

Selección del léxico disponible: propuesta metodológica con fines didácticos

INMACULADA CLOTILDE SANTOS DÍAZ

Universidad de Málaga

Recibido: 2 octubre 2015 / Aceptado: 29 noviembre 2016

ISSN: 1697-7467

RESUMEN: La disponibilidad léxica está asociada desde sus orígenes a la selección del léxico para facilitar la adquisición y el aprendizaje de la lengua francesa (Gougenheim, Michéa, Rivenc y Sauvageot, 1956). Sin embargo, el cambio metodológico de listas abiertas propuesto por Dimitrijević (1969) conlleva un número demasiado elevado de lexías en los listados de palabras disponibles. La heterogeneidad en las pautas de selección léxica, basadas sobre todo en el índice de disponibilidad y la frecuencia de aparición, provoca que no haya consenso entre los investigadores. En este artículo, se realiza una propuesta metodológica para seleccionar el léxico disponible, correspondiente a un nivel de usuario básico del Marco Común Europeo de Referencia para las Lenguas, mediante la aplicación del modelo matemático de conjuntos difusos al índice de compatibilidad léxica.

Palabras clave: disponibilidad léxica, lingüística aplicada a la enseñanza de lenguas, competencia léxica, método de enseñanza

Methodological proposal for the selection of the available vocabulary

ABSTRACT: Lexical availability is associated from the beginning to the lexical selection to facilitate the acquisition and learning of French (Gougenheim, Michéa, Rivenc and Sauvageot, 1956). Nevertheless, the methodological change of opening lists proposed by Dimitrijević (1969) implies a high number of lexical units in the lists of available words. The heterogeneity in the selection rules, based on the level of availability and apparition frequency, provokes the lack of consensus between researchers. In this article, we present a methodological proposition to select the vocabulary –belonging to the level of basic user of the European Framework of Reference for Languages– thanks to the application of the mathematical model of fuzzy expected value to the level of lexical compatibility.

Keywords: lexical availability, applied linguistics to language teaching, lexical competence, teaching method

1. INTRODUCCIÓN

La disponibilidad léxica surge en Francia bajo la demanda del Ministerio de Educación Nacional en 1951 de crear un vocabulario de base para facilitar el aprendizaje de la lengua gala a inmigrantes y habitantes de los países que conformaban la Unión Francesa. Tras la publicación de varios volúmenes, en 1964 aparece una versión mejorada y ampliada, *L'élaboration du français fondamental (1^{er} degré)* (Gougenheim, Rivenc, Michéa y Sauvageot, 1964). Esta investigación servirá, no solo para enseñar ese vocabulario de base, sino también

para conformar en sí misma la base de un nuevo campo de estudio que cuenta en nuestros días con una vasta literatura y que, según Bartol Hernández (2006: 379-380), tiene su auge en la década de los 90 sobre todo debido a la creación de numerosos trabajos y equipos de investigación. La mayoría de las investigaciones se han realizado en español como lengua materna. El máximo exponente, Humberto López Morales (1973)¹, realiza una de las primeras investigaciones sobre el léxico en español analizando el léxico de San Juan de Puerto Rico.

La metodología de recogida de datos de la obra original (Gougenheim, Rivenc, Michéa y Sauvageot, 1964) consistía en solicitar a escolares que escribieran veinte palabras relacionadas con un centro de interés dado. Dimitrijević (1969) propone cambiar el sistema y recoger todas las palabras que se le viniesen a la mente de los informantes en un tiempo dado. A partir de entonces, esa pauta está presente en la mayoría de estudios de disponibilidad léxica en diferentes lenguas y países: Mena Osorio (1986) en Chile, Benítez Pérez (1992) en España, López Morales (1999) en Puerto Rico, Rodríguez (2007) en Canadá, entre otros. En el ámbito hispano, la aportación del Proyecto Panhispánico de Disponibilidad Léxica (PPHDL) supone la puesta en común de los aspectos metodológicos en estudios de zonas de habla hispana para la creación de diccionarios que muestren las diferentes sintopías y permitan su contraste. El tiempo estipulado en estos trabajos es de dos minutos por centro de interés.

Sin embargo, esta nueva metodología provoca que el diccionario del léxico disponible contenga un repertorio lingüístico demasiado elevado para fines relacionados con la enseñanza de una lengua. Uno de los investigadores que cuestionan la idoneidad de incluir todo el vocabulario de las encuestas es Mackey (1971: 39). En un estudio realizado en la región canadiense de Acadie, el autor analiza cuáles son los vocablos escritos por un solo informante. Concluye que de los 10 000 vocablos presentes en los dieciséis centros de interés tradicionales, más del 44 % habían sido aportados por una sola persona.

Ruiz Basto (1987: 97) establece la clasificación entre vocabulario activo y pasivo dividiéndolo entre vocablos con un índice de disponibilidad superior a 0,10 e inferior. Además, añade que para que el vocabulario se considere activo, hace falta que sea mencionado por al menos el 20 % de los informantes. De esta forma, la clasificación del léxico de escolares en centros educativos mexicanos la realiza tomando dos fuentes diferentes: el índice de disponibilidad y el porcentaje de aparición.

Este criterio ha sido utilizado y modificado en estudios posteriores. Así, Carcedo González (2000: 157-211) compara el léxico con un índice de disponibilidad superior o igual a 0,1 de cuatro comunidades hispanohablantes y de estudiantes finlandeses de español. En un estudio ulterior, Carcedo González (2001: 62) selecciona el vocabulario aportado en Asturias con un índice de 0,1 o superior y un porcentaje de aparición del 30 %.

Este rango de vocablos ha sido más laxo en investigaciones posteriores. Gómez Devís (2004: 125-127) concluye que, ante la disparidad de los resultados obtenidos según los criterios de frecuencia acumulada y el índice de disponibilidad léxica mencionados anteriormente, prefiere incluir las palabras que superan el índice de disponibilidad de 0,05. Bellón Fernández (2011: 182-283) establece el criterio de corte en el índice de disponibilidad igual o superior a 0,05. Por una parte, considera que en todos los centros de interés el número de vocablos

¹ El primer estudio en español lo realiza Bailey Victory en 1971 sobre una muestra de estudiantes de 16 a 18 años: 33 monolingües de Monterrey (México), 33 monolingües y 33 bilingües de Houston (Texas).

seleccionado es similar y estable. Por otra parte, menciona que las palabras resultantes suelen estar en torno a las 50 lexías idóneas propuestas por Alba Ovalle (1998, 2000). Además, esas palabras son habitualmente producidas por al menos un 10 % de los hablantes.

Samper Padilla (1999) restringe el criterio de selección al 75 %, del índice acumulado para comparar el léxico de Puerto Rico y Gran Canaria. De esta forma, considera que trabaja con una muestra representativa del vocabulario obviando palabras aportadas por pocos informantes en posiciones finales. En cambio, Bartol Hernández (2001: 227-230) amplía la muestra estableciendo como premisa que el índice de disponibilidad fuese igual o superior a 0,02 y hubiese sido mencionado por más de un informante. Asimismo, cuestiona la relevancia del criterio propuesto por Samper Padilla (1999) por estimar que es bastante restrictivo.

2. DISEÑO DE LA INVESTIGACIÓN

2.1. Objetivo

Este artículo tiene como objetivo establecer unas pautas comunes para la selección del léxico disponible y probar empíricamente su idoneidad. Para ello, propone y argumenta la elección de un novedoso criterio de selección del léxico disponible a partir de la aplicación del índice de compatibilidad léxica elaborado por Ávila Muñoz y Sánchez Sáez (2010, 2011). Este índice da cuenta del número de palabras aportadas por los informantes, de la posición en los listados y de su relación con el centro de interés. A continuación, se aplica el modelo matemático de conjuntos difusos y se establecen seis criterios de corte que muestran el grado de pertenencia del vocabulario al léxico aportado por el total de la población objeto de estudio.

2.2. Hipótesis

Las siguientes hipótesis se plantearon como elemento de conexión entre la teoría y la observación de los resultados:

- a. Los criterios de selección léxica presentes en estudios anteriores dan lugar a listados de palabras muy dispares.
- b. La teoría de los conjuntos difusos puede establecer criterios de corte del léxico que den lugar a un vocabulario correspondiente a un nivel A1 y A2 del MCERL.

2.3. Muestra

La población objeto de estudio está formada por 171 estudiantes de postgrado que cursan el Máster en Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas en la Universidad de Málaga durante el curso 2010/2011. Este postgrado habilita al ejercicio de la profesión docente en esos niveles educativos y exige como requisito previo la acreditación de un nivel B1 en una de las lenguas habladas en la Unión Europea diferente al español.

2.4. Recogida y edición de los datos

Este estudio ha utilizado el sistema de listas abiertas en el que los informantes tienen que escribir todas las palabras que se le vengan a la mente en un tiempo límite de dos minutos por cada centro de interés o estímulo. Se han seleccionado ocho centros de interés tradicionales y uno de nueva incorporación: *Partes del cuerpo* (C1); *La ropa* (C2); *Comidas y bebidas* (C3); *La escuela: muebles y materiales* (C4); *La ciudad* (C5); *Medios de transporte* (C6); *Juegos y distracciones* (C7); *Profesiones y oficios* (C8); *Ordenadores e internet* (C9).

2.5. Procesamiento matemático de los datos

El establecimiento de los criterios de corte del léxico se ha realizado a partir del índice de compatibilidad léxica creado por Ávila Muñoz y Sánchez Sáez (2010, 2011). Los autores se basan en la hipótesis de que el índice de disponibilidad léxica permite conocer y acceder al vocabulario del grupo concreto que ha realizado el estudio y, por ende, al prototipo léxico creado para cada centro de interés. Este hecho se observa al contrastar los índices de disponibilidad de una misma palabra en estudios diferentes ya que se ven alterados, como se expuso en la fórmula López Chávez y Strassburger Frías (1987), por el número de informantes, número de menciones y posiciones de la palabra.

A modo de ejemplo, la palabra *hígado* tiene una disponibilidad de 0,24165 en el estudio realizado en Málaga (Ávila Muñoz y Villena Ponsoda, 2010) a una población adulta estratificada mientras que esa misma palabra en esta investigación presenta una disponibilidad de 0,09071. En el primer estudio ese vocablo es más disponible porque la accesibilidad a la misma es mayor y, por tanto, ha venido a la mente del hablante en más ocasiones y en posiciones más adelantadas. Asimismo, la palabra se encuentra más cerca del núcleo o prototipo puesto que ocupa el rango 14 en el listado de palabras ordenadas según el índice de disponibilidad mientras que, en el segundo caso, ocupa el puesto 39. La diferencia en el porcentaje de aparición de ambos términos es igualmente significativa: ya que el vocablo había sido mencionado casi por la mitad de los hablantes en el primer estudio mientras que en el estudio actual no llega a una cuarta parte.

La estructura del léxico en cada centro de interés vendrá determinada por la asociación entre las diferentes unidades léxicas y su grado de accesibilidad. Según Ávila Muñoz y Sánchez Sáez (2011: 257) para medir la compatibilidad de un vocablo hace falta considerar cada uno de los vocablos aportados por los informantes y el concepto que para la comunidad estudiada representa cada uno de los centros de interés. Para establecer diferentes grados de compatibilidad o accesibilidad entre un elemento y su conjunto se basan en el concepto matemático de ‘conjunto difuso’.

El primer paso para calcular el índice de compatibilidad consiste en determinar el espectro de un centro de interés mediante un conjunto difuso a través de un modelo cuantitativo similar al de la ley de Zipf. En este caso, (t) hace referencia a la valoración de cada término obtenida a partir de la división del número que ocupa en la lista un vocablo (n) y una constante para cada caso (k) que representa el tamaño relativo de los conjuntos de términos que superan un nivel de compatibilidad.

$$t = \frac{k}{n}$$

A continuación, se construye el modelo colectivo basado en el valor de nula disponibilidad para todos los términos a partir de la siguiente ley probabilística:

$$a + b - a \cdot b$$

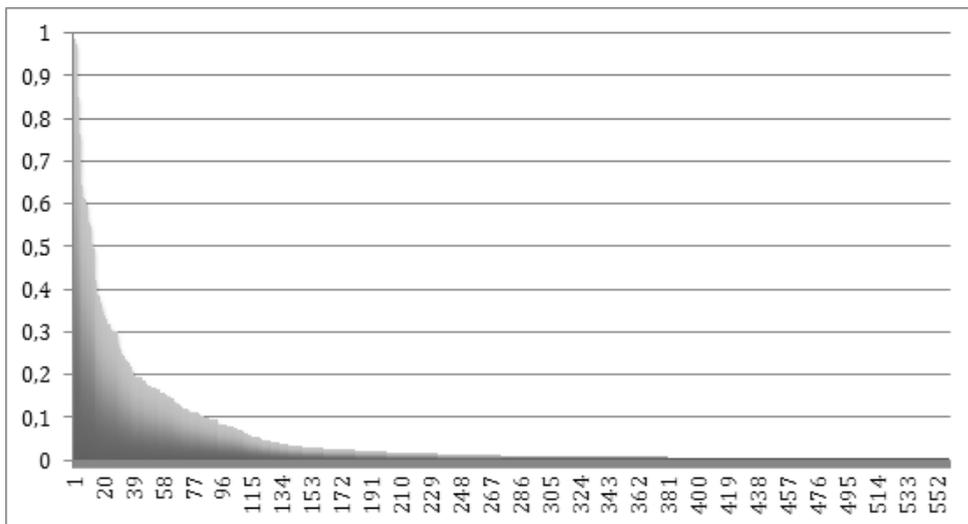
Donde

a= se refiere a la disponibilidad de un término

b= indica la disponibilidad personal de ese término

El índice de compatibilidad muestra la accesibilidad de los vocablos en un rango que oscila entre 0 y 1. En el Gráfico 1 las diferentes palabras corresponden a valores numéricos ordenados según el índice de compatibilidad. Los vocablos más compatibles se acercan al valor 1 mientras que aquellos cercanos a 0 han sido aportados por un número reducido de informantes y en posiciones poco relevantes. En este sentido, se constata que a un cierto punto las unidades léxicas pierden compatibilidad y apenas son accesibles.

Gráfico 1. Número de vocablos ordenados según el índice de compatibilidad en español. Centro de interés 09. Ordenadores e internet.



El índice obtenido al aplicar el valor del conjunto difuso sirve de base para la realización de los análisis posteriores que permiten seleccionar el léxico más accesible y conocer la compatibilidad de cada vocablo de la secuencia léxica aportada por un informante. Esa selección del vocabulario vendrá dada por criterios de corte que establecerán, por niveles, la mayor o menor pertenencia de los vocablos al núcleo, es decir, la prototipicidad de los mismos.

3. ANÁLISIS DE LOS RESULTADOS

3.1. Criterios tradicionales

Al igual que en estudios anteriores (Mackey, 1971), el porcentaje de palabras que han sido actualizadas por un solo informante supera el 40 %. La diferencia entre los tres idiomas es poco significativa y está influenciada por el número de informantes en cada caso ya que todos han realizado la prueba en español, 150 en inglés y 21 en francés. Se observa que a mayor número de informantes menor proporción de vocablos aportados una sola vez. Las unidades léxicas mencionadas por un único sujeto son fácilmente identificables gracias al porcentaje de aparición: en español 0,58 %, en inglés 0,67 % y en francés 4,76 %.

Tabla 1. Vocablos aportados por un único informante y su proporción en el centro de interés con respecto al total de vocablos en lengua española

CI	Español			Inglés			Francés		
	Total	UI	%	Total	UI	%	Total	UI	%
C1	244	69	28,28	115	38	33,04	78	29	37,18
C2	238	84	35,29	151	58	38,41	80	32	40
C3	530	207	39,06	287	117	40,77	166	84	50,6
C4	392	192	48,98	244	112	45,9	113	64	56,64
C5	483	220	45,55	343	169	49,27	135	61	45,19
C6	190	62	32,63	165	85	51,52	53	26	49,06
C7	469	207	44,14	310	134	43,23	125	75	60
C8	478	192	40,17	280	128	45,71	119	66	55,46
C9	555	286	51,53	460	253	55	127	72	56,69
Total	3579	1519		2355	1094	402,85	996	509	450,82
Media	397,67	168,78	42,44	261,67	121,56	44,76	221,33	56,56	50,09

A continuación, se analiza la selección del léxico teniendo en cuenta el índice de disponibilidad, la frecuencia acumulada y el porcentaje de aparición de cada vocablo. Las cuatro primeras columnas de la Tabla 1 seleccionan el léxico a partir de diferentes niveles de corte basados en el índice de disponibilidad léxica presentes en investigaciones previas: índice igual o superior a 0,1 en Carcedo González (2000), a 0,05 en Gómez Devís (2004) y Bellón Fernández (2011), a 0,02 y 0,01 en Bartol Hernández (2001). La cuarta columna se basa en el criterio propuesto por Carcedo González (2001) que selecciona el léxico con un índice de disponibilidad igual o superior a 0,10 y un porcentaje de aparición del 30 % de los informantes. La quinta columna recoge todas aquellas palabras con una frecuencia

acumulada del 75 % (Samper Padilla, 1999). Por último, se ha incluido el total de vocablos aportados en todos los centros de interés.

El hecho de que se hayan presentado en tres idiomas y con un número de informantes variable permite detectar qué factores influyen en esta selección léxica. De forma general, se observa que el porcentaje de vocablos incluidos en cada clasificación varía de forma considerable, lo que daría lugar a comparaciones muy diferentes según el criterio empleado. El apartado que cuenta con menos palabras es el de frecuencia acumulada mientras que los criterios del índice de disponibilidad son más integradores e incluyen una mayor cantidad de vocablos.

En español, el número de unidades léxicas extraídas con un 75 % de frecuencia acumulada solo cuenta con 90 vocablos de los 3579 que los informantes han actualizado, lo que representa un total del 2,51 %. En el lado opuesto, estarían los vocablos según el índice de frecuencia. Según el criterio menos restrictivo, el cual recoge las unidades léxicas con un índice mayor o igual al 0,01, la proporción de vocablos seleccionados asciende al 39,15 % mientras que la selección por medio del índice igual o superior a 0,10 se reduce al 8,27 % del total de vocablos. Si además, se considera el criterio del 30 % de los informantes, solo se incluiría el 4,30 % de las unidades léxicas.

Tabla 2. Vocablos en español, inglés y francés según el índice de disponibilidad, frecuencia acumulada y porcentaje de aparición

Idioma		ID ≥ 0,10	ID ≥ 0,05	ID ≥ 0,02	ID ≥ 0,01	ID ≥ 0,10 y 30 % informantes	Frecuencia Acumulada 75 %	Vocablos
Español	Total	296	527	928	1401	154	90	3579
	Media	32,89	58,56	103,11	155,67	17,11	10	397,67
	Porcentaje	8,27 %	14,72 %	25,93 %	39,15 %	4,30 %	2,51 %	100 %
Inglés	Total	202	355	600	917	99	62	2355
	Media	22,44	39,44	66,67	101,89	11	6,89	261,67
	Porcentaje	8,58 %	15,07 %	25,48 %	38,94 %	4,20 %	2,63 %	100 %
Francés	Total	230	395	778	959	113	604	996
	Media	25,56	43,89	86,44	106,56	12,56	67,11	110,67
	Porcentaje	23,09 %	39,66 %	78,11 %	96,29 %	11,35 %	60,64 %	100 %

En inglés, los datos referidos a los vocablos son inferiores no solo por contar con 21 informantes menos, sino por tratarse de una lengua extranjera. No obstante, el porcentaje en cada apartado es muy similar al español. Los centros más compactos cuentan con un menor número de vocablos y, por ende, también son menos las unidades léxicas seleccionadas según los criterios aplicados.

Las pruebas en francés han sido realizadas por 21 informantes lo que influye en todos los resultados. Así, el porcentaje de vocablos aportados por el 30 % de los informantes as-

ciende a 11,35 % mientras que en español es de tan solo el 4,30 % y en inglés el 4,20 %. Lo mismo ocurre con la diferencia porcentual marcada por el índice de disponibilidad, oscilando entre 23,09 % y el 96,29 %. Como la muestra es muy pequeña, muchos vocablos han sido aportados por un único individuo. De hecho, el francés cuenta en el apartado anterior con el porcentaje más amplio en este sentido. Por tanto, si se mantiene ese criterio, el número de vocablos se reduciría considerablemente.

3.2. Selección a partir de la teoría de los conjuntos difusos

La teoría de los conjuntos difusos, *Fuzzy Expected Value*, desarrollada por Zadeh (1965)², explica el nivel de pertenencia de algo a un conjunto en un continuo gradual. Ha sido aplicada a numerosos campos (ingeniería, medicina, biología) incluida a los estudios de disponibilidad léxica (Ávila Muñoz y Sánchez Sáez, 2010, 2011; Ávila Muñoz y Villena Ponsoda, 2010, 2011). A grandes rasgos, su cálculo permite diferenciar en varios niveles la compatibilidad del léxico con respecto al centro de interés. Cada nivel recoge un número reducido de vocablos que aumenta a medida que disminuye el grado de compatibilidad. Además, esos parámetros permiten identificar claramente cuáles son las unidades léxicas que cuentan con una alta disponibilidad en los informantes y, por tanto, que no deberían faltar en un diccionario.

Este FEV es capaz de proponer un valor de pertenencia que establece un equilibrio entre el número de términos que lo supera y el propio valor. Además, si ponderamos en diferente grado la relevancia del tamaño del conjunto de corte, se pueden establecer grados de restricción en su interior que nos permitirían construir conceptos como «muy representativo» o «poco representativo» (Ávila Muñoz y Villena Ponsoda, 2010: 63).

Los resultados obtenidos permiten diferenciar hasta seis niveles de compatibilidad con el área temática. Cuanto más se acercan los valores a 1, mayor es su representatividad en el área temática. El corte en niveles en cada centro vendrá determinado por el FEV y variará en función del estudio en cuestión y del tipo de centro analizado.

En lengua española, de los 244 vocablos actualizados en el primer centro de interés, solo 43 forman parte de los seis niveles considerados centrales, es decir, un 17,62 %. En ese mismo centro, en inglés han sido seleccionados 26 vocablos de los 115 aportados, lo que corresponde a un 22,61 % del total. En francés, ese porcentaje es similar, concretamente un 28,21 %, ya que de las 78 léxias diferentes se han extraído 22. Estos datos son de gran utilidad en la selección del léxico, sobre todo a nivel pedagógico, ya que estas serían las primeras palabras en relación al cuerpo humano que el estudiante debería aprender.

3.3. Léxico nuclear en los diferentes idiomas del estudio

El léxico nuclear o más compatible varía en función de las respuestas de cada muestra. El listado en español es fruto del uso de la lengua en un contexto natural mientras que, en

² A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. The notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of fuzzy sets are established. In particular, a separation theorem for convex fuzzy sets is proved without requiring that the fuzzy sets be disjoint (Zadeh, 1965: 338).

la mayoría de las ocasiones, el léxico aportado en inglés y francés ha sido adquirido en el ámbito académico. A continuación, se presentan los listados con el léxico nuclear del centro de interés *Comidas y bebidas* en las tres lenguas para comparar los resultados.

Su selección se debe al marcado carácter sociocultural presente en este centro de interés con respecto a otros, como puede ser *Partes del cuerpo humano*. Como muestra de ello, el inventario denominado *Saberes y comportamientos socioculturales* del PCIC dedica un apartado a las comidas y bebidas destacando no solo las unidades léxicas concretas que lleva este centro asociadas (*tapa*, entre otras) sino también los conocimientos específicos relativos al comportamiento compartido por los hablantes. Por ejemplo, qué es la merienda y cuándo se toma, cuál es el horario de cada comida, cómo se organiza el menú y, en definitiva, cuáles son las convenciones sociales y los comportamientos en la mesa.

El análisis del léxico nuclear se presenta en español, inglés y francés para comprobar si el alumno tiende a reproducir las mismas palabras en lengua materna y lengua extranjera o se ve influido por otros contextos de aprendizaje, como puede ser el manual de aprendizaje. Dado que el número de vocablos aportados es superior en español con respecto al inglés y el francés, también lo son las unidades léxicas nucleares. Este hecho se produce como consecuencia del mayor número de informantes (ya que todos han contestado en español) y por tratarse de la lengua materna. Por contraposición, la proporción de los vocablos con respecto al conjunto es inferior en español así como el nivel de compatibilidad que delimita el léxico más compatible.

Tabla 3. Léxico nuclear en español, inglés y francés del centro de interés 03. Comidas y bebidas

Nivel	Idioma	Compatibilidad	%	Vocablos	
1	Español	0,84	0,19 %	1	agua
	Inglés	0,86	0,7 %	2	water, meat
	Francés	0,85	0,6 %	1	eau
2	Español	0,76	1,13 %	6	cerveza, carne, pan, Coca-Cola, zumo
	Inglés	0,77	1,39 %	4	Coke, fish
	Francés	0,79	1,81 %	3	fromage, chocolat
3	Español	0,62	2,07 %	11	leche, pescado, patata, vino, café
	Inglés	0,65	2,79 %	8	wine, potato, tomato, beer
	Francés	0,65	3,01 %	5	omelette, bière
4	Español	0,46	4,33 %	23	tomate, pasta, arroz, lechuga, paella, refresco, fruta, huevo, ensalada, lenteja, chocolate, verdura
	Inglés	0,51	6,62 %	19	milk, orange, hamburger, apple, pizza, chip , bread, vegetable, egg, rice, juice
	Francés	0,53	7,83 %	13	jambon, lait, croissant , poisson, vin, pâte, pain, frite

Tabla 3. Léxico nuclear en español, inglés y francés del centro de interés 03.
Comidas y bebidas. (Continuación)

5	Español	0,28	7,35 %	39	manzana, tortilla, naranja, té, sopa, pollo, pizza, hamburguesa, potaje, filete, puchero, Fanta, espagueti, tortilla de patatas, bocadillo, plátano
	Inglés	0,34	11,15 %	32	coffee, banana, carrot, soup, chicken, cake, Coca-Cola, fruit, lemon, tea, salad, lettuce, onion
	Francés	0,4	15,66 %	26	boeuf, oignon, canard , pomme, viande, thé, choucroute , citronnade , dîner , fruit, pomme de terre, orange, Coca-Cola
6	Español	0,14	13,74 %	73	macarrón, batido, zanahoria, jamón, galleta, ternera, limón, cebolla, güisqui, tarta, queso, fresa, legumbre, melón, atún, ron, pimienta, cereal, pera, aceite, chorizo, cerdo, sal, cocido, gazpacho, salchicha, calabaza, maíz, sandía, mantequilla, garbanzo, solomillo, azúcar, yogur
	Inglés	0,18	17,42 %	50	spaghetti, sandwich , beef, cheese, chocolate, bean , pasta, watermelon, sugar, pineapple , omelet, steak, whisky, strawberry, salt, pepper , soda, butter
	Francés	0,23	22,29 %	37	vin blanc , pizza, crêpe , café, confiture , champagne , beurre, œuf, poulet, légume, salade

El listado en lengua española comprende palabras que no están presentes en inglés ni en francés y viceversa. En inglés, aparecen cinco términos que no se encuentran en español:

- chip* que hace referencia a las patatas fritas;
- sandwich* podría considerarse que se encuentra en su forma española *bocadillo*. En cambio, se ha marcado porque en español *sándwich* se refiere también al tipo de bocadillo que se prepara con pan de molde;
- bean* que tiene mayor relevancia en la dieta anglosajona;
- pineapple* puede ser más frecuente porque su venta suele ser en almíbar y el envase incluye la traducción en varios idiomas (a diferencia de otras frutas y verduras);
- pepper* suele enseñarse al mismo tiempo que *salt* aunque su uso no sea tan frecuente, razón por la cual en español no aparece en el listado.

En francés, a pesar de la mayor similitud lingüística, existen diez vocablos que no están incluidos en el listado en español. Cinco vocablos son de procedencia francesa: *champagne*,

crêpe y *croissant*, palabras típicas de la gastronomía francesa que forman parte del léxico disponible español pero con un menor grado de compatibilidad; *chocroute*, que no aparece en español; *confiture*, que aparece como *mermelada* pero podría traducirse también como *confitura*. En inglés aparece la palabra *chip* que hace referencia a las patatas fritas envasadas y a las patatas fritas tiernas. En cambio, en francés aparece *frite*, que quiere decir *patata frita* pero que solo se refiere a las patatas fritas tiernas y normalmente alargadas ya que para las patatas fritas envasadas utilizan el término de procedencia inglesa, *chip*. Las otras tres palabras aparecen en el léxico disponible en español pero no muestran la compatibilidad necesaria para ser seleccionadas en el léxico nuclear: *citronnade*, *dîner* y *canard*.

En el ámbito de la enseñanza de español o francés como segunda lengua o lengua extranjera, resultaría interesante enseñar el vocablo *pato* y *canard* incluso a alumnos principiantes por dos razones: por ser más frecuente su consumo en Francia y por tratarse de un falso amigo. Si en el léxico disponible no se enseña *canard* como *pato*, el estudiante español cuando viaje a puede pedir un plato de *pato* creyendo que está pidiendo *carne* y viceversa. De hecho, probablemente sea la razón que explique una compatibilidad similar de *canard* y *carne*.

3.4. Léxico nuclear en el Plan Curricular del Instituto Cervantes

Las nociones específicas del Plan Curricular del Instituto Cervantes (PCIC) incorporan un tema denominado *Alimentación* que se subdivide a su vez en siete apartados: *dieta y nutrición*; *bebida*; *alimentos*; *recetas*; *platos*; *utensilios de cocina y mesa*; y *restaurante*. Los dos subcampos relacionados con el centro de interés *Comidas y bebidas* serían el segundo y tercero (*bebidas y alimentos*) cuyas unidades léxicas, expresiones idiomáticas y colocaciones van acompañadas del nivel establecido en el MCER.

En el inventario de saberes y conocimientos socioculturales se incluye información adicional que complementa las nociones específicas. Por ejemplo, aparecen algunos platos típicos por zonas o regiones con independencia del nivel del MCER: *gazpacho* (Andalucía), *fabada* (Asturias), *migas* (Castilla-La Mancha), *bacalao al pilpil* (País Vasco), *paella* (Valencia); el nombre de los condimentos, hierbas y especias más utilizadas en la cocina española (*aceite de oliva*, *sal*, *vinagre*, *pimentón*, *ajo*, *azafrán* y *perejil*); los tipos de platos según su forma de cocina (*cocido*, *guisado*, *a la plancha*, *asado*, *frito*); entre otras nociones que ayudan a mejorar la competencia comunicativa del hablante en este tema concreto.

La Tabla 4 presenta el léxico más prototípico del centro de interés *Comidas y bebidas* organizado según los seis niveles de compatibilidad. A la derecha se ha incluido el nivel según el MCER y el tema específico que asigna el PCIC. Se han marcado en gris claro los niveles A1 y A2 y en gris oscuro los niveles B1 y B2. No hay ninguna palabra que se corresponda con los niveles C1 y C2.

Tabla 4. Nivel de referencia del léxico nuclear según el PCIC

Nivel	Vocablo	Compatibilidad	Nivel del MCER	Noción específica del PCIC
Nivel 1	agua	0,955196	A1	5.2. Bebida
Nivel 2	carne	0,81709	A1	5.3. Alimentos
	cerveza	0,835115	A1	5.2. Bebida
	Coca-Cola	0,787496		
	pan	0,791448	A1	5.3. Alimentos
	zumos	0,766884	A2	5.2. Bebida
Nivel 3	café	0,631503	A1	5.2. Bebida
	leche	0,736062	A1	5.2. Bebida
	patata	0,682504	A2	5.3. Alimentos
	pescado	0,699956	A1	5.3. Alimentos
	vino	0,675749	A1	5.2. Bebida
Nivel 4	arroz	0,569067	A2	5.3. Alimentos
	chocolate	0,465749	A2	5.3. Alimentos
	ensalada	0,501331	A1	5.5. Platos
	fruta	0,515579	A1	5.3. Alimentos
	huevo	0,512949	A1	5.3. Alimentos
	lechuga	0,53807	A2	5.3. Alimentos
	lenteja	0,470055	B1	5.3. Alimentos
	paella	0,532476	A1	5.5. Platos
	pasta	0,584875	A2	5.3. Alimentos
	refresco	0,526194	B2	5.2. Bebida
	tomate	0,615376	A2	5.3. Alimentos
	verdura	0,461414	A1	5.3. Alimentos
	Nivel 5	bocadillo	0,284398	A1
espagueti		0,2944		
Fanta		0,297434		
filete		0,317963	A2	5.3. Alimentos
hamburguesa		0,327463	A1	5.3. Alimentos
manzana		0,444405	A2	5.3. Alimentos
naranja		0,401905	A2	5.3. Alimentos
pizza		0,357435		
plátano		0,282128	A2	5.3. Alimentos
pollo		0,358437	A2	5.3. Alimentos
potaje		0,326473		
puchero		0,314485		
sopa		0,369157	A1	5.5. Platos
té		0,376824	A1	5.2. Bebida
tortilla		0,432196	A1	5.5. Platos
tortilla de patatas	0,288065	En el nivel A1: aparece tortilla		

Tabla 4. Nivel de referencia del léxico nuclear según el PCIC. (Continuación)

Nivel 5	bocadillo	0,284398	A1	5.3. Alimentos
	espagueti	0,2944		
	Fanta	0,297434		
	filete	0,317963	A2	5.3. Alimentos
	hamburguesa	0,327463	A1	5.3. Alimentos
	manzana	0,444405	A2	5.3. Alimentos
	naranja	0,401905	A2	5.3. Alimentos
	pizza	0,357435		
	plátano	0,282128	A2	5.3. Alimentos
	pollo	0,358437	A2	5.3. Alimentos
	potaje	0,326473		
	puchero	0,314485		
	sopa	0,369157	A1	5.5. Platos
	té	0,376824	A1	5.2. Bebida
tortilla	0,432196	A1	5.5. Platos	
tortilla de patatas	0,288065	En el nivel A1: aparece tortilla		
Nivel 6	aceite	0,196677	B1	5.3. Alimentos
	atún	0,206803	A2	5.3. Alimentos
	azúcar	0,140417	A2	5.3. Alimentos
	batido	0,274646		
	calabaza	0,172616		
	cebolla	0,234217	A2	5.3. Alimentos
	cerdo	0,18687	A2	5.3. Alimentos
	cereal	0,202538	A2	5.3. Alimentos
	chorizo	0,188728	B1	5.3. Alimentos
	cocido	0,180492		
	fresa	0,228662	En el nivel A2 aparece helado de fresa	
	galleta	0,256655	A2	5.3. Alimentos
	garbanzo	0,148535	B1	5.3. Alimentos
	gaspacho	0,180183	Inventario de referentes culturales. Apartado 1.4.1. Platos típicos de las demarcaciones territoriales. <i>Gaspacho</i> (España)	
	güisqui	0,230566		
	jamón	0,256852	A2	5.3. Alimentos
	legumbre	0,214103	B2	5.3. Alimentos
	limón	0,24784	En el nivel C1 aparece corteza de limón	
		macarrón	0,276936	
maíz		0,165424	En el nivel B2 aparece harina de maíz	
mantequilla		0,153964	A2	5.3. Alimentos
melón		0,207278		
pera		0,196799		
pimiento		0,204132		
queso		0,228864	A2	5.3. Alimentos
ron		0,204901	A2	5.2. Bebida
sal		0,183262	A2	5.3. Alimentos
salchicha		0,173112		
sandía		0,162259		
solomillo		0,140622	B1	5.3. Alimentos
tarta		0,229348	A2	5.3. Alimentos
ternera		0,248955	A2	5.3. Alimentos
yogur		0,140062	A2	5.3. Alimentos
zanahoria	0,263661	A2	5.3. Alimentos	

A simple vista se observa la relación del léxico compatible con el nivel de usuario básico (A1 y A2) según el MCER. Tan solo se destacan las siguientes unidades léxicas correspondientes a niveles superiores:

– Nivel B1: *lenteja, aceite, chorizo, garbanzo* y solomillo.

– Nivel B2: *refresco* y *legumbre*.

– El resto de palabras que no se encuentran marcadas en ningún color no aparecen clasificadas en el PCIC.

– Nombres de marcas: *Coca-Cola* y *Fanta*. Cabe destacar que en el apartado dedicado a la red de transportes aparecen los nombres propios *IBERIA* y *RENFE* (Red Nacional de Ferrocarriles Españoles). En cambio, en la alimentación han obviado los nombres de marcas con independencia de la lexicalización de las mismas.

–Vocablos concretos: *espagueti, pizza, potaje, puchero, batido, calabaza, cocido, güisqui, macarrón, pera, pimienta, salchicha, sandía*.

–Unidades léxicas que aparecen de forma parcial en el inventario: *tortilla* se localiza en el nivel A1 pero no aparece *tortilla de patatas*; *fresa* se ha marcado en gris claro porque en el nivel A2 aparece la entrada *helado de fresa*; *gazpacho* aparece como un plato típico en el inventario de referentes culturales sin indicar el nivel según el MCER; *limón* solo aparece con la locución *corteza de limón* en el nivel C1; *maíz* se incluye en el nivel B2 como *harina de maíz*.

Del cotejo del léxico compatible surgen los siguientes interrogantes: ¿Realmente aceite o refresco son vocablos que corresponden a un usuario competente? ¿Deberían enseñarse en niveles inferiores? ¿Cómo reconoce un docente el nivel correspondiente a una palabra que no aparece en el PCIC? ¿Cuál es la relación más frecuente entre las palabras de un mismo ámbito y nivel?

4. CONCLUSIONES

Las preguntas abiertas precedentes sirven para fundamentar la necesidad de la selección del léxico disponible con las bases teóricas y empíricas expuestas anteriormente para poder dar respuesta a las necesidades detectadas en la práctica docente. Resulta de gran interés para un profesional de la enseñanza de idiomas conocer a qué nivel del MCER corresponde cada vocablo. Un hablante nativo puede asociar una palabra a un nivel básico porque se emplee con frecuencia en su comunidad de hablante y un docente no hispanohablante puede considerar una palabra más o menos fundamental por su influencia con la lengua materna o por el manual seleccionado y ninguna de ellas pertenecer a un nivel de usuario básico.

Tras realizar un cotejo de los listados del léxico nuclear o más compatible en los tres idiomas estudiados, se ha comprobado que a pesar de que el número de lexías nucleares se vea influenciado por el número de informantes, existe una gran similitud entre el léxico nuclear de los listados en lengua materna y lengua extranjera. En inglés, tan solo se incluyen cinco palabras que no están presentes en el listado español y en francés se mencionan diez diferentes. No obstante, la mayoría aparece en el diccionario de léxico disponible en español pero en posiciones inferiores y con un índice de compatibilidad más bajo. Estos resultados constatan la presencia de un prototipo o núcleo en cada centro de interés con independencia del idioma de estudio. También se corrobora la incidencia de los factores socioculturales en

los listados del léxico, con palabras más propias de una cultura gastronómica que de otra: *choucroute* en francés, *bean* en inglés o *paella* en español.

Para conocer la adecuación del índice de compatibilidad en la selección del vocabulario disponible, se ha localizado el nivel según el MCER al que corresponde cada unidad léxica que conforma el listado de léxico compatible del centro de interés *Comidas y bebidas*. Como era de esperar, la mayoría de vocablos pertenecen a los niveles de usuario básico (A1 y A2) confirmando la hipótesis que el léxico nuclear fija los límites del vocabulario básico de una lengua. En cambio, existen ciertas palabras como *aceite* que el PCIC propone en un nivel B1 y cuenta con una compatibilidad equiparable a un A2. El PCIC se basa en clasificaciones realizadas en otros idiomas y, por tanto, herederas de otras culturas culinarias donde el aceite quizás no juega el mismo papel que en la gastronomía española.

El siguiente paso para hacer del léxico disponible una herramienta clave para la selección de las unidades léxicas presentes en los manuales e incluso en la categorización de las nociones específicas del PCIC, pasa por una revisión de los centros de interés. Con anterioridad, autores como Bartol Hernández (2010), Sánchez-Saus Laserna (2011) y Paredes García (2014) han propuesto renovar los dieciséis centros tradicionales por otros más adaptados a la situación actual. Es bien sabido que existen centros poco productivos, como Trabajos del campo y del Jardín e Iluminación, calefacción y medios de airear un recinto, y que existen lagunas en los diccionarios del léxico disponible que dejan de lado aspectos básicos como la dimensión perceptiva y anímica del individuo.

Dada la importancia de la publicación del Marco Común Europeo de Referencia para las Lenguas por el Consejo de Europa en 2001i, se ha comparado la propuesta de nociones específicas presente en el Plan Curricular del Instituto Cervantes con los centros de interés. Se ha constatado que los centros de interés solo cubren algunas parcelas de las nociones imprescindibles en el aprendizaje de una lengua. Además, los centros de interés que autores como Paredes García (2014) proponen sustituir se encuentran dispersos en la clasificación del PCIC.

De esta forma, se abren nuevas vías de investigación para crear el léxico disponible a partir de la selección de las 20 nociones específicas del PCIC ya que son las que guían la elaboración de materiales y posterior evaluación para los Diplomas de Español como Lengua Extranjera del Instituto Cervantes. No obstante, no se puede olvidar que el contexto de aprendizaje de la lengua influye en la selección del léxico eliminando palabras que carezcan de interés e incluyendo otras que sean determinantes. Por ello, el docente ha de saber que no sería recomendable que una de las primeras palabras relacionadas con la bebida que aprendiesen los escolares menores fuese cerveza y, más aún en países donde las bebidas alcohólicas están más restringidas.

Cabe destacar que si el objetivo es la creación de un léxico con fines específicos sería conveniente tomar como estímulo otros campos. Por ejemplo, para enseñar español administrativo o económico sería conveniente obtener el léxico a partir del subcampo del PCIC, *Financias y bolsa, renta, comercio, entidades y empresas*. En este caso es imprescindible seleccionar una muestra conocedora del tema y con una competencia léxica suficiente para obtener un léxico nuclear acorde a las necesidades de los estudiantes, como es el caso de los estudios llevados a cabo en el ámbito de las matemáticas por Urzúa, Sáez y Echeverría (2006) y Ferreira Salcedo, del Valle (2014).

En síntesis, este artículo ha presentado y mostrado la idoneidad de la selección del vocabulario disponible a partir de la aplicación de la teoría de los conjuntos difusos. Las unidades léxicas resultantes están organizadas en seis niveles de compatibilidad con respecto al núcleo o tema que se corresponden a un nivel de usuario básico según el MCERL. Además de la propuesta pedagógica en la selección léxica, se ha realizado una revisión de los centros de interés tradicionales a partir del contraste con las nociones específicas del PCIC para saber cuáles son las posibles lagunas léxicas que dejarían ese tipo de pruebas.

5. REFERENCIAS BIBLIOGRÁFICAS

- Alba Ovalle, O. (1998). “Variable léxica y dialectología hispánica”, en *La Torre: Revista de la Universidad de Puerto Rico*, 3, 7-8: 299-316.
- Alba Ovalle, O. (2000). “Variable léxica y comparación dialectal”, en O. Alba Ovalle, *Nuevos aspectos del español en Santo Domingo*. Santo Domingo: Librería La Trinitaria y Brigham Young University, 99-132.
- Ávila Muñoz, A. M. y Sánchez Sáez, J. M.^a. (2010). “La disponibilidad léxica. Antecedentes y fundamentos”, en A. M. Ávila Muñoz y J. A. Villena Ponsoda (Eds.), *Variación social del léxico disponible en la ciudad de Málaga*. Málaga: Editorial Sarriá, 35-81.
- Ávila Muñoz, A. M. y Sánchez Sáez, J. M.^a. (2011). “La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga”, en *ELUA. Estudios de Lingüística Universidad de Alicante*, 25: 45-74.
- Ávila Muñoz, A. M. y Sánchez Sáez, J. M.^a. (2014). Fuzzy sets and Prototype Theory: Representational model of cognitive community structures based on lexical availability trials, en *Review of Cognitive Linguistics*, 12, 1: 133-159.
- Ávila Muñoz, A. M. y Villena Ponsoda, J. A. (Eds.) (2010). *Variación social del léxico disponible en la ciudad de Málaga*. Málaga: Editorial Sarriá.
- Bartol Hernández, J. A. (2006). “La disponibilidad léxica”, en *Revista española de lingüística*, 36: 379-384.
- Bartol Hernández, J. A. (2010). “Disponibilidad léxica y selección del vocabulario”, en R. M.^a Castañer Martín y V. Lagüéns Gracia (Eds.), *De moneda nunca usada: Estudios dedicados a José M^a Enguita Utrilla*. Zaragoza: Institución Fernando el Católico, 85-107.
- Bartol Hernández, J. A. (2001). “Reflexiones sobre la disponibilidad léxica”, en José Antonio Bartol Hernández, Salvador Crespo Matellán, Carmen Fernández Juncal, Carmen Pensado, Emilio Prieto de los Pozos y Nieves Sánchez González de Herrero et alii (eds.), *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas*. Salamanca: Luso-Española Ediciones, 221-236.
- Bellón Fernández, J. J. (2011). *Léxico disponible de la provincia de Córdoba*. Córdoba: Servicio de Publicaciones de la Universidad de Córdoba.
- Benítez Pérez, P. (1992). “Disponibilidad léxica en la zona metropolitana de Madrid”, en *Boletín de la Academia Puertorriqueña de la Lengua Española*, 1,1: 71-102.
- 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.
- Carcedo González, A. (2001). *Léxico disponible de Asturias*. Turku: Universidad de Turku.

- Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Traducido por el Instituto Cervantes. Madrid: Ministerio de Educación Cultura y Deporte y Anaya.
- Dimitrijević, N. (1969). *Lexical Availability. A new aspect of the lexical availability of secondary school children*. Heidelberg: Julius Gross Verlag.
- Gómez Devís, M.^a B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Valencia: Universidad de Valencia.
- Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. (1956). *L'élaboration du français élémentaire*. Paris: Didier.
- Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. (1964). *L'élaboration du français fondamental (I degré). Étude sur l'élaboration d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Instituto Cervantes. (2006). *Plan Curricular del Instituto Cervantes (PCIC)*. Madrid: Biblioteca Nueva.
- López Chávez, J. y Strassburger Frías, C. (1987). "Otro cálculo del índice de disponibilidad léxica" en *Presente y perspectivas de la investigación computacional en México. Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada*. México: Universidad Nacional Autónoma de México.
- López Morales, H. (1973). *Disponibilidad léxica en escolares de San Juan*. MS.
- López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco/Libros.
- Mackey, W. (1971). *Le vocabulaire disponible du français (2 vols.)*. Paris, Bruxelles, Montreal: Didier.
- Mena Osorio, M. (1986). *Disponibilidad léxica infantil en tres niveles de enseñanza básica* (Tesis de maestría inédita). Universidad de Concepción.
- Paredes García, F. (2014). "A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE" [en línea], en *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 16. Disponible en <http://www.nebrija.com/revista-linguistica/a-vueltas-con-la-seleccion-de-centros-de-interes-en-los-estudios-de-disponibilidad-lexica-para-una-propuesta-renovadora-a-proposito-de-la-disponibilidad-lexica-en-ele>, recuperado el 15 de septiembre de 2015.
- Rodriguez, L. (2007). Ruralité et acquisition lexicale au Manitoba: le vocabulaire disponible dans les écoles Saint-Eustache (milieu rural) et Provencher (milieu urbain). *Cahiers franco-canadiens de l'Ouest*, 19,2: 141-158.
- Samper Padilla, J. A. (1999). "Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria", en E. Forastieri Braschi, J. Cardona, H. López Morales y A. Morales de Walters (Coords.), *Estudios de Lingüística Hispánica (Homenaje a María Vaquero)*. San Juan de Puerto Rico: Editorial de la Universidad de Puerto Rico, 551-573.
- Sánchez-Saus Laserna, M. (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*. (Tesis doctoral). Universidad de Cádiz. Disponible en <http://rodin.uca.es/xmlui/handle/10498/15862>. Recuperado el 15 de septiembre de 2015.
- Victory, J. B. Jr. (1971). *A study of Lexical Availability among Monolingual-Bilingual Speakers of Spanish and English*. (Master Thesis). Houston (Texas): Rice University. Disponible en <http://hdl.handle.net/1911/27482>, recuperado el 17 de septiembre de 2016.
- Zadeh L.A. (1965). "Fuzzy set" en *Information and control*, 8: 338-353.