EL VOCABULARIO FUNDAMENTAL: HISTORIA, DEFINICIÓN Y NUEVAS PROPUESTAS APLICADAS A LA ENSEÑANZA DE ELE[1]

XANDRA SANTOS PALMOU
Instituto Politécnico do Porto (ISCAP-CICE)

RESUMEN

En esta comunicación se define el concepto de Vocabulario Fundamental (VF) y se aporta la metodología para delimitarlo, en este caso con el ejemplo de la lengua española y con el objetivo de la enseñanza de ELE.

Para conocer el estado de la cuestión, parto del estudio de las anteriores propuestas de léxicos fundamentales, así como de los léxicos frecuentes y los diccionarios de léxico disponible, ambos de forma independiente. Se profundiza en el estudio del léxico fundamental: definición del concepto, historia desde la primera propuesta hasta las más actuales, metodologías seguidas, resultados, aplicaciones, retos, etc.

La hipótesis es que el vocabulario fundamental refleja de forma más completa la competencia léxica de una lengua que el léxico frecuente y el disponible por separado, y de que la unión de ambos puede resultar altamente útil como propuesta de selección léxica para la docencia de ELE.

Palabras clave: selección del vocabulario, disponibilidad léxica, léxico frecuente, léxico fundamental, ELE.

ABSTRACT

This paper deals with the concept of Fundamental Vocabulary (FV). We expose the methodology that we have developed to delimit it, in this case with the example of the Spanish language and with the aim of teaching this language.

We start from the study of the previous proposals of FV as well as of the dictionaries of frequencies and the dictionaries of available vocabulary, both of independent form. We deepen in the study of the fundamental lexicon: definition of the concept, history from the first proposal until the most current, followed methodologies, results, applications, challenges, etc.

We hypothesized that the fundamental vocabulary reflects more fully the lexical competence of a language than frequent and lexical availability separately, and

that the union of both can be highly useful to lexical selection for teaching Spanish as Foreign Language (or SFL).

Keywords: vocabulary selection, lexical availability, frequent lexicon, fundamental lexicon, Spanish as Foreign Language (or SFL).

1. INTRODUCCIÓN

En el ámbito hispano conocemos la expresión y el concepto de Vocabulario Fundamental (VF) gracias a la línea de investigación de la Disponibilidad Léxica (DL). Aunque en este contexto básicamente se retoman y reproducen las palabras de Humberto López Morales, director y coordinador del *Proyecto Panhispánico de Léxico Disponible* (PPHDL), que define el Léxico Fundamental como la unión del Léxico Disponible y el Léxico Básico.

A pesar de que no se haya escrito demasiado sobre el concepto y la metodología del Vocabulario Fundamental[2] (VF), su nacimiento fue parejo al de la Disponibilidad Léxica (DL), si bien fue esta última la que nació dentro del primero. Cuando los autores franceses Gougenheim, Michéa, Rivenc, y Sauvageot hicieron su propuesta de "français élémentaire" en 1954 y establecieron el repertorio léxico que enseñar a los aprendices extranjeros de francés, en realidad estaban poniendo la primera piedra de lo que poco más tarde definieron como "français fondamental". Esta aportación pretendía suplir las deficiencias léxicas que constataron al usar exclusivamente los listados de frecuencia. Ofrecían usar la DL, no como método alternativo e independiente, sino recurriendo a la unión de ambas fórmulas para la selección del vocabulario: frecuencia y disponibilidad. Así, si por una parte se estaba desarrollando toda una línea de investigación de gran profusión en español (gracias al PPHDL), su apuesta era sumar a las ventajas de los recuentos de frecuencias léxicas las virtudes que se descubrieron en las encuestas de disponibilidad, pues ambos procedimientos reducían sus limitaciones si se complementaban. De tal modo que el tipo de unidades léxicas que se conseguía seleccionar mediante esta nueva vía eran justo las que se echaban en falta en los estudios de frecuencias, y viceversa. Por lo tanto, estos investigadores franceses vieron muy pronto el potencial de aquella unión que empezaron a denominar Léxico[3] Fundamental, y que más tarde se llevaría a cabo en otras lenguas como el irlandés, alemán, español y portugués.

La bibliografía sobre DL en español a día de hoy es numerosa, pero son pocas las ocasiones en las que se centra en desarrollar la posibilidad de aplicar a la enseñanza del español el léxico disponible, tanto por separado como en conjunto con el léxico frecuente, y no pasan de algunas líneas en las que se comenta esta posibilidad.

2. VOCABULARIO FRECUENTE

El aprovechamiento de los corpus de frecuencias léxicas puede ser altamente útil aplicado a la enseñanza de lenguas extranjeras, como así lo demuestra su recurrente uso por parte de la tradición anglosajona en diccionarios, lecturas adaptadas, manuales, etc. aplicados a la enseñanza del inglés. Y es que realmente con un número reducido de palabras podemos obtener una cobertura textual muy amplia si las escogemos adecuadamente. Con "cobertura textual", nos referimos al porcentaje de palabras que componen un texto (oral o escrito). Esta es la principal ventaja de las palabras frecuentes, que aparecen de forma recurrente independientemente del tipo de discurso y contenido, con lo que permiten crear numerosos textos con pocos elementos.

De los corpus de frecuencias podemos extraer los léxicos básicos, seleccionando aquellas palabras más usuales y aplicando una criba que pondere su aparición en los diferentes tipos de textos que conforman el corpus (orales, escritos, por registros, por tipología textual, etc.). El uso de métodos lexicométricos para obtener las palabras más usuales reduce la arbitrariedad en la selección del vocabulario y permite además establecer cortes según su índice de uso, lo cual posibilita una gradación del léxico por niveles educativos.

Sabemos también que mediante los recuentos de frecuencias vamos a priorizar la obtención de un tipo de unidades léxicas, mientras que otras van a estar fuera de esas 5 o 7 mil más frecuentes, que son las que se suelen considerar para un recuento de léxico básico. Así, las palabras frecuentes que se obtienen mayoritariamente suelen ser las consideradas vacías, las gramaticales, los nombres y adjetivos de significado muy general (como *cosa, casa, tiempo, lugar, persona, sitio*...). Quizás esta sea una de las grandes debilidades que presenta recurrir solo a la frecuencia como método de selección léxica y que se pretendía corregir con el uso de la metodología de la DL. Los ejemplos de palabras que se consideran cotidianas y muy conocidas por cualquier hablante pero no aparecen entre

las más frecuentes son múltiples. Un caso altamente significativo es el que Gougenheim, Michéa y Sauvageot (1964: 138-139) usaron para ilustrar este tipo de vocablos que son altamente conocidos por los hablantes pero pueden ser poco usadas fuera de un contexto concreto: la palabra "fourchette" (*tenedor*). También lo percibió Paredes García (2015), quien rescata el ejemplo y apunta que en el CREA ocupa la posición 17.942, lo que evidencia que estaría lejos de introducirse en la enseñanza de esta lengua si este fuera en único principio. Por consiguiente, el porcentaje de uso de las unidades léxicas en los corpus no siempre es sinónimo de utilidad didáctica, ya que hay otras palabras que no alcanzamos con este criterio que tendremos que perseguir con otros, entre ellos la DL.

Al mismo tiempo, existen determinadas cuestiones metodológicas que la lingüística del corpus debe actualizar y perfeccionar para conseguir responder a algunas necesidades a las que todavía no ofrece soluciones. En concreto, sería útil etiquetar todo tipo de unidades léxicas, no solo las palabras gráficas, para poder obtener también las pluriverbales y rescatar las colocaciones, de gran importancia a la hora de aprender una lengua extranjera, ya que es una cuestión que presenta dificultades. Si los corpus nos permitieran saber cuáles son las colocaciones más frecuentes, podríamos incorporarlas con mayor eficacia desde un inicio y graduar también su aprendizaje.

Otra cuestión sobre la que todavía se debe reflexionar son los parámetros y fórmulas con las que decidimos qué número y cuáles palabras usuales se seleccionan para cada nivel de lengua. En mi caso, he tomado como referencia los diferentes estudios sobre cobertura textual y la existencia de un "umbral léxico" hechos sobre el inglés, lengua en la que abunda el uso de la lexicometría aplicada a la enseñanza de la lengua. En este idioma existen corpus que organizan el vocabulario por listas, con familias de palabras, en función de su nivel de frecuencia y uso, distinguiéndolas del vocabulario académico o técnico, así como análisis de comprensión de diferentes tipos de textos (orales y escritos) por parte de aprendices en función de cuál es el grado de conocimiento del vocabulario. Concluyeron que existe un nivel mínimo de conocimiento léxico que los aprendices de una L2 deben alcanzar para lograr la comprensión efectiva: es el denominado umbral léxico. Estos niveles serían el 95% de cobertura textual para que el nivel de comprensión sea aceptable y el 98% para que podamos considerarlo óptimo (Cartaya Febres 2011). Tomando como referencia las listas del *British Nacional Corpus* (BCN), un aprendiz que

conoce las primeras 3.000 familias de palabras más frecuentes y a la vez está familiarizado con un amplio inventario de nombres propios puede alcanzar el 95% de cobertura textual. Esto es lo que se considera el umbral léxico 'mínimo' para alcanzar un nivel 'aceptable' de comprensión de lectura en la L2. Sin embargo, para poder alcanzar el nivel 'óptimo' de comprensión, que se estima alrededor del 98%, se sugiere el dominio de las primeras cinco listas o bandas de frecuencia del BCN, esto es, unas 5.000 familias de palabras de alta frecuencia. Se debe tener en cuenta que estos son los niveles de comprensión que se necesitan para que se produzcan las inferencias léxicas contextuales, que nos permiten deducir el significado de las palabras desconocidas por el contexto. En español no hay estudios de este tipo, aunque un inicio podrían ser los llevados a cabo por Pellicer-Sánchez (2015) sobre el acceso léxico en el ámbito de la psicolingüística, los enfoques de enseñanza del léxico (incidental o explícito) y el nivel de lectura en español como L2.

3. VOCABULARIO DISPONIBLE

Frente a los puntos débiles que se observan al recurrir a la frecuencia de las palabras para escoger las unidades léxicas más útiles para la enseñanza de una lengua extranjera, la DL se erige como un proceso, más que alternativo, complementario. Entre sus virtudes se encuentra la objetividad que ofrece ser un método lexicométrico, que reduce la arbitrariedad en la selección y recurre a fórmulas científicas de trabajo. Los índices matemáticos mediante los que se organizan sus resultados (índice de disponibilidad, frecuencia absoluta, frecuencia acumulada y porcentaje de aparición) permiten ordenar los datos y establecer cortes en las unidades léxicas, basándose en un tratamiento matemático de los resultados de las encuestas.

Por otra parte, si bien usada en exclusividad sus resultados también se manifiestan parciales, cuando lo sumamos a las listas de frecuencias léxicas se muestran altamente útiles, pues obtenemos precisamente el tipo de términos que se echaban en falta: sustantivos, adjetivos, algunos verbos y en general palabras temáticas de alto valor significativo que están en relación directa con las situaciones comunicativas. A pesar de que el léxico disponible obtenido no es un vocabulario actualizado sino potencial, permite conocer las palabras reales y actuales de la población de un lugar concreto, por lo que se tiene en cuenta la variación léxica de la lengua en

sus diferentes vertientes. Las encuestas se organizan por ámbitos temáticos, denominados Centros de Interés, lo que aporta también sus beneficios. En primer lugar, obtener las palabras asociadas a campos semánticos ayuda a conocer la tipicidad de los resultados obtenidos en relación con el CI propuesto. El orden de las unidades ofrece la posibilidad de estudiar la organización del lexicón mental y las asociaciones que se establecen en él entre sus elementos. En segundo lugar, estos Centros de Interés (CI) que se suele usar guardan relación con las nociones específicas del *Plan Curricular del Instituto Cervantes* (PCIC) y del *Marco Común Europeo de Referencia para las Lenguas* (MCER), aunque recientemente se está trabajando para poder mejorar este aspecto (Tomé Cornejo 2015). Otra posibilidad es crear Centros de Interés más concretos, que sirvan para obtener vocabulario de un ámbito más específico, pensando por ejemplo en la enseñanza de lenguas de especialidad (Español para Fines Específicos, EFE). Ampliar y mejorar los CI es uno de los retos principales que hoy afronta esta línea de investigación, y parece que para ello la referencia a seguir son las 20 nociones específicas del PCIC.

Una vez que hemos obtenido y tratado las unidades léxicas, debemos fijar los parámetros de selección de las unidades que conforman los diccionarios de léxico disponible para cada nivel educativo, para lo cual no existen todavía propuestas suficientemente asentadas y probadas. Esto se debe a que, a pesar de la profusión y madurez de los estudios sobre DL en los últimos 20 años en el seno del PPHDL, no se han aplicado apenas a la enseñanza-aprendizaje de ELE, cuestión que se ha retomado más recientemente. De ahí que algunas soluciones importantes, como los CI precisos para dar respuesta a las nociones específicas del PCIC o el modo en que seleccionamos las unidades de los diccionarios de léxico disponible para ELE, se hayan afrontado en fechas muy próximas. Nos referimos a las propuestas de Tomé Cornejo (2015), para la primera cuestión, y a la presentada por Ávila Muñoz (2016), para la segunda. Este último recurre a modelos matemáticos de lógica difusa para estudiar el grado de compatibilidad de las palabras dentro de cada CI, con lo que sigue principios objetivos para jerarquizar el léxico. Aun así, se deben validar estos modelos en otras muestras.

4. VOCABULARIO FUNDAMENTAL (VF)

En definitiva, el objetivo final es hallar el Vocabulario Fundamental. Es decir, para poder afrontar el reto de unir de forma eficaz el vocabulario frecuente y el disponible, debemos solucionar en la medida de lo posible las cuestiones metodológicas que he venido comentando. Así, si ambos métodos por separado suponen soluciones parciales, establecer un VF del español supone, pues, una alternativa más completa y que se puede orientar directamente a la enseñanza-aprendizaje de ELE.

La primera iniciativa de este tipo, en cuyo seno se creó la línea de la DL, fue la llevada a cabo en francés por Gougenheim, Michéa, Rivenc, Paul y Sauvageot en 1954 en *L'élaboration du français élémentaire: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, con una nueva edición revisada en 1964 donde ya se usa la expresión de "francés fundamental" en lugar de "elemental" (*L'élaboration du français fondamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Didier, Paris). Su propuesta pretendía perfeccionar las deficiencias que los listados de frecuencia ofrecían, no usar la DL como método paralelo.

Atendiendo al contexto, ambas publicaciones nacieron de un proyecto con miras más amplias y bien orquestado, cuyo objetivo era mantener el prestigio del francés frente al ascenso del inglés como lengua internacional tras la II Guerra Mundial. Francia emprende en los 50 una reforma concienzuda de la enseñanza y la difusión de su lengua, que afecta tanto a la metodología como a los materiales. De hecho estas obras están auspiciadas por el Ministerio de Educación de Francia y dirigidas por el *Centre de Recherche et d'Étude pour la Diffusion du Français* (C.R.E.D.I.F.). De este centro dependía la Comisión para el Francés Fundamental, que coordinaba las encuestas y tomaba las decisiones para el proceso de selección del vocabulario. Un equipo de expertos -entre los que figuran el lingüista G. Gougenheim y el pedagogo P. Rivenc- adscritos al C.R.E.D.I.F. elabora la lista de palabras frecuentes y disponibles, así como una gradación gramatical. La referencia para ello era el *Basic English* (compuesto por solo 850 palabras) confeccionado por C.K. Ogden, que junto con otras obras en lengua inglesa y francesa de léxico frecuente son citadas por estos autores. El equipo persigue elaborar un vocabulario y gramáticas básicos y se centra en el francés oral, ya que el método educativo que iba a llevar a la práctica sus contenidos era el método

audiovisual (promocionado por Robert Lado en *Language teaching. A scientific approach* de 1964). Son las concreciones didácticas de la lingüística estructuralista y las teorías conductistas. Su éxito fue rotundo, también a nivel comercial.

En la bibliografía sobre DL se han repetido algunos datos sobre este origen, pero casi nada se ha profundizado en su aprovechamiento a la enseñanza de lenguas extranjeras, casi nada se comenta sobre el resto de obras que surgieron en otras lenguas ni su relación con los diferentes métodos didácticos de cada época. Sin embargo, al estudiar el origen e historia del VF se puede apreciar que existen propuestas en irlandés, alemán, español, portugués, etc. En español, los trabajos sobre léxico fundamental se usaron para los manuales *Vida y diálogos de España. 1er grado. Método audiovisual* (Rojo Sastre, J. A., Rivenc P. y Ferrer, A. 1968-1969) publicados en Paris por Didier y en Nueva York por Chilton Books. Cuentan con 4 volúmenes: V.1 *Libro del profesor*; V. 2. *Textos de lectura. Ortografía del español, preparación para el dictado*; V. 3. *Lista alfabética de formas empleadas*; V. 4. *Libro de imágenes*. Estos materiales inauguran la enseñanza de ELE y en su día fueron muy difundidos. Su autor, Rojo Sastre, presentaría en 1979 su tesis doctoral, *Bases lexicológicas en una metodología audiovisual de la lengua española para extranjeros*, en la que da cuenta por extenso de las bases teóricas, métodos de selección léxica y resultados pedagógicos que le habían servido para los anteriores tomos. Según su estudio, el "español fundamental hablado", como él lo denominó, estaba formado por 1.560 palabras, teniendo en cuanta las frecuentes y las disponibles.

El proyecto portugués, igualmente de gran envergadura, ve la luz en tres tomos en la década de los 80, cuando el método didáctico que estaba popularizándose ya era el comunicativo.

5. VOCABULARIO FUNDAMENTAL Y ELE

Para la enseñanza-aprendizaje de ELE tener un VF jerarquizado por niveles supondría un referente valioso a la hora de elaborar los programas de los cursos y sus materiales. Además, se parte de la misma hipótesis que los investigadores franceses al defender que el VF refleja de forma más completa la competencia léxica de los hablantes nativos y su lexicón mental que ambos géneros de vocabulario por separado. A partir de ahí, los siguiente pasos son definir qué y cuántas unidades léxicas se incluyen

en él, qué métodos de selección se siguen (léxico frecuente, disponible o algún tipo más), qué parámetros se usan para escoger cada tipo de vocabulario, como las unimos, cómo las organizamos y presentamos para su manejo y aprovechamiento, etc.

Por mi parte, considero que el VF es aquel necesario para que un hablante pueda desenvolverse con eficacia en las situaciones comunicativas cotidianas y tratando temas habituales. Está compuesto por diferentes unidades, incluidas las pluriverbales. Esto nos sitúa en un nivel umbral, es decir, un B1 del MCER, aunque se puede intentar jerarquizar de tal forma que se obtenga una clasificación para los niveles iniciales con un menor número de unidades y otra para los intermedios, dejando para los superiores el resto. Considerando los recursos con los que hoy contamos en español, he valorado los vocabularios básicos y diccionarios de léxico disponible actuales. Así, he optado por *A Frequency Dictionary of Spanish: core vocabulary for learners* (Davies 2006) para extraer la base de palabras frecuentes y *El léxico de frecuencia del español hablado en la ciudad de Málaga* (Ávila Muñoz 1999) y para complementarlo con algunas colocaciones y locuciones más comunes que este sí contempla. Con respecto al léxico disponible, he escogido los de Galicia, Alicante, Salamanca y Ceuta. Determino cuántas unidades léxicas de las más frecuentes necesitamos aproximadamente para cada nivel recurriendo a los estudios de cobertura léxica que antes se mencionaron. La meta es un curso de nivel A2 e inicio de un B1 del MCER, para lo que he tomado las 2.000 palabras más frecuentes (según el índice de frecuencia establecido con la fórmula usada por Davies) a las que sumo las más disponibles de los 16 CI clásicos, fijando como parámetro de corte las que tengan un Índice de Disponibilidad superior o igual a 0’1. He definido estos materiales y parámetros para escoger las unidades léxicas con las que establezco un núcleo léxico de no más de 3.000 palabras como una propuesta piloto de Vocabulario Fundamental del español. En este momento la investigación se centra en comparar este vocabulario con el propuesto por el PCIC para estos niveles, así como la búsqueda de la organización más adecuada.

Por último, para la docencia de ELE debe considerarse también el vocabulario instrumental y recurrir a un análisis contrastivo cuando la lengua meta y la lengua del alumnado son próximas (Izquierdo Gil 2014). Además, a estos métodos objetivos de selección que sientan la base léxica deben sumarse otras aportaciones más subjetivas que tengan en cuenta las circunstancias de enseñanza concretas.

NOTAS

[1] Este trabajo se inscribe en el proyecto COMBIDIGILEX, ref. FFI2015-64476-P, financiado por el Ministerio de Economía y Competitividad (IP: Inmaculada Mas Álvarez y Meike Meliss, Universidade de Santiago de Compostela).

[2] Cabe indicar que no tiene nada que ver con el uso de vocabulario fundamental que hizo en su día García Hoz (1953) en su trabajo sobre frecuencias léxicas titulado *Vocabulario usual, común y fundamental.*

[3] Usamos "léxico" y "vocabulario" indistintamente, aunque somos conscientes de su diferencia: léxico es todo el caudal léxico de una lengua, frente a vocabulario que es el conjunto de unidades léxicas que un hablante actualiza en el discurso. El vocabulario es el léxico potencial, por lo tanto. La diferencia está en la línea de lo que Saussure (1976) estableció entre lengua/habla, o lo que es lo mismo, entre el sistema y el uso de la lengua.

REFERENCIAS BIBLIOGRÁFICAS

Ávila Muñoz, A. M. 1999. *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Málaga: Universidad de Málaga.

Ávila Muñoz, A. M. 2016. "El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de los conjuntos difusos", *Journal of Spanish Language Teaching,* Vol.3, Iss.1: 31-43.

Bacelar do Nascimento, M. F., P. Rivenc, M.L. Segura da Cruz. 1984. *Português Fundamental, Vocabulário e Gramática, tomo 1, Vocabulário*. Lisboa: INIC.

Bartol Hernández, J. A. 2010. "Disponibilidad léxica y selección del vocabulario". En R. Castañer Martín y V. Lagüéns Gracia (eds.), *De moneda nunca usada. Estudios filológicos dedicados a José M. Enguita Utrilla.* Zaragoza: Institución Fernando el Católico. 85-107.

Cartaya Febres, C. N. 2011. "La hipótesis del umbral léxico en la comprensión de lectura en L2". *Lingua Americana,* Año XV, Nº 28: 103 – 118.

Davies, M. 2006. *A Frequency Dictionary of Spanish: core vocabulary for learners*. New York: Routledge.

Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. 1954. *L'élaboration du français élémentaire: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier. (Con

una nueva edición revisada en 1964: *L'élaboration du français fondamental:étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.)

Gougenheim, G., Michéa, R., Rivenc, P. y Sauvageot, A. (1958). *Dictionnaire fondamental de la langue française*. Paris: Didier. (Tiene una nueva edición revisada en 1977).

Izquierdo Gil, Mª C. 2005. *La selección del léxico en la enseñanza del español como lengua extranjera*. Málaga: ASELE, Colección Monografías nº 8.

López Morales, H. 1996. "Los estudios de disponibilidad léxica: pasado y presente". En *Homenaje a Rodolfo Oroz*, BFUCh XXXV: 245-259.

Paredes García, F. 2015. "Disponibilidad Léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula". *Lingüística en la Red*.

Pellicer-Sánchez, A. 2015. "Developing automaticity and speed of lexical access: the effects of incidental and explicit teaching approaches", *Journal of Spanish Language Teaching*, 2:2, 126-139.

Rojo Sastre, A. J. 1979. *Bases lexicológicas en una metodología audiovisual de la lengua española para extranjeros*, Tesis Doctoral, Madrid: Universidad Complutense. [Dirigida por Víctor García Hoz. Texto manuscrito e inédito.]

Tomé Cornejo, C. 2015. *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE.* Tesis Doctoral defendida en el curso 2014-2015 en la Universidad de Salamanca.