

LA ESTRUCTURA ESTADÍSTICA DEL LÉXICO DISPONIBLE DE INFORMANTES NATIVOS E INFORMANTES DE ELE

María Victoria Galloso Camacho
Josefina Prado Aragonés
Universidad de Huelva

Los primeros estudios de disponibilidad léxica, que han cumplido ya casi 60 años, se iniciaron en Francia, ante la necesidad de planificar de una forma rápida y eficaz la enseñanza del francés como lengua extranjera, fundamentalmente a la población de las excolonias francesas de origen africano y asiático. Estos estudios surgieron como consecuencia de los sorprendentes resultados obtenidos de los trabajos léxico-estadísticos que los lingüistas Michéa, Gougenheim y Sauvageot (1964) estaban realizando con el fin de elaborar el corpus del francés elemental contemporáneo, basados en principio en la frecuencia de aparición de las unidades léxicas, pensando que dichas palabras frecuentes eran las más útiles y las más usadas, y, por tanto, las más necesarias para la comunicación. Pero estos investigadores advirtieron pronto que palabras que parecían muy frecuentes en esa lengua no se encontraban en las listas de frecuencias. Así, mientras unas palabras muy usuales aparecían constantemente en el *corpus* obtenido, otras, en cambio, tan usuales como aquéllas, no lo hacían (era evidente que palabras como *metro*, *autobús*, *timbre* o *carta* no podían faltar en un léxico «fundamental» del francés). De ello se dedujo que no había una clara relación entre frecuencia y uso y que la frecuencia de los sustantivos no depende del número de palabras analizadas, sino de los temas tratados en las situaciones de comunicación. De manera que determinadas palabras usuales sólo se actualizan en la mente del hablante mediante un estímulo temático, es decir, si el tema de la comunicación lo requiere. Esto nos lleva a distinguir entre *palabras frecuentes* y *palabras disponibles*. Las primeras se actualizan en cualquier situación comunicativa, con independencia del tema tratado (de ahí que también se denominen *atemáticas*); la mayoría de ellas son palabras gramaticales y las que no lo son aparecen en este orden decreciente de frecuencia: verbos, adjetivos y sustantivos de significado general como *cosa*, *hombr*e, *persona*; frente a las *palabras disponibles*, de uso frecuente y común en una comunidad, pero que sólo se actualizan con estímulos temáticos en situaciones comunicativas específicas en que se necesita transmitir información sobre un tema determinado (son las llamadas *temáticas* porque dependen del tema) y pertenecen principalmente a la clase de los sustantivos que aluden a realidades concretas.

Surge así el concepto de *disponibilidad léxica*, que estaría integrada por aquellas palabras que acuden antes a la mente del hablante como reacción a un estímulo temático concreto (López Morales 1986, 1999) y constituyen el *vocabulario disponible* de ese hablante o de esa comunidad de hablantes.

En este punto de la investigación fue cuando se vio la necesidad de actualizar el léxico de un modo artificial y obtener en superficie las palabras de que dispone un determinado hablante o grupo de hablantes, mediante la propuesta de pruebas asociativas basadas en centros de interés, que funcionasen como estímulos para que el hablante expresase, escribiéndolas en una lista en un tiempo determinado, aquellas palabras que conoce y utiliza en torno a ese tema concreto. De esta manera, el rango de disponibilidad léxica va en función de las palabras que acuden antes a la mente del hablante en relación con un tema determinado. Serán las más disponibles las que ocupen los primeros lugares.

Gougenheim *et al.* (1964) utilizaron como estímulos en las primeras investigaciones de disponibilidad léxica los dieciséis centros de interés siguientes: *partes del cuerpo*, *vestidos*, *la casa*, *muebles de la casa*, *alimentos y bebidas*, *objetos que hay sobre la mesa en las comidas*, *la cocina con sus muebles y utensilios*, *la escuela y material escolar*, *el calor y el alumbrado*, *la ciudad*, *el pueblo o el mercado*, *medios de transporte*, *trabajos del campo y del jardín*, *animales*, *juegos y distracciones*, y *oficios*.

En investigaciones posteriores sobre disponibilidad léxica, partiendo de los centros de interés anteriores, se han llevado a cabo algunas modificaciones, adaptándolos al objeto de investigación y al contexto geográfico donde se lleva a cabo la misma.

Los estudios sobre la disponibilidad léxica de una comunidad de hablantes, basados en fórmulas matemáticas que permiten obtener el grado de disponibilidad de las palabras, ha abierto nuevos ca-

minos a diversas disciplinas lingüísticas como la psicolingüística, la sociolingüística y, muy especialmente, la lingüística aplicada y la didáctica de la lengua, para las que tiene un valor fundamental por sus implicaciones en la enseñanza de la lengua materna y lenguas extranjeras, a la hora de planificar de forma rigurosa la enseñanza del vocabulario (Prado, 1999), como han puesto de manifiesto, entre otros, López Morales (1999), el cual coordina en la actualidad un ambicioso proyecto Panhispánico de disponibilidad léxica del nivel preuniversitario en fase de realización bastante avanzada. Ya son muchas las provincias y comunidades españolas e hispanoamericanas que tienen publicados los vocabularios disponibles de dicho nivel preuniversitario¹, e incluso se está extendiendo dicha investigación a otros niveles de enseñanza² que, sin duda, aportan notables posibilidades para su aprovechamiento en las disciplinas mencionadas.

En esta comunicación trataremos de presentar algunas de las posibilidades que estas investigaciones pueden brindar en la enseñanza de ELE, aprovechando los estudios que sobre disponibilidad léxica y ELE ya han publicado Carcedo (2000), Samper Hernández (2001) y Sánchez Gómez (2004), en los que se analizan y comparan los resultados de disponibilidad léxica obtenidos en estudios realizados a estudiantes de español como lengua extranjera, que han sido también comparados, concretamente por Carcedo, con los obtenidos en otros llevados a cabo a informantes españoles.

La aplicación de estos estudios a la enseñanza-aprendizaje del español como lengua extranjera es aún incipiente. Las líneas de investigación que sigue dicha aplicación tratan de reflejar las peculiaridades detectadas en los léxicos contrastados, como por ejemplo, poner de manifiesto el aprovechamiento que de orden etnolingüístico se puede hacer de aquellos inventarios al tiempo que se aplican a la didáctica de segundas lenguas³.

Aprovechando los resultados obtenidos en el estudio realizado por Sánchez Gómez (2004) a un grupo de estudiantes de ELE, hemos comparado dichos resultados con los obtenidos en el estudio que recientemente se ha llevado a cabo a estudiantes preuniversitarios de Huelva y provincia, con el fin de comprobar si el léxico disponible de estos alumnos extranjeros en la lengua meta se aproxima al de alumnos nativos, a medida que asciende su dominio lingüístico del español, y de esta forma obtener nuevos datos que contribuyan en el avance de esta línea de investigación centrada en la planificación léxica en la enseñanza de ELE.

La investigación realizada por Sánchez se ha llevado a cabo sobre una muestra empírica integrada por un total de 44 informantes distribuidos en cuatro niveles de estudio⁴: Intermedio I, Intermedio II, Avanzado, Superior.

Siguiendo las directrices de Samper Hernández (2001), en este trabajo tampoco se ha contemplado la edad de los informantes, pero sí se ha tenido en cuenta el nivel de conocimientos de español que tenían. Las variantes valoradas han sido: sexo, nivel de dominio de la lengua extranjera, nacionalidad de los padres, nacionalidad de los informantes y otros países visitados. Y, como en los trabajos citados, se ha utilizado la metodología del proyecto Panhispánico: dos minutos para que cada informante respondiera por escrito todas aquellas palabras que le vinieran a la mente sobre los doce centros de interés seleccionados: 01. *El cuerpo humano*; 02. *La ropa*; 03. *Partes de la casa (sin muebles)*; 04. *Alimentos y bebidas*; 05. *La educación*; 06. *La ciudad*; 07. *El campo*; 08. *Medios de transportes*; 09. *Animales*; 10. *Juegos y distracciones*; 11. *Profesiones y oficios*; 12. *La salud*.

Tras resultados obtenidos, deseamos plantear las siguientes consideraciones:

1. En primer lugar, hay que destacar la dificultad que ha representado la localización de un grupo más o menos homogéneo de estudiantes de español como lengua extranjera; y nos referimos con el adjetivo homogéneo tanto a distintos niveles de estudio de español, como a edad o número de informantes de cada grupo. De ahí que hubiese que desplazarse a Sevilla, aun con los inconvenientes previamente advertidos de que no todos los grupos contaban con el mismo número de informantes y que a menudo el nivel de los alumnos puede ser engañoso ya que a veces la prueba de nivel que se realiza al principio del curso no se corresponde exactamente con su nivel de conocimientos.
2. En segundo lugar, nos gustaría poner de manifiesto la importancia que las distintas investigaciones de disponibilidad léxica y, por ende, los diccionarios de léxico disponible pueden tener en la enseñanza

¹ El proyecto pretende alcanzar todos aquellos territorios donde se hable español, de hecho ya contamos con el diccionario de léxico disponible de Puerto Rico, República Dominicana, y se está trabajando en ello en México, Costa Rica y en otros países latinoamericanos.

² En la Universidad de Huelva estamos realizando dicha investigación también en los niveles de enseñanza obligatoria, con el fin de valorar el incremento cuantitativo, además de la calidad cualitativa, de los escolares en su proceso de formación.

³ Sería muy interesante la ampliación de estos estudios, pues su aprovechamiento en la exploración de rasgos culturales falta por completo.

⁴ Las encuestas se realizaron en Sevilla, en la Academia I.S.A., que colabora con la Universidad Menéndez Pelayo, dado que en la Universidad de Huelva, donde trabajamos, no existía ningún grupo que cumpliera las características que buscábamos para realizar nuestro trabajo.

de segundas lenguas, en este caso del español. Como señala Calero Fernández (2003: 179):

Uno de los instrumentos más útiles en el aprendizaje de una lengua, tanto materna como extranjera, es el diccionario. Sin embargo, de todos los materiales didácticos empleados para tal fin, ha sido el menos cuidado hasta hace muy poco, no sólo porque la técnica lexicográfica – y la propia metalexigrafía – no estaban suficientemente desarrolladas, sino también porque los diccionarios escolares no eran más que una reproducción en pequeña escala de los diccionarios generales, sin atender las necesidades propias de cualquier aprendiz de una lengua.

Sin embargo, las investigaciones de léxico disponible van en la dirección de ofrecer un corpus de léxico actual y verdaderamente usado por los hablantes de una comunidad y, por lo tanto, necesario en los actos de comunicación que se le presentan al hablante en su vida cotidiana, puesto que la información ha sido obtenida a partir del empleo que los encuestados realizan de su propia lengua, con el valor añadido de que se trata de un trabajo que proporciona información dialectológica, sociolingüística y de frecuencia, además de su carácter panhispánico. Se trata, pues, de un auténtico diccionario de uso, no solo fiable por la metodología empleada, sino por su actualidad que debería ser imprescindible para los lexicógrafos y para el material lexicográfico en el que se trabaja hoy día, que hasta el momento parece que no se ha valorado como se merece y que atendería a las imperfecciones o insuficiencias que destaca Calero (2003) en su trabajo sobre los diccionarios de enseñanza materna y extranjera actuales. Se superaría de esta forma la sistematización de los datos que aportan los diccionarios y podríamos saber con rigurosa exactitud el carácter dialectal o sociolectal de los mismos.

3. En tercer lugar, queremos presentar la valoración obtenida de la disponibilidad léxica de los estudiantes preuniversitarios onubenses y el grupo de estudiantes de español como segunda lengua, que se desprende del estudio comparativo realizado entre ambos grupos, y su análisis cuantitativo y cualitativo.

3.1 Aspectos cuantitativos:

Dado que los datos obtenidos en los diccionarios de léxico disponible nos permiten realizar comparaciones entre sí a fin de significar las diferencias y semejanzas entre ellos, en esta comunicación nos disponemos a cotejar los léxicos disponibles de cinco de los centros de interés obtenidos, por un lado, en la provincia de Huelva y, por otro, en las encuestas realizadas a los estudiantes de ELE. El objetivo es el análisis del porcentaje de afinidad de cada uno de estos cinco centros de interés, tomando como referencia aquellos vocablos o palabras diferentes, de ambos grupos, que no superan el 75% de frecuencia acumulada.⁵ Los centros de interés seleccionados han sido los siguientes: 01. *El cuerpo humano*; 02. *La ropa*; 03. *Alimentos y bebidas*; 04. *La ciudad*; 05. *Profesiones y oficios*.

En el siguiente cuadro se aprecia el cómputo de vocablos con una frecuencia acumulada inferior al 75% en informantes de ELE:

CENTROS DE INTERÉS	Nivel Inter. I	Nivel Avanz.	Nivel Sup.
01. Partes del cuerpo	18	23	27
02. La ropa	14	21	27
03. Alimentos y bebidas	49	48	60
04. La ciudad	64	54	61
05. Profesiones y oficios	31	39	50
Totales	176	185	225

Cuadro 1. *Vocablos por centros de interés*

CENTROS DE INTERÉS	HUELVA	ZAMORA	CÁDIZ
06. Partes del cuerpo	38	36	40
07. La ropa	36	33	34
08. Alimentos y bebidas	96	84	82
09. La ciudad	100	95	93
10. Profesiones y oficios	118	111	98
Totales	388	359	347

Cuadro 2. *Vocablos por centros de interés*⁶

⁵ Seguimos de esta forma el modelo de cotejo que anteriormente han utilizado otros investigadores como González Martínez y Orellana Ramírez (2000) o Carcedo (2003). Como ellos y como indica Samper (1999: 554): «creemos que hay razones suficientes para justificar la decisión de poner un límite y no utilizar todos los datos recogidos en las encuestas: a partir de un determinado punto de los listados de disponibilidad aparece un conjunto de palabras que han sido aportadas por muy pocos encuestados, que, además, las han incluido en sus hojas de respuesta en posiciones poco relevantes. Podríamos encontrarlos entonces con que no estamos comparando hechos de norma, sino fenómenos particulares del habla de un grupo de informantes o incluso de un único hablante».

3.2. Aspectos cualitativos:

Los resultados que muestran los cuadros anteriores son prácticamente los esperables. Las cifras de vocablos del cuadro 2 son semejantes, la coincidencia es absoluta tanto en los centros que aportan más vocablos como en los que aportan menos. Los resultados de los estudiantes extranjeros (cuadro 1) no alcanzan los de los preuniversitarios, algo bastante lógico pues no podemos esperar que actualicen el mismo número de unidades que los nativos. Sin embargo, «la diferencia entre unos y otros disminuye cuando el objeto de análisis es el más elevado de los niveles de estudios que la investigación contempla. Las diferencias disminuyen a medida que asciende el nivel de estudios» (Carcedo, 2000: 154). Aunque el número de vocablos es razonablemente menor en los estudiantes de ELE, coinciden con los estudiantes españoles en la mayor y menor productividad de los centros, salvo en el caso de *Profesiones y oficios*, en el que sorprendentemente los extranjeros no producen un gran número de vocablos, a diferencia de lo que sucede con los finlandeses, que sí coinciden con los nativos (Carcedo, 2000: 155). Ambos grupos, los preuniversitarios y los estudiantes extranjeros producen lógicamente menos en los centros más concretos (como en *Partes del Cuerpo* o *La Ropa*) y producen más en los centros de áreas referenciales más amplios (*Alimentos y Bebidas* y *La Ciudad*). El único dato cuantitativo que no se corresponde con lo esperado es el de *La Ciudad* del Nivel Intermedio I, puesto que contradictoriamente supera a los otros dos niveles. Ello se comprende rápidamente cuando se analizan las respuestas, entre las cuales aparecen 15 nombres de ciudades. Por lo tanto, si restamos ese número al que figura en el Cuadro 1 obtenemos el razonable resultado de 49, más acorde con el conjunto.

3.2.1. A continuación expondremos los 20 primeros vocablos que aparecen en las listas de las dos zonas cotejadas, los datos obtenidos de los estudiantes preuniversitarios de Huelva y los informantes de ELE⁷, con el fin de que puedan advertirse, de este modo, las diferencias y/o semejanzas entre cada uno de ellos.

01. PARTES DEL CUERPO				02. LA ROPA			
	Huelva		ELE		Huelva		ELE
1	cabeza	1	dedo	1	pantalón	1	pantalón
2	brazo	2	rodilla	2	camisa	2	camiseta
3	ojo	3	pie	3	calcetín	3	zapato
4	mano	4	brazo	4	chaleco	4	chaqueta
5	pierna	5	nariz	5	camiseta	5	sombrero
6	nariz	6	mano	6	falda	6	camisa
7	oreja	7	pierna	7	jersey	7	falda
8	pie	8	ojo	8	chaqueta	8	abrigo
9	dedo	9	oreja	9	braga	9	calcetín
10	boca	10	cabeza	10	zapato	10	pantalón corto
11	rodilla	11	espalda	11	bufanda	11	suéter
12	cuello	12	pelo	12	blusa	12	media
13	codo	13	tobillo	13	calzoncillo	13	cinturón
14	uña	14	cara	14	sujetador	14	bufanda
15	muñeca	15	muñeca	15	guante	15	blusa
16	hombro	16	boca	16	chaquetón	16	vestido
17	espalda	17	uña	17	abrigo	17	ropa interior
18	tobillo	18	piel	18	chándal	18	jean
19	pelo	19	pecho	19	media	19	manga larga
20	pecho	20	diente	20	sudadera	20	guante

⁶ Los datos de Zamora y Cádiz han sido tomados de la investigación de González y Orellana (2000: 142).

⁷ Los resultados de los estudiantes extranjeros que presentamos corresponden a los del nivel superior, teniendo en cuenta, como señalábamos antes, puesto que la diferencia entre éstos y los nativos no debería ser tan significativa al tratarse del objeto de análisis el más elevado de los niveles de estudios que la investigación contempla. Las diferencias disminuyen a medida que asciende el nivel de estudios.

03. ALIMENTOS Y BEBIDAS
Huelva ELE

04. LA CIUDAD
Huelva ELE

1	agua	1	café	1	coche	1	tienda
2	Coca Cola	2	vino	2	tienda	2	calle
3	güisqui	3	pan	3	calle	3	coche
4	lenteja	4	leche	4	edificio	4	edificio
5	Fanta	5	zumo	5	casa	5	casa
6	patata	6	carne	6	parque	6	parque
7	zumo	7	naranja	7	semáforo	7	autobús
8	tomate	8	manzana	8	carretera	8	taxi
9	pan	9	agua	9	contaminación	9	gente
10	garbanzo	10	mantequilla	10	autobús	10	restaurante
11	lechuga	11	queso	11	piso	11	banco
12	arroz	12	pollo	12	plaza	12	carro
13	jamón	13	cerdo	13	gente	13	bar
14	carne	14	uva	14	bar	14	río
15	vino	15	cerveza	15	avenida	15	hotel
16	pescado	16	cubata	16	ruido	16	discoteca
17	huevo	17	refresco	17	moto	17	oficina
18	cerveza	18	sangría	18	discoteca	18	camino
19	ron	19	pescado	19	tráfico	19	muchas gente
20	leche	20	Fanta	20	ayuntamiento	20	piso

05. ALIMENTOS Y BEBIDAS
Huelva ELE

1	profesor/a	1	profesor/a
2	médico/a	2	negocio
3	albañil	3	abogado
4	abogado/a	4	jefe
5	arquitecto	5	camarero/a
6	agricultor	6	doctor
7	policia	7	taxista
8	maestro	8	enfermero/a
9	camarero	9	médico
10	ama de casa	10	estudiante
11	secretario/a	11	maestro
12	enfermero/a	12	dentista
13	carpintero	13	policia
14	empresario	14	vendedor
15	ingeniero	15	conductor
16	jardinero	16	pescadero
17	futbolista	17	piloto
18	periodista	18	ingeniero
19	farmacéutico	19	actor
20	camionero	20	presidente

En el siguiente cuadro podemos observar en cifras las semejanzas y/o diferencias cuantitativas entre ambos léxicos de cada uno de los cinco centros de interés:

01. Partes de la casa	17
02. La ropa	12
03. Alimentos y bebidas	8
04. La ciudad	10
05. Profesiones y oficios	8

Cuadro 3. Número de vocablos comunes en las posiciones del 1-20. Convergencia Huelva/ELE

El cuadro refleja la notable coincidencia entre las dos sintonías, coincidencia esperable al haberse establecido la comparación entre las unidades más altas de cada uno de los listados. Carcedo (2003: 215) concluye igualmente, en el análisis de los léxicos disponibles de Asturias, Cádiz y Costa Rica, que «la comparación de las unidades más disponibles de cada léxico arroja un grado de convergencia bastante elevado, pues una gran parte de los vocablos se demuestran comunes para las sintonías cotejadas». Asimismo, coincide con nuestro trabajo en que observamos la mayor coincidencia en los centros más concretos, como ya comentáramos anteriormente, especialmente en *Partes del cuerpo*, pues «dichos campos están constituidos por un número menor de clases lógicas, lo que reduce el tipo de evocaciones que puede despertar en el hablante. En otras palabras, el mayor o menor grado de coincidencia de la comparación interdialectal parece guardar relación muy estrecha con el carácter cerrado o difuso del área nocional que se sondea» (Carcedo, 2003: 209). En algunos casos coinciden los mismos vocablos en las mismas posiciones del léxico onubense y del español de los estudiantes extranjeros: *muñeca* (15); *pantalón* (1); *edificio* (4); *casa* (5); *parque* (6); *profesor/a* (1). Y en el centro *La ciudad* coinciden exactamente los seis primeros vocablos en los dos listados, pero en distinto orden los tres primeros: *coche* (1) (3), *tienda* (1) (2), *calle* (2) (3), *edificio* (4), *casa* (5), *parque* (6).

BIBLIOGRAFÍA

- Calero Fernández, M. A. (2003): «Información sociolingüística en los diccionarios de enseñanza de la lengua: el *DSLE*», en Francisco Moreno Fernández, Francisco Gimeno Menéndez, José Antonio Samper, M. Luz Gutiérrez Araus, María Vaquero y César Hernández (eds.), *Lengua, Variación y Contexto, Estudios dedicados a Humberto López Morales* 1, 179-197, Madrid, Arco/Libros.
- Carcedo González, A. (2000): *Disponibilidad léxica en español como lengua extranjera: el caso finlandés*, Turku, Universidad de Turku.
- Carcedo González, A. (2003): «Unidad y variedad diatópico de la disponibilidad léxica del español: comparación de los inventarios de Puerto Rico, Cádiz y Asturias», en Francisco Moreno Fernández, Francisco Gimeno Menéndez, José Antonio Samper, M. Luz Gutiérrez Araus, María Vaquero y César Hernández (eds.), *Lengua, Variación y Contexto, Estudios dedicados a Humberto López Morales* 1, 199-225, Madrid, Arco/Libros.
- González Martínez A. y Orellana Ramírez, P. (2000): «Cotejo de los léxicos disponibles de Cádiz y Zamora», en *Documentos de Español Actual (DEA)*, 139-160, Turku, Dpto. de Lengua Española-Universidad de Turku.
- Gougenheim, G., Michéa R., Rivenc, P. y Sauvageot, A. (1964): *L'elaboration du Français Fondamental*, Paris, Didier.
- López Morales, H. (1986): *Enseñanza de la lengua materna*, Río Piedras, Plaza Mayor.
- López Morales, H. (1999): *Léxico disponible de Puerto Rico*, Madrid, Arco/Libros.
- Prado Aragonés, J. (1999): «El léxico y sus implicaciones en la didáctica de la lengua», *Seminarios de Lingüística*, 3, Universidad de Faro (Portugal), págs. 145-162.
- Samper Hernández, M. (2001): *Disponibilidad léxica en alumnos de español como lengua extranjera*, Málaga, ASELE.
- Sánchez Gómez, C. (2003): *Disponibilidad léxica en alumnos de español como lengua extranjera*, Tesina inédita, Universidad de Huelva.