxicas, como los trabajos de Pedro Benítez (1994, 1997), Francisco García Marcos y María Victoria Mateo (1994), y Jenny Norman (1994). Alberto Carcedo González ha hecho numerosos estudios con alumnos de ELE finlandeses comparando sus resultados con los de la nómina léxica de Madrid (1999b), analizando errores léxicos (1999) y también abordando algunos aspectos cualitativos como "la otredad cultural" (2000). Marta Samper Hernández (2002), por su parte, ha trabajado con alumnos de ELE de diferentes nacionalidades en la Universidad de Salamanca y ha comparado sus resultados con algunos de los estudios de Carcedo.

2. Disponibilidad léxica y escritura del ELE: comparación en curso de dos corpus

La investigación¹ que realizo actualmente y que deseo presentar en esta comunicación se inscribe dentro del estudio del léxico disponible del ELE, pero se encuentra también vinculada con la adquisición de vocabulario y con la escritura, ya que el corpus que manejo, además de contar con las encuestas de léxico disponible necesarias para obtener una nómina inicial de disponibilidad léxica, incorpora también una serie de textos, de temática personal y no personal, escritos por los mismos informantes, y, finalmente, una segunda encuesta de léxico disponible.

Mediante el estudio de este corpus se busca cumplir los siguientes objetivos de análisis de vocabulario:

1.- Determinar en qué medida el léxico disponible de los alumnos de ELE se asemeja al que usan en escritos de temática personal y no personal para saber si el léxico disponible de esta población es mayoritariamente pasivo o activo.

2.- Analizar si el vocabulario adquirido en un contexto de inmersión se ve reflejado en las encuestas de léxico disponible y en los escritos de los alumnos. De ser así, se estudiará también en qué tipo de escritura se evidencia más la adquisición: en la personal o en la no personal.

3.- Analizar si los errores léxicos cometidos en las encuestas de adquisición son del mismo tipo que los cometidos en sus escritos.

El primer problema que plantea una investigación de este tipo consiste en la reunión de un corpus adecuado en el que deben considerarse un buen número de factores y variables. Los explicaré a continuación conforme exponga la manera en la que he elaborado tanto el corpus de léxico disponible como el de escritos.

2.1. La muestra y la recolección de datos

Para recolectar encuestas suficientes y obtener la nómina léxica de una población determinada, normalmente se acude a centros educativos que suelen facilitar la aplicación de la encuesta a los sujetos que cumplan con las características requeridas, por ejemplo: edad o grado de estudios. En cambio, para reunir el corpus de esta investigación no sólo era necesario aplicar una encuesta de léxico disponible, sino que era preciso reunir a un grupo de estudiantes de ELE que estuvieran dispuestos a responder la primera encuesta, a escribir en un tiempo determinado ciertos textos, a ceder esos textos para la investigación y, además, a responder una segunda encuesta de léxico disponible. Esta población se reunió de dos formas gracias a la colaboración Alcalingua, el centro de la Universidad de Alcalá encargado de los cursos de lengua y cultura españolas para extranjeros, que ha facilitado la impartición de los talleres con sus alumnos y en sus instalaciones. La primera fue invitando a un grupo de 6 alumnos estadounidenses, que tenía que tomar un curso de escritura para obtener ciertos créditos, a responder las encuestas de léxico y a ceder sus escritos. La segunda fue ofreciendo un par de talleres de escritura gratuitos dirigidos a los alumnos de niveles avanzado y superior.² En la primera sesión se explicaba que tendrían que responder a ciertas encuestas y que los textos que escribieran formarían parte del corpus de esta investigación. La mayoría de ellos accedió a ceder sus textos, pero no todos asistieron a todas las sesiones por lo que sus expedientes quedaron incompletos y fueron descartados para la investigación. Hasta ahora he logrado reunir el material de 15 alumnos de niveles avanzado y superior de los cuales 11 tienen como lengua materna el inglés y los cuatro restantes el francés, portugués, sueco y japonés respectivamente. Estos no han sido descartados, en favor de una muestra más uniforme debido a que en un futuro cercano planeo ampliar el corpus con el material de más alumnos con otras lenguas maternas.

2.1.1 Las variables

Al trabajar con pruebas de léxico disponible aplicadas a alumnos de ELE ha sido necesario considerar y resolver algunas variables que no habían sido previstas por la encuesta del Proyecto Panhispánico, que es en la que me he basado. Aunque se han recogido todas las variables que este Proyecto toma en cuenta (sexo, edad, lugar de nacimiento, estudios en instituciones públicas o privadas, la procedencia urbana o rural y nivel sociocultural), sólo la de sexo y edad se han incorporado a la presente investigación. Se han añadido, en cambio, los datos sobre nacionalidad, nivel y método de estudio de español, lengua materna y conocimiento de otras lenguas.

2.1.2 Las encuestas de léxico disponible

La encuesta de léxico disponible que respondieron los alumnos se conforma de dos partes. La primera consiste en una página recoge las variables antes mencionadas; la segunda parte consiste en cinco páginas, cada una con tres columnas de renglones numerados, sin los nombres de los centros de interés. Los números en las columnas son sólo para facilitar la localización de las palabras, ya que se trata de una prueba de listas abiertas en la que no se requería un número de palabras determinado. Los centros de interés no se encontraban al inicio de cada columna para evitar que los alumnos se adelantaran a escribir palabras del centro siguiente cuando hubieran terminado de escribir las del centro en turno, o bien que regresaran a columnas anteriores. Después de llenar la primera página, se les explicaba a los informantes que se les diría el nombre de un centro de interés y que ellos debían escribir en la columna correspondiente todas las palabras que recurrieran a su memoria, relacionadas con el centro que habían escuchado, sin prestar atención a la ortografía para que este aspecto no representara un motivo para no escribir alguna palabra o bien para que no se detuvieran a pensar en las reglas ortográficas, ya que esto podría, por el tiempo que pudiera tomar, resultar en detrimento de la cantidad de palabras escritas. El tiempo que los alumnos tenían para escribir palabras en cada centro fue de un minuto y medio, a diferencia de los dos minutos que maneja la mayoría de las encuestas de léxico disponible que se realizan actualmente. Esta reducción de tiempo se debe al comportamiento observado durante algunas encuestas realizadas a alumnos de ELE anteriormente. En ellas pude observar que a diferencia de los encuestados nativos que empiezan a escribir en cuanto se les dice el centro de interés, los alumnos de ELE tardaban entre 5 y 10 segundos en empezar y continuaban con periodos breves de descanso durante el primer minuto. Después dejaban de escribir y casi todos añadían alguna otra palabra al minuto y medio. En los últimos 30 segundos no todos añadían palabras. Como mi interés es recoger las palabras que resulten espontáneamente de un estímulo dado y no producto de una reflexión más detenida, decidí reducir el tiempo de respuesta a un minuto y medio.

Los centros de interés que he elegido para la encuesta son una selección de los que actualmente considera el Proyecto Panhispánico. He eliminado algunos centros que tenían poca probabilidad de verse reflejados en los textos escritos como es el caso de "trabajos de campo y de jardín" o "iluminación, calefacción y ventilación". Los centros que he conservado son los siguientes:

- | | |
|----------------------------|---------------------------------|
| 1. Las partes del cuerpo. | 8. Los animales. |
| 2. La ropa. | 9. Juegos y distracciones. |
| 3. Alimentos y bebidas | 10. Profesiones y oficios. |
| 4. La escuela ³ | 11. Colores. |
| 5. El campo. | 12. Sensaciones y sentimientos. |
| 6. La ciudad. | 13. Adjetivos. |
| 7. Medios de transporte. | 14. Medios de comunicación. |

2.1.3 La edición de los datos de las encuestas de léxico disponible

Una vez que se tienen las encuestas respondidas hay que transcribir y editar los datos para poder luego procesarlos en *Lexidisp*. En este caso, además de la edición normalmente requerida para unificar los datos y hacerlos comparables con otras nóminas léxicas, el manejo de listas de vocablos escritas por extranjeros requiere otro tipo de edición que contemple los casos específicos que no se dan en las listas de léxico de lengua materna, por ejemplo cuando un alumno escribe un extranjerismo adaptándolo al español, como "hacer *hiking*" donde añade a una palabra del inglés un verbo en español a semejanza de "hacer pesas" por ejemplo. Para la edición general seguí los lineamientos propuestos por Samper Padilla (1998), por ejemplo: la eliminación de términos repetidos o la corrección y unificación ortográfica. Para la edición específica del léxico de ELE seguí, la mayoría de las veces, los lineamientos de Samper Hernández (2002).

Sin embargo, este tipo de edición implica, casi siempre, cierta pérdida de información, pues es necesario regularizar ciertas palabras para poder considerarlas en las encuestas de léxico disponible. Para resolver este problema he diseñado una base de datos en la que es posible recoger todas las variables de las palabras que se han cambiado u omitido en las listas de léxico disponible. Un poco más adelante describiré el diseño y funcionamiento de esta base.

2.1.4 Los escritos

El segundo corpus se conforma por los escritos de los mismos alumnos que respondieron las pruebas de léxico disponible. Estos talleres, para no influenciar, al menos directamente, el vocabulario de los alumnos, estaba enfocado a la planeación de la estructura de un texto y a la localización y corrección de estructuras gramaticales fosilizadas. En ningún caso se facilitaron ni se estudiaron listas de vocabulario ni ningún otro material que pudiera influir directamente en el léxico de los informantes. Los alumnos de los tres cursos tienen en común dos textos escritos exactamente en las mismas condiciones. El primero de ellos fue realizado en la primera sesión del curso, en la que tenían 15 minutos para escribir, sin uso de diccionario, una composición titulada "Mis primeras impresiones de España". El segundo fue escrito en la última sesión del taller con las mismas condiciones y se trataba de una carta, de contenido libre, para algún familiar o amigo. Los alumnos escribieron otros textos además de los ya mencionados de temática personal o no personal, pero con títulos diferentes. Estos detalles se recogen también en el corpus para tenerlos en cuenta en los resultados de la investigación.

Este corpus, se ha almacenado de dos formas distintas: la primera es una transcripción íntegra, del texto, sin correcciones. Ésta recoge la mayor cantidad posible de información del original, incluyendo por ejemplo texto tachado o anotaciones al margen que se transcriben mediante un sistema convencional de signos. Este formato tiene la desventaja de que es sumamente difícil realizar búsquedas de vocablos precisos y hacer comparaciones entre ellos. Para resolver este problema, la segunda forma de almacenar este corpus consiste en la exportación, también sin correcciones, de todas las palabras de los textos a la misma base de datos que contiene las listas de léxico disponible.

2.2 La base de datos: almacenamiento y manejo de información

Una vez recopilado este corpus doble de léxico disponible y textos, es necesario plantear un instrumento adecuado para manejar y comparar la información obtenida. Para ello he diseñado una base de datos relacional que permite recoger, en una misma estructura, toda la información, esto es: las variables de los informantes, sus listas de léxico disponible, y sus escritos. Pero además de esta información, la base de datos permite recoger todas las variables de las palabras que por algún motivo fueron editadas u omitidas. Esta base de datos está estructurada en cuatro tablas independientes relacionadas entre sí:

1. La primera llamada "Alumnos" le asigna un código a cada informante y recoge sus datos: el taller al que asistió, sexo, nacionalidad, edad/generación, nivel de estudios de español, lengua materna, conocimiento de otras lenguas, tiempo de residencia en España y fecha de la prueba de léxico disponible. El registro de cada alumno cuenta también con un espacio para notas.
2. La segunda tabla llamada "vocabulario" se conforma de cinco campos: código del alumno, código del texto, vocablo, posición del vocablo en el texto y disortografía.
3. La tercera y cuarta tabla llamadas "léxico disponible 1" y "léxico disponible 2" respectivamente, tienen la misma estructura que la tabla anterior, pero en el campo de "código de texto" aparece el centro de interés al que pertenece el vocablo, y, en el campo de "posición" aparece el lugar que éste ocupa en la lista original de ese informante.

Tanto la tabla de "vocabulario" como las de "léxico disponible 1" y "léxico disponible 2", tienen, vinculado al campo de "vocablo" un campo más llamado "disortografías". Éste es el campo más complejo de todos y el que requiere de un trabajo más minucioso y personal ya que en él se registra la catalogación de errores. Por ejemplo, el vocablo "marrón" presenta la disortografía "maron" del centro de interés colores. En el campo de "disortografías" seleccionaremos las categorías "Ausencia de tilde en aguda" y "Error de letra o grafía r/rr". Para determinar los criterios de catalogación de errores hemos partido además del trabajo de Carcedo (1999), del de Florentino Paredes (1999) *La ortografía en las encuestas de disponibilidad léxica* en el cual contempla prácticamente todos los errores que un hablante nativo puede cometer, desde la ausencia de tildes hasta los errores de concepto ("fosforescente" por "fluorescente"). Aún así, he añadido algunas clasificaciones para casos que aparecen en las listas de los alumnos extranjeros y que no caben en ninguna de las clasificaciones propuestas por Paredes.

También en este campo se pueden registrar los "extranjerismos", los vocablos que han aparecido en un centro de interés sin tener relación con él, y finalmente los vocablos que no hemos logrado descifrar. De esta manera queda registrada toda la información al transcribir, ya sea las listas de léxico disponible, o los textos de los alumnos y, al mismo tiempo, se estructura la información de una manera que permite manejarla por medio de consultas.

Una vez expuesta la estructura mediante la cual se han almacenado estos datos es necesario regresar a los objetivos de la investigación para explicar cómo es que se pretende alcanzarlos. Para llevar a cabo el primer objetivo de la investigación –saber en qué medida el léxico disponible de los alumnos de ELE se asemeja al que usan en escritos de temática personal y no personal– se realizará una búsqueda para comparar las palabras en las encuestas de léxico disponible con las de los escritos de los alumnos. Ésta búsqueda mostrará cuántas y cuáles palabras son comunes a los dos corpus haciendo posible un análisis cuantitativo y cualitativo.

Para analizar si la adquisición de vocabulario de ELE en inmersión se ve reflejada en las encuestas de léxico disponible, se compararán la primera y segunda encuesta para encontrar los vocablos comunes y diferentes. Si en la segunda encuesta aparecen palabras nuevas, se evaluarán dos posibilidades: que se trate de palabras adquiridas en el periodo entre la primera y la segunda encuesta; o bien, que se trate de una desviación del léxico disponible.⁴ En ambos casos, los nuevos vocablos de la segunda encuesta se buscarán también en los textos de los alumnos para ver si éstos aparecen también en ese corpus y con qué frecuencia. Ésta búsqueda en los dos corpus permitirá ver si existe una relación entre dos tipos de vocabulario normalmente considerados distintos: uno como pasivo (el del léxico disponible) y el otro como activo (el de la escritura).

Finalmente se analizará si los errores que aparecen en la primera y la segunda encuesta son del mismo tipo y después se compararán con los errores que aparecen en los escritos de los alumnos, para determinar si los alumnos cometen los mismos tipos de errores en una prueba de asociación (lista de léxico disponible) que en contexto (sus escritos).

La importancia de almacenar los dos corpus en una misma base de datos radica no sólo en la cantidad de datos consultables, sino también en que los resultados de las búsquedas nos permitirán encontrar semejanzas y diferencias en dos contextos distintos de actuación de la interlengua, tanto en los errores como en los aciertos. Las búsquedas planeadas en esta investigación nos permitirán conocer mediante la comparación de la encuesta inicial y final de léxico disponible el grado de constancia o de variación tanto de los vocablos, como del total de palabras y de los errores. La comparación del corpus de léxico disponible con el de textos mostrará la relación entre estos dos vocabularios y nos permitirá determinar si las pruebas de léxico disponible pueden llegar a usarse como un instrumento para medir los niveles de adquisición de LE.

Esta investigación plantea sólo algunas de las posibilidades de estudio al comparar varios corpus que recogen diferentes aspectos de la interlengua. He querido plantear en esta ponencia cómo se pueden aprovechar no sólo los datos estadísticos que se obtienen de un conjunto de pruebas de léxico disponible, sino también la información que ofrece cada una de las pruebas,

ya sea analizando los errores que presenta, o bien vinculándola con los escritos del mismo informante.

Notas

¹ Esta investigación doctoral se realiza con el apoyo de de una beca del Consejo Nacional de Ciencia y Tecnología, México, y también forma parte del proyecto "Investigación y desarrollo de materiales didácticos para la enseñanza del español a inmigrantes y del español con fines específicos" (HUM 2005-07303-02-01), España.

² Tanto el curso como los talleres de escritura se impartieron entre los meses de agosto y noviembre de 2005.

³ Sin la especificación de "muebles y materiales".

⁴ Por desviación del léxico disponible me refiero a las diferencias que puedan presentarse entre la primera y la segunda encuesta, mismas que en su momento deberán calcularse a partir de la evaluación de los datos obtenidos.

Bibliografía

- BENÍTEZ PÉREZ, P., "Consideraciones en torno a la enseñanza del vocabulario", MORENO FERNÁNDEZ F. et al (coords.), *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*, Madrid, Arco Libros, 2003, vol. I, 145-156.
- , "Convergencia y divergencia en el léxico de alumnos de COU", *Revista de Estudios de Adquisición de la Lengua Española*, 1994, 39-45.
- , "El vocabulario enseñado en los manuales de ELE ¿es el adecuado?", *Problemas actuales en la enseñanza del español como lengua extranjera: gramática, pragmática, vocabulario y cultura*, Utrecht, UFSIA, 1997, 64-76.
- CARCEDO GONZÁLEZ, ALBERTO, "Análisis de errores léxicos del español en la interlengua de los finlandeses", T. JIMÉNEZ JULIÁ et al. (eds.), *Español como lengua extranjera: enfoque comunicativo y gramática*, Santiago de Compostela, Universidad de Santiago, 1999, 465-472.
- , "Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas", *Documentos de español actual*, 1999, 1, 73-87.
- , "La lengua como manifestación de otredad cultural (o convergencia intercultural)", AGUIRRE ROMERO, *Especulo. Monográfico "Cultura e Intercultura en la enseñanza del español como lengua extanjera"*, Madrid, Universidad Complutense de Madrid, 2000. Disponible en la red 25/abril/06: www.ucm.es/info/especulo/ele/carcedo.html
- GARCÍA MARCOS, F. Y MARÍA VICTORIA MATEO "Disponibilidad léxica y enseñanza de lenguas extranjeras", *Docencia e Investigación*, Universidad de Castilla-La Mancha, 1996.
- GOUGENHEIM, G., MICHÉA, RENÉ, RIVENC, PAUL Y SAUVAGEOT, AURÉLIEN, *L'elaboration du français fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de bas (Nouvelle édition refondue et augmenté)*, Paris, Didier, 1964.
- LÓPEZ CHAVES, JUAN, Y STRASSBURGUER CARLOS, "El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable", *Anuario de Letras*, 38, 2000, 227-251.
- LÓPEZ MORALES, HUMBERTO, *Léxico disponible de Puerto Rico*, Madrid, Arco Libros, 1999.
- MICHÉA, RENÉ, "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage", *Les langues modernes*, 47, 1953, 338-344.

NORMAN, JENNY, "El vocabulario español en los manuales suecos. Comparación entre el estudio de disponibilidad léxica y seis manuales suecos", *Revista de Adquisición de la Lengua Española*, 1994.

SAMPER HERNÁNDEZ, MARTA, *Disponibilidad léxica en alumnos de español como lengua extranjera*, Málaga, ASELE, 2002.