Cálculo de disponibilidad léxica en ELE de universitarios marroquíes de Agadir: el programa *LexiDisp*

Mohamed Serfati
Universidad de Las Palmas de Gran Canaria

1. LA APLICACIÓN DEL PROGRAMA *LEXIDISP*

Para informatizar el léxico disponible recogido en las encuestas de nuestros informantes de Agadir y su zona de influencia, nos hemos servido del programa *LexiDisp* (versión 1.02), elaborado por José Enrique Moreno Fernández y Antonio García de las Heras (auspiciado por la Asociación de Lingüística y Filología de la América Latina, el Instituto Cervantes y la Universidad de Alcalá de Henares), sobre la fórmula desarrollada por López Chávez y Strassburger (1991: 92-93).

Como sabemos, *LexiDisp* es una aplicación para Windows por lo que debe utilizarse desde este entorno. Con esta herramienta se pretende cubrir la necesidad de los investigadores de disponibilidad léxica en el mundo hispánico. Tiene la posibilidad de relacionar la frecuencia de una unidad léxica con su lugar de aparición en las listas. Ya que cuando una palabra está más disponible para un hablante, este la incluye en su lista respuesta porque acude rápidamente a su memoria (Samper, Bellón y Samper, 2003: 103). Nuestra intención es, pues, dar una información a propósito de la utilización de este *software* presentando sus características generales y servir de ayuda para aquellos que quieren utilizarlo.

2. EL PROCESAMIENTO DE LOS DATOS

Hemos de subrayar que el léxico disponible de nuestros sujetos se calcula sin aislar ninguno de los factores sociales que se han tenido en cuenta para este estudio por medio del programa *LexiDisp*, por lo que se tiene en cuenta todo el conjunto léxico producido por los estudiantes en las pruebas de disponibilidad.

Entonces, para poder volcar[1] con éxito los resultados de las encuestas en el banco de datos de la disponibilidad léxica (*LexiDisp*), es condición *sine qua non* que el documento de origen presente las siguientes características:

[1] Quisiera agradecer aquí la ayuda siempre disponible de mis profesores Marta Samper Hernández, José Antonio Samper Padilla, así como la ayuda infatigable de los profesores Clara Eugenia Hernández Cabrera, Florentino Paredes García y la recién doctora Bianca Manuela Sandu en el manejo de esa base de datos a lo largo de la investigación.

a) Las variables sociológicas tienen que estar codificadas.

Cuadro 1. Muestra de variables codificadas

Variables	**Codificación**			
Sexo	Hombre	Mujer		
Lengua materna	Árabe 1	Tamazight 2		
Nivel sociocultural	Alto 1	Medio 2	Bajo 3	
Lengua extranjera	Francés 1	Inglés 2		
Procedencia	Agadir 1	Periferia 2	Warzazat 3	Otros 4

Además de las variables codificadas, en cada línea del documento que corresponde a cada centro de interés debe aparecer el número del informante y el del centro de interés al que corresponden las respuestas, que naturalmente será el mismo para todos los informantes recogidos en un mismo fichero.

Cuadro 2. Muestra de fichero de texto: código de características del informante + número del centro de interés + palabras recogidas mediante cuestionario

12331 001 01[2] garganta, mano, dedo, ojo, oreja, ceja, mejilla, rodilla, pierna
21431 002 01 garganta, cabeza, mano, cuello, oreja, pecho, boca, mejilla
11132 003 01 mano, pelo, garganta, ceja, oreja, cara, nariz, labio, diente, pie, uña, dedo

b) Es necesario que no haya líneas vacías ni al comienzo, al medio ni al final del documento.

c) Tras la última palabra de una línea no debe haber ningún signo de puntuación.

d) Los signos de puntación utilizados para la separación de las palabras no deben estar duplicados, tampoco deben aparecer dobles espacios entre una palabra y otra.

[2] Los cinco primeros números representan las variables sociológicas que hemos subrayado en el cuadro anterior: sexo, lengua materna, nivel sociocultural, lengua extranjera y procedencia. Las tres cifras siguientes corresponden al número de informante, que abarca desde el 001 hasta el 40, y los dos últimos números representan el centro de interés (1-17).

3. EL CÁLCULO DEL ÍNDICE DE DISPONIBILIDAD LÉXICA

La última etapa de esta fase metodológica es la obtención del índice de disponibilidad léxica (IDL) del vocabulario recopilado. Antes, se establece el orden de las unidades léxicas y su importancia a partir del número de veces que estas aparezcan en las muestras, lo cual lleva a establecer el índice de frecuencia.

Sin embargo, en los estudios de disponibilidad léxica no solo asumen como criterio el número de apariciones de la unidad léxica –frecuencia–, sino que, además, tienen el orden de aparición de estas dentro de las listas[3]; la correlación de estos dos elementos –frecuencia + orden– permite establecer el índice de disponibilidad léxica.

En realidad, los primeros trabajos en que se empezó a relacionar la frecuencia de una palabra con la posición que ocupa en cada campo nocional son de Roberto Lorán Santos y Humberto López Morales (1983). Sin embargo, y a pesar de la novedad de esta aportación, su fórmula tenía muchas limitaciones tal como lo indica Marta Samper Hernández (2009: 109):

> En realidad, la posibilidad de calcular mediante una fórmula matemática el grado de disponibilidad de cada palabra dentro de su correspondiente lista es una aportación de los estudios hispánicos, ya que los investigadores anteriores (Gougenheim y sus colaboradores, Mickey, Dimitrijévic…) […] solo trabajaron con el factor frecuencia.

Dichas limitaciones fueron solucionadas más tarde por los investigadores López y Strassburger (1987) mediante el desarrollo de un nuevo artificio matemático. Su fórmula ha demostrado su superioridad y objetividad científica para la lingüística y es, por supuesto, la elegida para realizar todas las indagaciones de disponibilidad léxica del proyecto panhispánico. Según la investigadora Marta Samper Hernández (2002: 43):

> La (fórmula) que ha sido utilizada en todas las investigaciones del proyecto panhispánico es la de Juan López Chávez y Carlos Strassburger Frías, una vez demostrada su superioridad para la lingüística gracias a los diversos trabajos realizados en México. […] logra una adecuación descriptiva plausible, tanto en los cálculos relativos al grupo como a los individuos que lo componen. Además, frente al modelo que habían propuesto Roberto Lorán Santos y Humberto López Morales, permite la discriminación matemática entre aquellos vocablos que aparecen en las posiciones inmediatas y finales de las listas de disponibilidad de los grupos.

[3] El orden de aparición de la palabra en los listados, tanto individuales como grupales, recibe el nombre de índice de espontaneidad; este subraya el hecho de que unas palabras son evocadas por el hablante de forma más rápida que otras, y, en consecuencia, ocupan las primeras posiciones en los listados, lo que lleva a plantear que poseen un alto índice de disponibilidad.

Para llegar a resultados objetivos, dicha fórmula toma en consideración estos factores:

a) La frecuencia absoluta de la palabra;
b) La frecuencia absoluta de la palabra en cada posición;
c) El número de los informantes;
d) El número de posiciones alcanzadas en la encuesta del centro de interés que se analice;
e) Las posiciones en que figura la palabra.

Esta fórmula es la que ha utilizado el proyecto panhispánico de disponibilidad léxica, y la que también sirvió de base para el análisis de disponibilidad léxica de los estudiantes marroquíes.

En definitiva, para editar nuestros datos hemos utilizado el programa informático *LexiDisp*[4], (versión 1.02) que calcula los siguientes valores:

a) El índice de disponibilidad[5] de cada unidad léxica, es decir el número de veces que se ha dicho una palabra y el orden en que aparece;
b) La aparición, es decir el porcentaje de alumnos que han actualizado esa unidad léxica en las encuestas;
c) La frecuencia con respecto al total de las palabras, o sea el número de veces que se ha dicho un término en relación al total de palabras;
d) La frecuencia acumulada[6] –suma de las relativas–.

Si aplicamos este planteamiento a los resultados obtenidos al confeccionar las listas de disponibilidad léxica en la muestra de la ciudad de Agadir y su zona de influencia, especialmente al conjunto de estudiantes del departamento de hispánicas, podemos observar de manera muy práctica cómo se manifiesta este planteamiento teórico a la hora de calcular el índice de disponibilidad.

La lista de palabras que aparece a continuación constituye el centro de interés 17 «los colores» del diccionario del léxico disponible de Agadir. En ella se encuentran recogidas todas las palabras aportadas por los informantes ordenadas según el índice de disponibilidad desde la palabra más disponible a la menos disponible.

---

[4] *LexiDisp* es una aplicación para Windows y se puede descargar en ordenadores personales en línea en http://www.linguas.net/Proyectos/LexiDisp/tabid/73/%20language/%20es-ES/Default.aspx. Al respecto, hemos utilizado la versión recomendada por los responsables del programa a los directores de esta tesis José Antonio Samper Padilla y Marta Samper Hernández.

[5] El índice de disponibilidad es un valor entre 0 y 1 que se obtiene en función de dos criterios: a) el número de informantes que incorporaron la palabra correspondiente y b) la posición en la que fue producida. Cuanto más se acerque a 1 el índice, más disponible estará la lexía en la sintopía estudiada.

[6] La frecuencia acumulada, que se obtiene sumando la frecuencia de la palabra a las frecuencias parciales de los lexemas precedentes, muestra el porcentaje sobre el total de palabras emitidas en el campo asociativo.

Es de subrayar que la frecuencia % se refiere al número de veces que aparece esa palabra respecto al total de palabras; así, si *rojo* tiene una frecuencia del

11,859%, significa que el 11,859% de las 312 palabras dadas del centro de interés 17 «los colores», ha sido la forma *rojo*, lo que equivale a 37 menciones. La columna aparición % nos da el porcentaje de sujetos que han aportado esa forma en su encuesta; de esta manera, *rojo* ha aparecido en 37 informantes, que representan el 92,500%. La columna de frecuencia acumulada refleja la suma de frecuencias relativas de las palabras que la preceden.

El cuadro siguiente muestra los vocablos encontrados en el centro de interés 17 «los colores» con índice de disponibilidad (ID) ≥ 0,1, y empleados, al menos por el 75% de la muestra.

Cuadro 3. Vocablos con índice de disponibilidad (ID) ≥ 0,1, y empleados, al menos por el 75% de la muestra en la ciudad de Agadir

N.º	Palabra	Disponibilidad	Frecuencia (%)	Aparición (%)	Frec. acumulada
1	rojo	0,72477	11,859	92,500	11,859
2	blanco	0,71397	12,821	100,000	24,679
3	verde	0,65399	11,218	87,500	35,897
4	azul	0,64051	11,538	90,000	47,436
5	negro	0,58532	10,897	85,000	58,333
6	amarillo	0,52098	10,577	82,500	68,910
7	marrón	0,22790	5,769	45,000	74,679
8	rosa	0,15531	4,167	32,500	78,846
9	gris	0,13996	4,167	32,500	83,013

Fuente: Serfati (2010). C. I. 17. «Los colores».

Como puede observarse, los siete vocablos empleados por más del 75% de la población estudiada son los que alcanzan mayor disponibilidad léxica. Esto nos llevó a confirmar que las unidades diferentes con mayor índice de disponibilidad de cada campo asociativo formaran la categorización conceptual colectiva del estímulo en cuestión.

A medida que el índice de disponibilidad de los vocablos decrece y, por tanto, es empleado por menor número de informantes, las unidades léxicas se ven alejadas del núcleo prototípico del centro de interés.

De esta forma, el cuadro 4 nos revela los vocablos que en el mismo centro de interés consiguen el menor índice de disponibilidad y están alejados del núcleo de categorización colectiva del área temática.

Cuadro 4. Vocablos con índice de disponibilidad (ID) ≤ 0,1 en la ciudad de Agadir

N.º	Palabra	Disponibilidad	Frecuencia	Aparición (%)	Frec. acumulada
10	naranja	0,08872	2,885	22,500	85,897
11	café	0,05948	1,923	15,000	89,744
12	castaño	0,06003	1,923	5,000	87,821
13	violeta	0,05652	1,603	12,500	91,346
14	oscuro	0,02218	0,641	5,000	91,987
15	moreno	0,02015	0,962	7,500	92,949
16	celeste	0,01847	0,641	5,000	93,590
17	lila	0,01793	0,641	5,000	94,231
18	carmesí	0,01624	0,321	2,500	94,551
19	multicolor	0,01218	0,321	2,500	94,872
20	primario	0,01218	0,321	2,500	95,192
21	beige	0,01188	0,641	5,000	95,833
22	plateado	0,01055	0,321	2,500	96,154
23	rubio	0,00914	0,321	2,500	96,474
24	morado	0,00792	0,321	2,500	96,795
25	anaranjado	0,00686	0,321	2,500	97,115
26	blanco roto	0,00686	0,321	2,500	97,436
27	turquesa	0,00686	0,321	2,500	97,756
28	azul roto	0,00594	0,321	2,500	98,077
29	secundario	0,00594	0,321	2,500	98,397
30	barniz	0,00514	0,321	2,500	98,718
31	claro	0,00445	0,321	2,500	99,038
32	rosado	0,00445	0,321	2,500	99,359
33	débil	0,00334	0,321	2,500	99,679
34	cálido	0,00289	0,321	2,500	100,000

Fuente: Serfati (2010). C. I. 17. «Los colores».

4. BIBLIOGRAFÍA

López Chávez, J. y Strassburger, C. (1987). «Otro cálculo del índice de disponibilidad léxica». En *Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada, Presente y perspectiva de la investigación computacional en México*. México: UNAM.

— (1991). «Un modelo más para el cálculo de disponibilidad léxica individual». En *La enseñanza del español como lengua extranjera*. Ed. Humberto López Morales. Río Piedras: Universidad de Puerto Rico, pp. 91-112.

López Chávez, J. (1992). «Alcances panhispánicos del léxico disponible». En *Lingüística* 4: 26-124.

Lorán, R. (1983). *Un índice de disponibilidad léxica*. Mayagüez: Universidad de Puerto Rico (Departamento de Matemáticas).

Lorán, R. y López Morales, H. (1983). *Nouveau calcul de l'indice de disponibilité*. Puerto Rico: MS.

Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.

— (2009). *Evolución de la disponibilidad léxica en estudiantes grancanarios de enseñanza primaria y secundaria*. Las Palmas de Gran Canaria: Cabildo de Gran Canaria.

Samper Padilla, J. A., Bellón Fernández, J. J. y Samper Hernández, M. (2003). «El proyecto de estudio de la disponibilidad léxica en español». En *Pautas y pistas en el análisis del léxico hispano (americano)*. Ed. Worjak, G., Frankfurt – Madrid: Vervuert – Iberoamericana, pp. 27-140.

Serfati, M. y Lahoussine, A. (2010). *Disponibilidad léxica de ELE en Marruecos (Niveles de Secundaria y Enseñanza Superior en la región Souss Massa Drâa)*. Agadir: Facultad de Letras y Ciencias Humanas – Universidad Ibnou Zohr.