

Competencia léxica y oportunidades de aprendizaje léxico en el aula de lenguas: una aproximación teórica

Lexical knowledge and vocabulary learning opportunities in the language classroom: a theoretical approach

Pablo Robles García

Universidad de California, Davis

problesgarcia@ucdavis.edu

Resumen

El componente léxico es un aspecto esencial en el desarrollo y dominio lingüístico de un aprendiente. Diversas investigaciones argumentan que un número relativamente pequeño de palabras cubre una gran proporción del léxico utilizado en una lengua. Por lo tanto, es necesario definir y delimitar cuáles son estas palabras ya que estas unidades léxicas se encuentran en una amplia variedad de textos escritos y producciones orales y son de reconocida utilidad comunicativa. Además, debemos conocer si los cursos de enseñanza de lenguas y los materiales didácticos ofrecen a los estudiantes las condiciones necesarias para favorecer el aprendizaje de este vocabulario. En este trabajo (1) se lleva a cabo una descripción detallada de los conceptos de cobertura léxica, amplitud léxica y el tratamiento del vocabulario en los manuales tanto en inglés como en español, (2) se hace énfasis en la escasez de estudios relacionados con la adquisición de vocabulario en español y (3) se propone una agenda con una serie de pautas para subsanar las carencias observadas en materia de investigación.

Palabras clave: cobertura léxica, amplitud léxica, aprendizaje de lenguas, frecuencia, corpus

Abstract

Lexical knowledge is an essential element in the development and linguistic competence of L2 learners. Several studies show that a relatively small number of word-families appear more frequently and represent a large proportion of the words used in a given language. Therefore, it is important to define and delimit what these words are, since these lexical units can be found in a wide variety of written texts and oral productions and are extremely useful for communicative purposes. Furthermore, we must know if language teaching courses and language textbooks favor the acquisition of these lexical units. This study (1) brings a detailed description of the concepts of lexical coverage, vocabulary size and vocabulary treatment in language textbooks in both Spanish and English, (2) emphasizes a lack of studies related to the acquisition of vocabulary in Spanish and (3) proposes an agenda with a series of guidelines in order to solve this lack of empirical studies.

Keywords: lexical coverage, vocabulary size, language learning, frequency, corpora

1. INTRODUCCIÓN

El conocimiento léxico es un elemento fundamental en el desarrollo lingüístico de un aprendiente en una lengua extranjera. Su adquisición no solo influye en la mejora de las habilidades de comprensión oral y escrita (Hu y Nation, 2000) sino que también favorece en el desarrollo de la expresión oral (Hilton, 2008) y del rendimiento académico de un aprendiente (Stæhr, 2008). Esto nos sugiere que el aprendizaje del vocabulario debe ser un pilar fundamental en la enseñanza de lenguas.

Puesto que el lexicón de una lengua está compuesto por una gran cantidad de palabras, desde un punto de vista pedagógico, es necesario plantearse (1) cuáles son los criterios para decidir cuáles son las palabras que se deben enseñar prioritariamente y (2) si los materiales de enseñanza que se utilizan en este momento responden a esos criterios. En este contexto, la bibliografía parece concentrarse en que la frecuencia léxica debería ser el criterio principal para seleccionar vocabulario. Esto se debe a que las palabras más frecuentes son las que ofrecen mayor cobertura en los textos orales y escritos de una lengua. Por esta razón, el presente trabajo pretende ofrecer una visión de conjunto de los resultados de investigación obtenidos en el ámbito de la cobertura léxica, de la frecuencia léxica y de la selección léxica en materiales de enseñanza, con el objetivo de proponer mejoras en el ámbito de la enseñanza de español lengua extranjera (ELE).

El interés que despierta el aprendizaje léxico se manifiesta en la abundancia de monografías y artículos específicos que se han publicado durante las dos últimas décadas (Laufer, 1989; Nation, 2006; Horst, 2013; Schmitt y Schmitt, 2014). Destacan especialmente los trabajos aportados por Paul Nation, Norbert Schmitt, Batia Laufer y Marlise Horst, referentes internacionales en el estudio de la adquisición del vocabulario. Ahora bien, la mayoría de las investigaciones que hacen referencia a la cobertura y amplitud léxica y al tratamiento del vocabulario se han desarrollado predominantemente con aprendientes y manuales de inglés como segunda lengua (L2). En efecto, existe un desajuste evidente entre las investigaciones llevadas a cabo en inglés y en español:

- **Cobertura léxica:** Laufer (1989), Hu y Nation (2000), Nation (2006), Laufer y Ravenhorst-Kalovski (2010), Schmitt, Jiang y Grabe (2011), Van Zeeland y Schmitt (2012), todos ellos sobre el inglés;
- **Amplitud léxica:** Hirsh y Nation (1992), Nation (2006), Webb y Rodgers (2009a), Webb y Rodgers (2009b), Van Zeeland y Schmitt (2012), frente al trabajo de Davies (2005) para el español;
- **Tratamiento del vocabulario de alta frecuencia:** Meara, Lightbown y Halter (1997), Matsuoka y Hirsh (2010), Horst (2010), Alsaif (2011), Martini (2012), versus la investigación de Davies y Face (2006) para el español.

Como podemos observar, son escasos los estudios relacionados con la adquisición de vocabulario en español. Además, los resultados se basan en datos recopilados hace ya más de una década, situación que refleja un panorama obsoleto en materia de investigación. No obstante, esta carencia no se relaciona con la cantidad de hablantes nativos y aprendientes de español como lengua extranjera. Según el anuario de “El español en el mundo” publicado por el Instituto Cervantes (2017), el español es la segunda lengua del mundo con

mayor número de hablantes nativos (más de 572 millones de personas) y es el tercer idioma más estudiado como lengua extranjera. Además, la importancia del español en muchos países, como es el caso de Estados Unidos, es aún más evidente. Según el último censo realizado, más del 15 % de la población estadounidense es hispana (lo que la convierte en el grupo minoritario más numeroso del país) y el español es el idioma más estudiado como L2 por encima del francés o el alemán. Consecuentemente, creemos que la investigación debería estar acorde con la relevancia que tiene el español en el mundo actual.

Este trabajo se plantea, pues, como un estado de la cuestión sobre la cobertura léxica, la amplitud léxica y el tratamiento del vocabulario tanto en inglés como en español. Además, pretende sensibilizar a la comunidad docente e investigadora sobre la necesidad de llevar a cabo un plan riguroso de investigación que nos ofrezca cifras concretas y actualizadas sobre la realidad léxica del español actual.

Así pues, el primer apartado de este trabajo ofrece una descripción detallada sobre la importancia que tiene el conocimiento léxico en la competencia lingüística de un aprendiente. El segundo apartado ofrece cifras concretas sobre el porcentaje (cobertura léxica) y el número de palabras (amplitud léxica) que necesita un aprendiente para comprender la mayoría de los textos orales y escritos de una lengua. El tercer apartado analiza cuáles son las palabras más productivas e importantes que un aprendiente de L2 debe conocer, basándose en criterios de frecuencia y cobertura. Para terminar, el último apartado describe el tratamiento de estas palabras en los manuales de enseñanza de L2 y en el habla del instructor. Con esta revisión bibliográfica, pretendemos presentar los datos disponibles en inglés y en español y mostrar cuáles son las carencias observadas en los estudios sobre ELE. En efecto, con este artículo pretendemos dar los primeros pasos hacia una agenda de investigación sobre la enseñanza/aprendizaje de vocabulario en español, tratando de aumentar la visibilidad de los estudios de ELE y ponerlos al nivel de los trabajos que se han hecho en inglés.

2. LA IMPORTANCIA DEL VOCABULARIO EN UNA L2

El conocimiento léxico es un elemento clave en el desarrollo de la competencia lingüística en una L2 (Saville-Troike, 1984; Stæhr, 2008; Horst, 2013). Se ha demostrado que el conocimiento de una amplia variedad de vocabulario favorece la comprensión oral (Van Zeeland y Schmitt, 2012) y escrita (Schmitt, Jiang y Grabe, 2011) y juega un papel esencial en la competencia comunicativa (Hilton, 2008) y el rendimiento académico (Saville-Troike, 1984). En este contexto, el estudio de Saville-Troike (1984) es uno de los primeros trabajos que trata de forma exhaustiva la importancia del vocabulario en el desarrollo lingüístico y el éxito académico de un aprendiente. Concretamente, la autora analizó la relación entre varios aspectos del conocimiento lingüístico (i.e., producción léxica, complejidad gramatical y precisión morfológica) y el rendimiento de 19 estudiantes de primaria en el *Comprehensive Test of Basic Skills* (CTBS), un examen estadounidense estandarizado que se realiza a nivel nacional y en el que se incluyen pruebas de lectura, matemáticas, ciencias y estudios sociales.

Los resultados mostraron que -a excepción de la prueba matemática- existía una correlación significativa entre la amplitud léxica y la actuación de los estudiantes, es decir, cuanto más vocabulario conocían, mejores puntuaciones obtenían en las diferentes pruebas académicas,

especialmente en las pruebas de lectura. Este resultado es incluso más interesante cuando se ve a la luz de otro resultado sorprendente: contrariamente al efecto positivo de la amplitud léxica, la complejidad gramatical y la precisión morfológica no afectaron significativamente al rendimiento académico. Por ejemplo, cinco de los seis participantes que lograron los mejores resultados omitían de forma sistemática el plural, los artículos y los marcadores temporales (presente, pasado, futuro). Por el contrario, tres de los seis estudiantes que obtuvieron peores resultados tenían un dominio morfológico y gramatical del inglés muy superior al del resto de los participantes. Además, la influencia de la amplitud léxica se hizo más evidente en la prueba de comprensión lectora que en las demás pruebas, ya que esta fue la tarea donde se encontró una diferencia más significativa. En las pruebas de ciencias y estudios sociales, el vocabulario también proporcionó una mejora en el rendimiento de los estudiantes, aunque no fue tan evidente como en la prueba de lectura. En definitiva, la correlación existente entre el conocimiento léxico y la comprensión lectora se confirma en múltiples estudios como los de Hu y Nation (2000) o Schmitt, Jiang y Grabe (2011) y se observa también en el ámbito de la comprensión auditiva (Van Zeeland y Schmitt, 2012).

Más allá de la relación entre la amplitud léxica y las habilidades de comprensión, algunos trabajos recientes han estudiado también la contribución del léxico al desarrollo de la expresión oral. Por ejemplo, Hilton (2008) se interesó por el impacto del conocimiento léxico en la fluidez oral. Para ello, analizó el corpus PAROLE, que se compone de las transcripciones de una serie de monólogos de 56 estudiantes de tres L2 distintas (12 en francés, 11 en italiano y 33 en inglés) y 23 nativos de inglés. Los resultados mostraron que los aprendientes que disponían de una amplitud léxica mayor hablaban con mayor fluidez (130 palabras por minuto) que los que tenían un lexicón más reducido (52 palabras por minuto). Además, el habla de los aprendientes con un menor nivel de competencia léxica se caracterizaba por el empleo de pausas¹ que interrumpían el ritmo natural del habla de forma frecuente. En este contexto, Hilton observó que la mayoría de las pausas realizadas por los hablantes nativos se producían al final de una oración; y en raras ocasiones se daban en el medio de una cláusula. Sin embargo, en el caso de los aprendientes de L2, más de la mitad de las pausas que realizaban los aprendientes con un menor nivel de conocimiento léxico se encontraban en el medio de las oraciones y no al final de las mismas. Tras haber observado la posición en la cual los aprendientes producían más pausas, la autora analizó de forma detallada estos fragmentos con el fin de conocer cuáles eran los motivos que inducían a la producción de cláusulas discontinuas. Los resultados revelaron que la inmensa mayoría de las pausas venían propiciadas por la falta de repertorio léxico o por un intento de recuperación léxica prolongado:

- He's uh wearing the same [long pause] sweat [* lexical error for sweater] than [. . .] when he was a child (Hilton, 2008, p. 159).

En algunos casos, estas pausas iban acompañadas por una pregunta al interlocutor, pidiendo ayuda para recuperar la palabra olvidada:

¹ Por pausa nos referimos a los periodos de silencio que superan los dos segundos de duración, ya que en el marco de una conversación, el interlocutor tiende a intervenir cuando se produce una pausa de esta magnitud (Rieger, 2003).

- Sulla video vediamo [long pause] che [long pause] che una [long pause] uno [long pause] ...une machine comment on dit? (p. 160).

Como se puede observar, las pausas debidas a una falta de conocimiento léxico interrumpen la comunicación hasta tal punto que el propósito mismo del intercambio puede verse comprometido. Hilton observó que este fenómeno de disrupción comunicativa es específico de las pausas debidas a limitaciones léxicas mientras que las pausas que tenían una naturaleza gramatical se integraban sin problemas en el flujo del habla. Concretamente, la búsqueda de formas morfosintácticas se caracterizaba, en su mayoría, por el uso de reformulaciones, pausas cortas o meras repeticiones que no interferían en la comunicación de forma significativa.

En resumen, estos resultados ponen en evidencia la necesidad de ofrecer mayor importancia a la enseñanza del vocabulario en el aula. En el marco de una clase cuyo objetivo es la comunicación, este argumento es particularmente poderoso, ya que la falta de fluidez interrumpe y dificulta la comunicación. Sin embargo, la enseñanza de lenguas extranjeras tradicional se ha centrado predominantemente en el conocimiento gramatical. Esto se hace patente en los cursos de formación docente, donde el objetivo primordial es que los instructores aprendan a transmitir los contenidos gramaticales de forma efectiva (Horst, 2013), sin considerar que es el conocimiento léxico el que más influye en la comprensión y la expresión oral y escrita (Hu y Nation, 2000; Schmitt, Jiang y Grabe, 2011). En definitiva, como dice Wilkins (1972) en su trabajo “while without grammar little can be conveyed, without vocabulary nothing can be conveyed” (p. 111).

3. ¿QUÉ PALABRAS SE DEBEN ENSEÑAR?

Una vez reconocida la necesidad de enfatizar el componente léxico en las clases de L2 debemos conocer cuántas palabras hay que enseñar y en qué orden. Puesto que el vocabulario de una lengua está compuesto por una red de decenas de miles de palabras (Schmitt *et al.*, 2017) es necesario adoptar un criterio que nos ayude a delimitar esta cantidad. Uno de los argumentos más objetivos que se pueden proponer para efectuar dicha selección es el de la cobertura léxica, es decir, ¿qué porcentaje de palabras de un texto hay que conocer en una lengua para comprenderlo fácilmente? Y, una vez respondida esta pregunta, ¿cuáles son las palabras de una lengua que permiten que se dé dicho porcentaje de cobertura en la mayoría de textos de una lengua? Una vez establecido el número de palabras mínimo que aseguran la comprensión de la mayoría de textos escritos en una lengua, es razonable argumentar que son esas las que se deberían priorizar en la enseñanza de esta lengua. Por esta razón, el apartado 3.1. se centrará en los trabajos que han tratado de establecer el porcentaje mínimo de cobertura léxica que asegura la comprensión fluida, mientras que el apartado 3.2. se enfocará en los trabajos que han estudiado cuántas palabras (y cuáles) son las que pueden asegurar dicho nivel de cobertura.

3.1. Cobertura léxica

La mayoría de las investigaciones realizadas hasta la fecha (Laufer, 1989; Hu y Nation, 2000; Nation, 2006; Laufer y Ravenhorst-Kalovski, 2010; Schmitt, Jiang y Grabe, 2011; Van Zeeland y Schmitt, 2012) argumentan que la comprensión se produce de forma

satisfactoria cuando el vocabulario conocido ronda entre el 95-98 % del texto. Por lo tanto, si las palabras desconocidas en un texto superan el 2-5 % del léxico total, la comprensión se deteriora de forma significativa.

Existe cierto debate en torno a estos dos límites (i.e., 95 o 98 %). Los primeros estudios que abordaron este tema (Liu y Nation, 1985; Laufer, 1989) argumentaron que es necesario un conocimiento del 95 % del vocabulario en un texto para alcanzar un grado de comprensión satisfactorio. Esto supone enfrentarse a una palabra desconocida por cada dos líneas aproximadamente. Sin embargo, investigaciones posteriores sugieren que es necesaria una cobertura del 98 % para garantizar una comprensión eficaz. Hu y Nation (2000), por ejemplo, analizaron qué porcentaje de cobertura léxica es necesario para leer un texto de forma efectiva e independiente. Con este fin, los autores crearon cuatro versiones del mismo texto con diferentes niveles de cobertura léxica: 100 %, 95 %, 90 % y 80 % respectivamente. Para conseguir estas proporciones, sustituyeron palabras de baja frecuencia por palabras inventadas. Una vez que los estudiantes leyeron el texto, se les proporcionaron dos actividades de comprensión: (1) un test de respuesta múltiple y (2) una prueba escrita de respuestas cortas. El test de respuesta múltiple estaba compuesto por 14 preguntas y la prueba escrita estaba compuesta por 27 preguntas. En total, la puntuación máxima en el test de respuesta múltiple era de 14 puntos y en la prueba de respuestas cortas de 124. Posteriormente, los autores delimitaron cuál sería la puntuación mínima para alcanzar un nivel de comprensión satisfactorio. Hu y Nation decidieron que los estudiantes debían obtener al menos 11 puntos en la tarea de respuesta múltiple y 70 puntos en la prueba escrita. Cualquier puntuación por debajo de estos límites sería considerada no satisfactoria. La tabla 1 muestra los resultados del estudio y confirma que la cobertura léxica afecta de forma sustancial a la comprensión lectora.

Tabla 1. Rango y media de respuestas correctas en el test de respuesta múltiple y en la prueba escrita para los distintos niveles de cobertura léxica (Hu y Nation, 2000)

	Test de respuesta múltiple					Prueba escrita de respuestas cortas			
Rango de respuesta	100 %	95 %	90 %	80 %	Rango de respuesta	100 %	95 %	90 %	80 %
13-14	5	5	3		100-109	1			
11-12	11	4	4		90-99	2			
9-10	1	4	5	2	80-89	5	2		
7-8		1	1	3	70-79	2	5	3	
Menos de 7		3	3	11	Menos de 70	7	10	13	16
Media de respuestas	12.24	10.18	9.5	6.06		77.17	61	51.3	24.6
Participantes	17	17	16	16		17	17	16	16

En el test de respuesta múltiple, 16 de los 17 participantes que leyeron el texto con un 100 % de cobertura obtuvieron una comprensión satisfactoria. Ahora bien, con una cobertura del 95 %, solo la mitad de los estudiantes fueron capaces de comprender el texto de forma eficaz. Esto nos sugiere que el 95 % de cobertura no garantiza que los estudiantes obtengan una comprensión adecuada. Asimismo, esta proporción disminuyó considerablemente con los participantes que leyeron el texto con la versión del 90 %; y con un 80 %, ningún estudiante pudo obtener una comprensión eficaz. En el caso de la prueba escrita, los resultados son aún más evidentes. La mayoría de los estudiantes que leyeron el texto con un 100 % de cobertura obtuvieron resultados satisfactorios. Además, la media de respuestas correctas de estos estudiantes se encuentra por encima (77.17) del umbral de comprensión adecuado propuesto por los autores (70). Por el contrario, solo una minoría de los estudiantes que participaron en la versión del 95 % pudo alcanzar una comprensión efectiva. En cuanto a los estudiantes del 90 %, solo uno logró una comprensión apropiada y ningún participante del 80 % obtuvo una comprensión satisfactoria. Después de analizar los resultados, Hu y Nation aplicaron un análisis de regresión lineal para observar la relación entre los niveles de comprensión y cobertura léxica y observaron que un 98 % de cobertura habría sido suficiente para que la mayoría de los estudiantes hubieran obtenido una comprensión satisfactoria. En efecto, este nivel de cobertura habría producido una media de 11.53 puntos en el test de respuesta múltiple y 70.82 en la prueba escrita.

Así, de acuerdo con estos autores, un 95 % de cobertura léxica no debería considerarse umbral de comprensión satisfactorio ya que solo a partir del 98 % de cobertura se puede empezar a considerar que la mayoría de los lectores son capaces de alcanzar una comprensión adecuada. Esta conclusión también corresponde con los resultados encontrados por Schmitt, Jiang y Grabe (2011) en su estudio de comprensión lectora con aprendientes de inglés como L2. En la tentativa de comprobar la relación existente entre comprensión y cobertura léxica, los autores utilizaron dos textos de carácter académico. El primer texto, titulado *What's wrong with our weather*, trataba sobre el cambio climático y el calentamiento global. El segundo texto, *Circuit training*, trataba sobre un estudio de esfuerzo y agudeza mental en ratones de laboratorio.

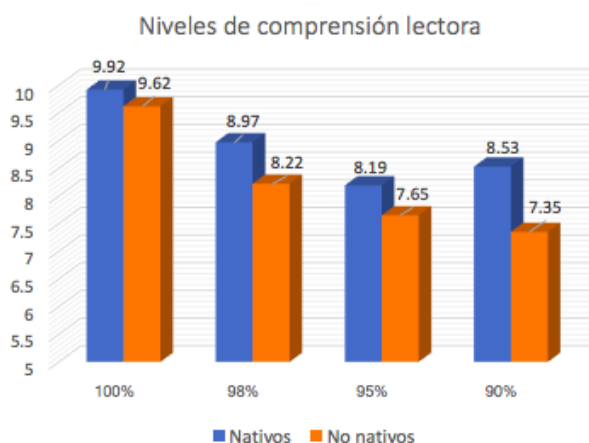
Se les administró un test de conocimiento de vocabulario para determinar cuál era la amplitud léxica que tenían de cada uno de los textos. Esta prueba reveló niveles de cobertura que oscilaban entre el 90 y el 100 %, por lo tanto, los autores decidieron analizar estos diez niveles (del 90 %, 91 %, 92 %, 93 %, etc.) de cobertura léxica de forma separada. Para evaluar la comprensión, los estudiantes realizaron dos pruebas para cada uno de los textos: un test de respuesta múltiple y una actividad de comprensión de la estructura del texto. El objetivo de esta última era que los estudiantes reconocieran la estructura organizativa del texto y proporcionaran respuestas que fueran compatibles con dicha organización. El test de respuesta múltiple estaba compuesto por 14 preguntas y la tarea de comprensión estructural del texto de 16. Así, cada una de las pruebas de comprensión tenía un total de 30 preguntas.

Los autores encontraron que a partir del 90 % existía una correlación entre cobertura léxica y comprensión lectora, es decir, cuanto más léxico conocían los estudiantes, mejores resultados obtenían en las pruebas de comprensión. No se observó ningún punto en el que la comprensión mejorara drásticamente, sino que el crecimiento se produjo de forma

gradual a medida que aumentaba la amplitud léxica. Esto demuestra que la cobertura léxica es un factor clave para la comprensión, ya que las puntuaciones no superaban el 50 % del total cuando sus niveles de conocimiento léxico eran bajos. Además, si se quiere que la totalidad de los estudiantes aspire a una comprensión de más del 60 %, es necesario poseer al menos un 98 % de cobertura léxica. Finalmente, los autores sugieren que es recomendable conocer entre un 98-99 % del vocabulario de un texto para comprenderlo de forma independiente. Del mismo modo, Nation (2006) argumenta que una cobertura del 98 % garantiza que los lectores no requieran de la utilización del diccionario, ya que el propio texto asegura la comprensión. En otras palabras, cuanto mayor es la cobertura léxica, más fácil es interpretar el significado de una palabra desconocida en un texto.

Debemos apuntar que la mayoría de las investigaciones que relacionan el grado de cobertura léxica con la comprensión se han centrado principalmente en la comprensión lectora (Schmitt, Cobb, Horst y Schmitt, 2017). Hasta la fecha, son escasos los estudios que abordan el efecto de la cobertura léxica en el marco de la comprensión oral. Nation (2006) menciona que el lenguaje hablado utiliza habitualmente una mayor proporción de palabras más frecuentes que el lenguaje escrito. Sin embargo, considera la posibilidad de que la cobertura léxica en la comprensión oral requiera un porcentaje superior al 98 %, debido al carácter continuo del lenguaje hablado que no hace posible que el receptor pueda retroceder para reciclar información. Ahora bien, los resultados de la investigación realizada por Van Zeeland y Schmitt (2012) no respaldan esta afirmación. En su estudio, los autores pretendían conocer el volumen léxico necesario para la comprensión oral tanto en hablantes nativos como en aprendientes de inglés como L2. Para ello, utilizaron cuatro pasajes narrativos de carácter informal en los que se introducían palabras inventadas con el fin de proporcionar a cada historia un porcentaje de cobertura léxica predeterminado (100 %, 98 %, 95 % y 90 % respectivamente). Los cuatro pasajes utilizados estaban compuestos por historias cortas contadas en primera persona y hacían referencia a anécdotas de diferentes personajes que se metían en situaciones peculiares. Todas las palabras que aparecían en los textos orales pertenecían a las 2000 familias de palabras más frecuentes. Para garantizar que los participantes conocían la totalidad de las palabras de los textos se les sometió a una prueba de vocabulario en la que se eligieron de forma aleatoria 30 palabras que pertenecían a las 2000 más frecuentes y se les pidió que las emparejaran con su definición correcta. Los estudiantes que no alcanzaron al menos 29 sobre 30 en esta prueba no pudieron participar en el experimento. Para evaluar la comprensión, los estudiantes tuvieron que realizar un test de respuesta múltiple compuesto por 10 preguntas. Cada una de estas tenía un total de cinco opciones (cuatro posibles respuestas y una respuesta intermedia tipo “no lo sé”). A continuación, en la tabla 2 se muestran los niveles de comprensión tanto de hablantes nativos como de aprendientes de L2 en las diferentes modalidades (100 %, 98 %, 95 % y 90 %).

Tabla 2. Niveles de comprensión de hablantes nativos y no nativos para los cuatro niveles de cobertura léxica (extraído de Van Zeeland y Schmitt, 2012)



Los participantes que conocían el 100 % de las palabras de la grabación obtuvieron resultados significativamente mejores, ya fueran hablantes nativos o no. Este resultado demuestra que una cobertura léxica del 100 % permite a los estudiantes la comprensión integral de la grabación, aunque no la garantiza en todos los casos, ya que algunos participantes no alcanzaron este nivel máximo de comprensión. Los participantes que escucharon la grabación con un 98 % de cobertura léxica alcanzaron también niveles de comprensión elevados, aunque existe un descenso significativo en los niveles de comprensión. Además, dentro de este nivel de cobertura, los aprendientes de L2 mostraron mayor variación en los resultados. Ocho de los 40 participantes respondieron correctamente a las 10 preguntas, mientras 11 de ellos no fueron capaces de responder correctamente a más de 7 preguntas, lo que sugiere que el descenso en la cobertura léxica afecta negativamente a la comprensión. Las historias con un 95 % y 90 % de cobertura también mostraron un descenso pronunciado en los niveles de comprensión y la dispersión entre participantes fue más pronunciada. No obstante, las puntuaciones en estos dos últimos niveles fueron todavía bastante elevadas y no se encontraron diferencias significativas en la comprensión entre los sujetos que leyeron los textos con un 95 % y 90 % de cobertura.

En conclusión, los autores argumentan que una cobertura del 95 % puede ser el objetivo a alcanzar en la enseñanza de la comprensión oral en una L2, ya que en este nivel la comprensión es ya relativamente alta (7.65) y se evita la gran variación que se producía en el nivel de cobertura del 90 % (p. 474). Además, proponen este mismo planteamiento para la comprensión lectora, al igual que los resultados aportados por el estudio de Laufer y Ravenhorst-Kalovski (2010). Así, una cobertura del 98 % sería recomendable para una comprensión total pero una cobertura del 95 % sería suficiente para garantizar una comprensión adecuada (Van Zeeland y Schmitt, 2012) siempre y cuando se pueda recurrir a un diccionario u otro soporte externo.

No obstante, hasta la fecha, no existe ningún estudio que determine la cobertura léxica necesaria para entender un texto en español. Consecuentemente, conocer dichas proporciones sería una de las labores más urgentes para poder así establecer pautas similares a las propuestas en inglés.

3.2. Amplitud léxica

Una vez establecidos estos porcentajes (98 % y 95 %), es necesario definir de forma precisa cuál es la magnitud de vocabulario que un aprendiente necesita conocer para cubrir estas proporciones. Una de las primeras publicaciones que aborda la intersección entre la cobertura léxica y la cantidad de vocabulario necesaria para alcanzar diversos niveles de cobertura (i.e. 90 %, 95 % y 98 %) es el trabajo realizado por Hirsh y Nation (1992). Estos autores trataron de descubrir el número de palabras necesario para leer novelas originales sin ningún tipo de soporte externo. Hirsh y Nation eligieron novelas para adolescentes porque se trataba de novelas bastante accesibles para los usuarios: *Alice in Wonderland*, *The Pearl* y *The Haunting*.

Michael West (1953), teniendo en cuenta la importancia de la frecuencia en la comprensión, desarrolló la *General Service List*, que incluye las 2000 familias de palabras más frecuentes del inglés. Esta lista de frecuencias supone un punto de partida para la enseñanza del vocabulario más útil y productivo a aprendientes de inglés como L2, ya que supone la base del idioma y las palabras que incluye se encuentran en una amplia variedad de textos escritos y producciones orales. Con esta lista como referente, Hirsh y Nation analizaron si las 2000 familias propuestas por West eran suficientes para garantizar el 97-98 % de cobertura necesario para comprender un texto sin recursos externos. Los resultados demostraron que el conocimiento de las 2000 familias de palabras de la *General Service List*, incluyendo los nombres propios, suponen una cobertura léxica del 95 % del léxico que aparece en las tres novelas. No obstante, esta proporción no garantiza una comprensión eficiente sin el uso del diccionario. Para conseguir el objetivo del 97-98 % de cobertura léxica es necesario tener un conocimiento de 5000 familias de palabras.

De todos modos, el estudio de Hirsh y Nation cuenta con varias limitaciones metodológicas: (1) sus resultados no son representativos, ya que el corpus utilizado únicamente está compuesto por tres libros; (2) la carga de vocabulario de estas novelas no refleja la realidad lectora de los estudiantes que intentan leer novelas auténticas en inglés, (3) no se tuvieron en cuenta otros géneros discursivos como la prensa y (4) no se contempló la modalidad oral. Consecuentemente, en un trabajo posterior, Nation (2006) decidió investigar el volumen léxico receptivo necesario para comprender novelas, periódicos, películas de dibujos animados y conversaciones. Para ello, desarrolló una tabla en la que ordenó las primeras 14000 familias de palabras del inglés -estos datos fueron extraídos del corpus del *British National Corpus* (BNC)- en bandas de frecuencia de 1000 palabras dependiendo de la frecuencia de uso de estas mismas. Así, las 1000 primeras palabras más frecuentes irían agrupadas en el primer nivel, las siguientes 1000 palabras en el segundo, y así sucesivamente. El 90 % del corpus del BNC está compuesto por textos escritos y el 10 % restante por transcripciones orales. El corpus oral contiene en su mayoría conversaciones y grabaciones de radio y el corpus escrito incluye una amplia variedad de novelas y periódicos.

A continuación, en la tabla 3 presentamos los dos umbrales de volumen léxico -98 % y 95 %- y las palabras necesarias para alcanzar dichos umbrales de comprensión en los distintos tipos de géneros discursivos y modalidades propuestos en este artículo.

Tabla 3. Género discursivo y volúmenes léxicos para un 95 % y un 98 % de cobertura léxica (Nation, 2006)

Género	95 % de cobertura	98 % de cobertura
<i>Lady Chatterley's Lover</i>	4000 + nombres propios	9000 + nombres propios
Novelas	4000 + nombres propios	9000 + nombres propios
Periódicos	4000 + nombres propios	8000 + nombres propios
Lecturas graduadas	2000 + nombres propios	3000 + nombres propios
Películas de dibujos animados	4000 + nombres propios	7000 + nombres propios
Lenguaje oral espontáneo	3000 + nombres propios	6000-7000 + nombres propios

Si queremos alcanzar un 98 % de cobertura léxica, es necesario conocer 9000 familias en la modalidad escrita y de 6000 a 7000 palabras para el lenguaje oral. En efecto, la modalidad escrita posee una variedad y complejidad léxica mayor que la modalidad oral, es decir, los textos escritos requieren un mayor volumen de vocabulario. Si, en cambio, consideramos el 95 % de cobertura léxica como el umbral necesario, el volumen léxico necesario disminuye considerablemente. Para cumplir con este nivel necesitamos conocer 4000 familias de palabras para la modalidad escrita y las películas infantiles, y 3000 para comprender el lenguaje oral espontáneo. Esta última cifra coincide con el estudio de Van Zeeland y Schmitt (2012) y con los resultados encontrados por Webb y Rodgers (2009a, 2009b). Webb y Rodgers realizaron una serie de estudios en los que querían conocer la amplitud léxica necesaria para comprender una amplia variedad de programas de televisión (2009a) y películas (2009b). Para ello, analizaron un total de 88 programas y 318 guiones de películas de diversa índole (acción, comedia, bélicas, animación, suspense, etc.). Los resultados mostraron que el conocimiento de las 3000 familias más frecuentes -incluyendo nombres propios, interjecciones y pausas dubitativas tales como “oh, uh, mmm, etc.”- suponían una cobertura léxica de más del 95 %. Por fin, en el caso de las lecturas graduadas, el volumen léxico es más reducido: un 98 % supondría conocer 3000 familias, mientras que el conocimiento de 2000 familias léxicas supondría una cobertura de un 95 %.

Ahora bien, antes de proceder, es necesaria una discusión sobre los distintos métodos a la hora de concebir la unidad léxica en la que se basan los análisis. La mayoría de los estudios que tratan el volumen léxico y la cobertura de vocabulario proporcionan una categorización basada en la noción de familia de palabra (Hirsh y Nation, 1992; Nation, 2006; Stæhr, 2008; Van Zeeland & Schmitt, 2012; Nation, 2014). Por ejemplo, Nation (2006) utilizó como unidad de medida la familia de palabra con un nivel 6 de acuerdo con la definición propuesta por Bauer y Nation (1993). Este nivel incluye todas las formas flexivas y más de 80 afijos derivativos, lo que lleva a incluir una gran cantidad de palabras dentro de una misma familia léxica. Nation argumenta que decidió utilizar familias de palabras porque asume que cuando los nativos o los aprendientes de una L2 conocen al menos un miembro de esta pueden entender el resto de los componentes de la misma familia si utilizan los mecanismos de formación de palabras del inglés.

Debemos tener en cuenta que la conciencia morfológica puede variar de forma significativa entre sujetos, es decir, muchos aprendientes de una L2 no poseen familias de palabras tan amplias como las de los hablantes nativos (Nation, 2006; Schmitt *et al.*, 2017) e incluso pueden tener problemas a la hora de identificar relaciones morfológicas entre palabras de una misma familia (Sánchez Gutiérrez, 2013). Por lo tanto, desde un marco de enseñanza de lenguas, la utilización de familias de palabras es posible que no represente la realidad lingüística de los aprendientes, ya que el conocimiento de tres palabras de una misma familia (i.e. flor, florero, floristería), por ejemplo, no implica el conocimiento de las demás formas de la misma (i.e. florista, florido, floral, etc.). Por esta razón, diversos autores (Davies, 2005; Schmitt *et al.*, 2017) proponen reemplazar el uso de familias por lemas debido a que no se puede predecir de forma precisa la cantidad de miembros que un aprendiente conoce dentro de una misma familia (Schmitt y Zimmerman, 2002; Schmitt *et al.*, 2017). La agrupación en lemas separa todos los miembros de una familia que constituyen una unidad semántica concreta y que tienen una entrada en el diccionario (i.e. pan, panadero, panadería, empanar) pero mantiene bajo el mismo lema todas las formas flexivas de una unidad léxica. Así, *jugar*, *jugador* y *jugada* tendrán cada uno un lema diferente pero *jugamos*, *jugó* y *jugaré* pertenecerán todos al mismo lema *jugar*.

En español, el único trabajo disponible hasta la fecha que aborda las cuestiones de cobertura y frecuencia léxica en español es el estudio realizado por Davies (2005). En la tentativa de proporcionar información sobre la frecuencia del léxico español, Davies creó un corpus (Corpus del Español, 2002) compuesto por textos y transcripciones de conversaciones de diversos géneros y registros. Cuenta con un total de 100 millones de palabras extraídas de documentos que comprenden del siglo XIII al siglo XX. El propósito de esta sección es ofrecer al usuario una visión panorámica de la historia del idioma español (Corpus del Español, 2002). Los datos recogidos pertenecientes al siglo XX contienen textos escritos y transcripciones orales a partes iguales. Además, incluyen una amplia variedad de géneros (textos académicos, novelas de ficción y periódicos) en ambas modalidades (oral y escrita). De esta forma, es posible hacer una comparación rigurosa de los distintos géneros dentro del español. Después de crear el corpus, Davies desarrolló una tabla en la que ordenó por frecuencia de aparición todos los lemas que se presentaban en el corpus. Así, los 1000 primeros lemas más frecuentes irían agrupados en la primera lista, los segundos 1000 lemas en la segunda, y así sucesivamente. Posteriormente, analizó todas las palabras para observar los niveles de cobertura léxica que ofrecían en dicho corpus. En la tabla 4 se presenta el número de lemas necesario para alcanzar una cobertura léxica del 95 % y 98 % en los textos escritos y el lenguaje hablado.

Tabla 4. Coberturas léxicas para el género oral y escrito en español (Davies, 2005)

Género	95 % cobertura	98 % cobertura
Lenguaje escrito		
• Ficción	9.401	14.349
• No ficción	11.119	18.989
Lenguaje oral	4.699	8.235

Si queremos alcanzar un 98 % de cobertura léxica es necesario conocer 8.235 lemas para el lenguaje oral, 14.349 para las novelas de ficción y 18.989 para las novelas de no ficción. Un 95 % de cobertura léxica supondría conocer 4.699 lemas para el lenguaje oral, 9.401 para las novelas de ficción y 11.119 para las novelas de no ficción. Aunque las 3000 primeras palabras no proporcionen una cobertura del 95 % en los textos orales, debemos notar que Davies (2005) utilizó el lema y no la familia como unidad de agrupación léxica; por lo tanto, una agrupación en familias habría supuesto una cobertura de más del 95 % tanto en los textos orales como escritos.

Ahora bien, las listas de frecuencias utilizadas para este estudio se han desarrollado con un corpus de hace más de 15 años. De hecho, el Corpus del Español (2002) no ofrece una realidad objetiva sobre el español actual, ya que no introduce de forma equitativa textos orales y escritos de las diferentes variedades del español. No obstante, la segunda versión de este corpus -publicada en el año 2016- consta de 2 billones de palabras provenientes de diversas páginas webs de 21 países de habla hispana. Gracias a esta nueva sección, podemos tener una visión global del uso del español en la actualidad, ya que las páginas utilizadas datan de los años 2013 y 2014 y comprenden una amplia variedad de dialectos. Además, esta nueva versión ha venido acompañada por la publicación de la segunda edición del diccionario de frecuencias (Davies y Hayward Davies, 2017), que ofrece pautas más amplias y actualizadas sobre el vocabulario del español. Por lo tanto, es necesario llevar a cabo un estudio que actualice las cuestiones de cobertura y frecuencia léxica y nos ofrezca datos concretos sobre la realidad del español actual.

Los resultados aportados por Davies (2005), Nation (2006) y Van Zeeland y Schmitt (2012) nos ofrecen argumentos evidentes para proponer ciertas orientaciones pedagógicas. Como dijimos previamente en esta sección, es necesaria una cobertura del 98 % de un texto para garantizar una comprensión oral o escrita satisfactoria. Sin embargo, este nivel de cobertura léxica sería difícil de implementar en el marco de la enseñanza de una lengua extranjera, ya que implica el conocimiento de un gran número de palabras. Si comparamos la amplitud léxica necesaria para cada uno de los umbrales -98 % y 95 %- podemos observar que una diferencia de un 3 % en la cobertura léxica supone un aumento drástico en la cantidad de palabras que un hablante debe conocer. Así, las 4000 familias más frecuentes suponen un reconocimiento del 95 % de todas las palabras que aparecen en el inglés oral, mientras que las siguientes 5000 familias conjuntamente aportan un mero aumento del 3 % en la cobertura léxica. Además, existen diferencias notables entre las bandas de frecuencia de las 4000 familias de palabras más frecuentes. Estas diferencias se ilustran en la tabla 5 propuesta por Nation (2006), en la cual se muestra el nivel de cobertura léxica perteneciente a cada una de las bandas de frecuencia definidas por el autor.

Tabla 5. Cobertura y amplitud léxica dividida en bandas de frecuencia (Nation, 2006)

Nivel de frecuencia	Porcentaje en textos escritos (%)	Porcentaje en textos orales (%)
1000	78-81	81-84
2000	8-9	5-6
3000	3-5	2-3
4000-5000	3	1.5-3
6000-9000	2	0.75-1
10000-14000	<1	0.5

Así, las primeras 1000 palabras más frecuentes representan alrededor del 80 % del léxico utilizado en inglés. A partir de esta banda de frecuencia, se puede observar un descenso drástico en la cobertura léxica: las segundas 1000 palabras más frecuentes suponen únicamente alrededor de un 7 % de la amplitud léxica, las terceras 1000 palabras ofrecen aproximadamente un 4 % de cobertura léxica y las cuartas y quintas 1000 palabras más frecuentes, combinadas, solamente suponen un 3 % de cobertura adicional. Estos resultados ponen en evidencia la existencia de palabras de alta frecuencia que son extremadamente útiles para los aprendientes de una L2.

En la siguiente sección explicaremos la importancia del conocimiento de estas palabras de alta frecuencia y explicaremos por qué corresponden con las 2000-3000 palabras más frecuentes.

4. CRITERIOS DE SELECCIÓN LÉXICA EN LA ENSEÑANZA DE UNA L2

Como hemos visto en los apartados 3.1. y 3.2., un número relativamente pequeño de familias de palabras cubre una gran proporción del léxico utilizado en el lenguaje oral y escrito (Stæhr, 2008; Horst, 2013; Schmitt y Schmitt, 2014). Esto se ve reflejado en diversas investigaciones que argumentan que el vocabulario de alta frecuencia comprende las primeras 2000 familias de palabras, ya que el conocimiento de estas supone una cobertura de más del 85 % de un idioma y tiene un gran impacto en la competencia lingüística de un aprendiente (Horst, 2013). Además, el conocimiento de las 2000 familias más frecuentes supone un punto de inflexión en el rendimiento de los aprendientes en las pruebas de comprensión auditiva y expresión escrita. Prueba de ello es el estudio realizado por Stæhr (2008) en el que se muestra la importancia que tiene el conocimiento de las familias de alta frecuencia para los estudiantes de nivel elemental. El autor quería conocer la relación existente entre la amplitud léxica y la competencia lingüística en la comprensión lectora y la expresión escrita en aprendientes daneses de inglés como lengua extranjera. De este modo, los estudiantes completaron diferentes pruebas que correspondían con cada una

de las destrezas analizadas en este estudio. Para la comprensión lectora, los aprendientes realizaron un test de respuesta múltiple. Este estaba compuesto por cuatro partes que incluían varios tipos de texto: (1) textos cortos (2) artículos de periódico y (3) textos narrativos de mayor extensión. Para la actividad de escritura, los aprendientes escribieron una carta (entre 350 y 450 palabras) en la que tenían que solicitar un trabajo. Asimismo, para conocer la amplitud léxica de los aprendientes, realizaron una prueba de vocabulario receptivo.

Los resultados demostraron que existe una estrecha relación entre amplitud léxica y competencia lingüística. Todos los estudiantes que mostraron un conocimiento de las 2000 familias más frecuentes alcanzaron resultados superiores a la media en las dos pruebas, lo que sugiere que el conocimiento de las 2000 palabras más frecuentes parece representar un umbral en el rendimiento de los estudiantes. Por el contrario, más del 60 % de los participantes que no conocían las 2000 familias más frecuentes no obtuvieron resultados por encima de la media de todos los sujetos.

Como podemos observar, son varios (West, 1953; Meara, 1980; Hirsh y Nation, 1992; Davies, 2005; Horst, 2013; Nation, 2006; Stæhr, 2008) los autores que consideran las 2000 familias como umbral de alta frecuencia léxica. De hecho, este umbral ha sido utilizado en numerosas investigaciones y ha servido de punto de partida para la creación de una amplia gama de materiales didácticos (Horst, 2013; Schmitt y Schmitt, 2014) y listas de frecuencia (West, 1953). Además, este umbral se ve reforzado con la creación de la primera prueba de conocimiento léxico propuesta por Nation -llamada “Vocabulary Levels Test” (VLT)- (1983), que fue diseñada para proporcionar una información más detallada sobre el conocimiento léxico receptivo que tienen los estudiantes y que evalúa, entre otras, las 2000 palabras más frecuentes del inglés. Esta prueba ha sido utilizada posteriormente en diversos estudios (Laufer, 1992; Hu y Nation, 2000; Webb y Kagimoto, 2010; Van Zeeland y Schmitt, 2012) y ha servido como punto de partida para la creación de distintas pruebas de nivel léxico (Nation y Beglar, 2007; Mancilla-Martínez y Lesaux, 2011; Schmitt, Schmitt y Clapham, 2011; McLean, Kramer y Beglar, 2015).

Ahora bien, Schmitt y Schmitt (2014) argumentan que la idea de que las 2000 palabras más frecuentes son las que se pueden considerar como palabras de alta frecuencia se basa en listas de frecuencias que se han realizado hace ya más de cinco décadas. Por lo tanto, sería necesario examinar si el vocabulario de alta frecuencia todavía comprende las 2000 familias o, por el contrario, sería conveniente modificar dicho umbral. El único trabajo que propone una revisión y actualización detallada del concepto de vocabulario de alta frecuencia es el estudio realizado por Schmitt y Schmitt (2014). En su investigación, cuestionan la validez de las 2000 primeras familias como límite de vocabulario más frecuente y argumentan que se debería ampliar el umbral de alta frecuencia a las 3000 primeras familias de palabras. Para justificar esta decisión, los autores ofrecen cuatro argumentos de peso: (1) la cobertura léxica, (2) las lecturas graduadas, (3) los diccionarios de lengua extranjera y (4) la frecuencia y el aprendizaje incidental.

Cobertura léxica- Un número reducido de familias de palabras constituye la mayoría del léxico utilizado tanto en el discurso oral como en el escrito (Davies, 2005; Nation, 2006; Stæhr, 2008; Schmitt y Schmitt, 2014). Por consiguiente, Schmitt y Schmitt argumentan que el vocabulario de alta frecuencia es el que aparece antes de que las proporciones de

cobertura desciendan de forma sustancial y se conviertan en niveles marginales. Nation (2006) argumenta que a partir de las 3000 familias más frecuentes la cobertura léxica desciende significativamente, ya que las siguientes 2000 familias únicamente proporcionan un 3 % de cobertura y las 4000 siguientes solo un 2 %. Por lo tanto, Schmitt y Schmitt (2014) argumentan que el vocabulario de alta frecuencia debería incluir las primeras 3000 familias. Esta conclusión también coincide con los resultados expuestos por Davies (2005) en los que se muestra que a partir de los 3000 lemas más frecuentes la cobertura léxica desciende de forma significativa. Además, en el caso del inglés, el conocimiento de las primeras 3000 familias supone una cobertura léxica del 95 % en los textos orales (Nation, 2006; Webb y Rodgers, 2009a, 2009b; Van Zeeland y Schmitt, 2012).

Lecturas graduadas- Las lecturas graduadas son uno de los métodos más efectivos a la hora de aprender vocabulario de forma incidental. Estos textos simplificados ofrecen a los aprendientes las herramientas necesarias para aprender vocabulario de forma gradual mientras disfrutan de un texto que se adapta a sus capacidades. Schmitt y Schmitt (2014) analizaron la amplitud léxica utilizada por cinco editoriales de lecturas graduadas y observaron que la mayoría de las lecturas graduadas finalizaban alrededor de las 3000 primeras familias más frecuentes.

Diccionarios de lengua extranjera- Schmitt y Schmitt (2014) también estudiaron la longitud y contenido de los diccionarios para estudiantes de inglés como L2, ya que ofrecen una perspectiva sobre las palabras que los lexicógrafos especializados en L2 consideran primordiales para los estudiantes de inglés. Así, analizaron la amplitud léxica de tres diccionarios de inglés como L2 y observaron que más del 90 % de las entradas pertenecían a las 3000 familias de palabras más frecuentes, y que las 1000 palabras siguientes solo ofrecían una cobertura del 5 %. Estos resultados refuerzan el hecho de que las 3000 familias más frecuentes deben suponer el umbral de vocabulario de alta frecuencia.

Frecuencia y aprendizaje incidental- Finalmente, otro método a la hora de averiguar el umbral de alta frecuencia es calcular la asiduidad con la que los estudiantes pueden encontrar determinado vocabulario de forma incidental (Schmitt y Schmitt, 2014) ya que el aprendizaje incidental es un proceso importante en la adquisición de nuevo vocabulario (Nassaji, 2006; Pellicer-Sánchez y Schmitt, 2010). Debemos notar que el aprendizaje incidental de vocabulario está estrechamente relacionado con la frecuencia léxica; es decir, la repetición afecta positivamente al aprendizaje (Horst, Cobb, y Meara, 1998; Webb, 2007; Horst, 2013); y en la bibliografía actual, la mayoría de los autores parecen sugerir que para el aprendizaje de varios aspectos léxicos -tales como el conocimiento receptivo de la forma y del significado- una palabra debe aparecer repetida al menos 10 veces.

Ahora bien, puesto que las 3000 primeras familias suponen la base de un idioma, tanto los instructores como los libros de texto deberían potenciar el aprendizaje de estas unidades léxicas y garantizar que se presentan con una frecuencia que garantice su adquisición. De este modo, es necesario conocer si los cursos de enseñanza de lenguas y los materiales didácticos ofrecen a los estudiantes las condiciones necesarias para favorecer el aprendizaje de este vocabulario. Por consiguiente, en la siguiente sección analizaremos en qué medida el instructor y los materiales didácticos promueven el aprendizaje del vocabulario de alta frecuencia.

5. TRATAMIENTO DEL VOCABULARIO DE ALTA FRECUENCIA

Los manuales de enseñanza gozan de una importancia considerable en la planificación de los currículos de enseñanza de lenguas. Su presencia en el aula condiciona tanto la metodología como las decisiones pedagógicas de los instructores (Macian, 1986; Issitt, 2004; LaBelle, 2011; Cubillos, 2014), juegan un papel fundamental en la enseñanza del contenido léxico (Van Patten, 2015) y suelen articular los cursos de lenguas extranjeras cuando se desarrollan varias secciones simultáneamente (Marcos Miguel, 2015). De hecho, Guerrettaz y Johnston (2013) demostraron que los manuales se pueden convertir en el centro de la mayoría de las interacciones entre los alumnos y el instructor, siendo estos el objeto de discusión y el soporte que estructura el transcurso de la clase y del aprendizaje. En conclusión, comprender cuál es el tratamiento del léxico en los manuales de enseñanza debería proporcionar una visión general sobre lo que ocurre en el aula a este respecto.

Matsuoka y Hirsh (2010) examinaron las oportunidades de aprendizaje léxico en un manual de inglés como L2 diseñado para estudiantes de nivel intermedio-alto. Los autores calcularon las repeticiones de las familias de palabras que pertenecían a las segundas 1000 familias, utilizando la lista de frecuencias propuesta por West (1953). Los resultados mostraron que el 64 % -si consideramos 7 repeticiones como frecuencia mínima para favorecer el aprendizaje de vocabulario receptivo- de las segundas 1000 palabras tienen un tratamiento insuficiente en el libro de texto, lo cual podría limitar las oportunidades de aprendizaje léxico de los estudiantes.

En este mismo sentido, Martini (2012) analizó el léxico de tres manuales de secundaria diseñados para estudiantes de inglés como lengua extranjera en Quebec, Canadá. La autora quería observar el nivel de cobertura léxica que ofrecían los manuales y si el vocabulario se repetía lo suficiente como para favorecer su aprendizaje, es decir, si las palabras se reciclaban al menos 10 veces a lo largo de los manuales. Para ello, filtró todas las palabras de los manuales por un software llamado RANGE (Heatley, Nation y Coxhead, 2002) que clasifica las familias de palabras dependiendo del nivel de frecuencia y proporciona el número de repeticiones de cada palabra.

Tabla 7. Distribución de la frecuencia, cobertura léxica y frecuencia de aparición (Martini, 2012)

Banda de frecuencia	Cobertura léxica en los manuales	Palabras con más de 10 repeticiones
1000	98 %	94 %
2000	95 %	65 %
3000	85 %	27 %

Los resultados demuestran que los libros de texto presentan una gran proporción del vocabulario de alta frecuencia ya que un 98 % de las primeras 1000 familias aparecen en los manuales de texto. Del mismo modo, los manuales incluyen un 95 % de las segundas

2000 familias. En el caso de las terceras 1000, aunque la proporción disminuye ligeramente, el número de familias representadas en los libros supone un 85 %.

Si nos centramos en la frecuencia de exposición al léxico de alta frecuencia, podemos observar que el 94 % del vocabulario de las primeras 1000 palabras se repiten al menos diez veces a lo largo de los manuales. Por el contrario, solo un 65 % de las segundas 1000 familias (alrededor de 605) se repiten 10 o más veces y esta proporción disminuye al 27 % para las familias del tercer nivel. Esto nos sugiere que los manuales no ofrecen a los estudiantes las repeticiones suficientes para favorecer el aprendizaje de palabras que van más allá de las 1000 primeras.

Estos resultados ponen en evidencia el tratamiento del vocabulario de alta frecuencia: aunque la mayoría de las palabras que aparecen en los manuales pertenecen a las 3000 familias más frecuentes, su representación no es consistente, ya que no aparecen lo suficiente como para garantizar su aprendizaje. Esto puede provocar que los estudiantes no vuelvan a encontrar muchas de las palabras que han visto previamente y, por lo tanto, es muy probable que con la mera exposición al libro de texto no se produzca la adquisición de nuevo vocabulario. Además, dado que los manuales de enseñanza guían las prioridades didácticas de los instructores (Guerrettaz y Johnston, 2013; Van Patten, 2015), si un contenido no se repite de forma continua, es probable que los profesores no lo revisen por decisión propia.

En el caso del español, el único trabajo disponible hasta la fecha es el de Davies y Face (2006). Los autores examinaron el vocabulario de seis manuales de enseñanza de español universitarios: tres pertenecientes al primer año y tres al segundo. Para ello, los autores analizaron el vocabulario activo, es decir, el que aparece en las listas de vocabulario de cada uno de los temas. Después de extraer estas palabras de los seis manuales, agruparon las unidades léxicas en lemas.

A continuación, en la tabla 8 se representan los niveles de cobertura léxica para los diferentes rangos de frecuencia.

Tabla 8. Rango de frecuencia y niveles de cobertura léxica en los manuales (Davies y Face, 2006)

Rango de frecuencia	Cobertura léxica en los manuales
1000	87.5 %
2000	66.5 %
3000	52 %
4000	38.5 %
5000	30 %

La representación del vocabulario en los manuales disminuye de forma sustancial a medida que asciende el rango de frecuencia. El 87.5 % de los primeros 1000 lemas aparecen incluidos en los manuales de enseñanza. Esta proporción disminuye al 66.5 % para los segundos 1000 lemas y al 52 % para los 3000. A partir de los 3000 lemas, la proporción desciende drásticamente hasta llegar a los 5000, donde no supera el 30 %. Estos resultados demuestran que los manuales no ofrecen las herramientas adecuadas para el aprendizaje de las palabras más frecuentes. Además, debemos notar que existen diferencias sustanciales entre los manuales de primer y segundo año. Los manuales de primer año ofrecen una cobertura léxica más amplia que los libros de texto de segundo año, ya que los libros del segundo nivel cubren temáticas que están relacionadas con cuestiones que requieren un vocabulario específico y, en muchos casos, de baja frecuencia. Esta última observación indica que el rol de los campos semánticos juega un papel fundamental en la introducción de vocabulario en un manual de enseñanza. En efecto, los libros de texto organizan su vocabulario en torno a estas temáticas previamente definidas (Davies y Face, 2006: 141) lo cual conlleva que se introduzcan términos sin tener en cuenta la frecuencia de estas palabras. Así, por ejemplo, en el campo semántico relacionado con la comida, es habitual encontrar vocabulario de baja frecuencia (calabaza, langosta, asado, etc.). Consecuentemente, los estudiantes están expuestos a un vocabulario artificial que no se ajusta al léxico que van a encontrar en situaciones comunicativas reales.

Además de los libros de texto, la otra fuente principal de *input* para los estudiantes es el habla del instructor. Como se ha explicado previamente, la exposición a *input* oral favorece el aprendizaje incidental de vocabulario (Brown, Waring, y Donkaewbua, 2008; Horst, 2010; Horst, 2013), lo cual complementa el aprendizaje que se debe a la enseñanza explícita a través de ejercicios. Así, el habla del instructor en el aula va a ser un buen indicador de las oportunidades de aprendizaje léxico de las que disponen los aprendientes, especialmente cuando la mayor parte de la exposición a su L2 tiene lugar en el aula (Meara, Lightbown y Halter, 1997).

Puesto que la frecuencia de aparición está estrechamente relacionada con el aprendizaje léxico, es necesario tener en cuenta cuáles son las oportunidades que tienen los estudiantes a la hora de aprender vocabulario mediante la exposición al habla del instructor. Hasta la fecha, el mejor trabajo que aborda la cobertura léxica y las oportunidades de aprendizaje léxico mediante el habla del instructor es el estudio realizado por Horst (2010). Aunque Meara, Lightbown y Halter (1997) ya habían investigado previamente sobre el habla del profesor, su corpus estaba compuesto por grabaciones de 10 instructores impartiendo 10 clases distintas. En consecuencia, su estudio no nos proporciona información sobre el comportamiento de un instructor en particular, lo que hace imposible determinar cuáles son las oportunidades de aprendizaje léxico real que ofrece un instructor a sus estudiantes en un curso concreto. En el caso de Horst (2010), la autora analizó el léxico de un corpus formado por el habla de un instructor en un curso de conversación de 9 semanas para aprendientes avanzados de inglés como L2. En particular, se centró en la cobertura léxica del habla del instructor, las oportunidades de aprendizaje de vocabulario desconocido que ofrecía el profesor, la frecuencia con la que este léxico se repetía a lo largo del curso y el espaciamiento del vocabulario utilizado. Además, examinó cuál era el tipo de discurso que proporcionaba más vocabulario desconocido: (1) gestión de la clase, (2) explicación de cuestiones relacionadas con la gramática, la pronunciación o la corrección de errores y (3)

tareas de comprensión. Posteriormente, Horst utilizó las listas de frecuencia creadas por Nation (2006) y el programa RANGE para clasificar las familias de palabras dependiendo de su nivel de frecuencia y proporcionar el número de repeticiones de cada una.

En cuanto a la cobertura léxica, las 2000 familias de palabras más frecuentes cubrían más del 95 % del vocabulario de la instructora y las 4000 más frecuentes el 98 % del total. Puesto que los estudiantes eran de un nivel intermedio-alto, se esperaba que tuvieran un conocimiento de las 2000 primeras familias y de la mayoría de las 2000 familias siguientes. Por lo tanto, se puede concluir que el habla de la instructora era comprensible y más que adecuado para el nivel de los aprendientes. En cuanto a la introducción de vocabulario desconocido, los resultados mostraron que los aprendientes estuvieron expuestos a una amplia variedad de familias. Si tenemos en cuenta el 95 % de cobertura léxica como umbral de comprensión, los estudiantes se encontraron con un total de 949 familias de palabras desconocidas. Si consideramos un 98 % como umbral de comprensión, esta cifra se reduce a 456 familias en total. Aunque el vocabulario de la instructora era variado, la frecuencia con la que se repetía a lo largo del curso no favorecía el aprendizaje incidental. En efecto, casi la mitad de las familias de palabras desconocidas se repitieron entre 1 y 2 veces. Por el contrario, solo un 13-14 % de ellas aparecieron repetidas entre 6 y 9 veces. Además, esta cifra disminuye al 11-12 % cuando la frecuencia asciende a 10 o más repeticiones. Sorprendentemente, un 65 % del vocabulario que se reciclaba 6 o más veces a lo largo del curso pertenecía a términos utilizados comúnmente en una clase de lengua; como es el caso de “verbo”, “gramática”, “pronunciación”, “adjetivo”, etc. Del léxico que se reciclaba, solo un número muy reducido de palabras parecía ser desconocido, ya que se esperaba que los estudiantes tuvieran un conocimiento de la mayoría de las 4000 primeras familias de palabras.

En cuanto a los tipos de discurso, Horst observó que más del 50 % del tiempo total se empleaba en dar explicaciones sobre el curso y sobre las tareas que los estudiantes tenían que realizar en el aula. Del mismo modo, la instructora utilizaba más de un 40 % del tiempo total en la explicación de cuestiones relacionadas con la gramática, la pronunciación o la corrección de errores. Por el contrario, solo un 5 % del total se empleó en tareas relacionadas con la comprensión lectora y auditiva. Si nos centramos en la exposición a nuevo vocabulario, se pueden observar diferencias significativas. Aunque la explicación de cuestiones relacionadas con la gramática y la corrección de errores es el género que proporcionó más familias desconocidas (149 en total), las tareas de comprensión (solo suponen un 5 % total del corpus) introdujeron un total de 122 familias desconocidas. Paradójicamente, a pesar de que más de la mitad de las intervenciones iban destinadas a la gestión de la clase, el vocabulario utilizado en esta categoría discursiva no incluía muchas palabras nuevas (alrededor de 54 familias de palabras desconocidas).

En definitiva, los resultados de estos estudios demuestran que los manuales de enseñanza y el habla del instructor ofrecen nuevo vocabulario pero no proporcionan las repeticiones suficientes para asegurar su adquisición y retención a largo plazo. Además, los estudiantes están expuestos a un vocabulario que no se ajusta al léxico más frecuente que van a encontrar en situaciones comunicativas reales (Davies y Face, 2006), por lo tanto, no es sorprendente que haya aprendientes que no dominen las 2000-3000 palabras más frecuentes tras varios años de instrucción (Alsaif, 2011).

Stæhr (2008), por ejemplo, menciona que en Dinamarca se espera que los estudiantes de escuela secundaria tengan un conocimiento léxico de las 2000 primeras familias de palabras antes de finalizar sus estudios. Sin embargo, sus resultados muestran que el 65 % de los estudiantes que participaron en el estudio no tenían un conocimiento del vocabulario más frecuente, a pesar de haber estudiado inglés durante siete años. Esta conclusión también corresponde con los resultados encontrados por Laufer (2000). En la tentativa de proporcionar el conocimiento léxico de los aprendientes de inglés como L2, Laufer recopiló los niveles de amplitud léxica de estudiantes de varios países del mundo (Tabla 9).

Tabla 9. Amplitud léxica de aprendientes de inglés como L2 (Laufer, 2000)

Contexto de aprendizaje	Amplitud léxica	Horas de instrucción
Indonesia: Universidad	1.220	900
Francia: Instituto	1000	400
China: Grado en inglés	4000	1.800-2.400
Grecia: Instituto	1.680	660
Omán: Universidad	2000	1.350+
Alemania: Instituto	1.200	400
Japón: Universidad	2000	800-1.200

Los resultados confirman que la mayoría de los aprendientes de inglés como L2 no dominan las 3000 familias de palabras más frecuentes ni siquiera tras haber sido expuestos a 1000 horas de instrucción. Además, debemos tener en cuenta que la amplitud léxica de la mayoría de los estudiantes no queda restringida únicamente a las familias de alta frecuencia. En efecto, es muy improbable que un aprendiente conozca todas las familias de alta frecuencia y, de hecho, es muy probable que muestre un léxico descentralizado, es decir, que su vocabulario esté repartido por las distintas bandas de frecuencia (Cobb y Horst, 2011). Prueba de ello es el estudio realizado por Cobb y Horst (2011) en el que se puede observar esta tendencia. Los autores observaron que los estudiantes mostraban un conocimiento léxico repartido por las distintas bandas de frecuencia. Así, por ejemplo, aunque un estudiante conozca 2.409 familias de palabras, el hecho de que todas estas no pertenezcan a las familias de palabras más frecuentes va a tener repercusiones en sus niveles de comprensión oral y escrita (Cobb y Horst, 2011). Estas conclusiones muestran que los aprendientes de una L2 necesitan una orientación más específica en el aprendizaje de vocabulario. Puesto que los libros de texto y el habla del instructor no fomentan el aprendizaje léxico, es necesario introducir estrategias pedagógicas complementarias que potencien la adquisición de vocabulario (Nation, 2001; Stæhr, 2008; Horst, 2013).

6. CONCLUSIÓN

El presente trabajo ofrece una visión de conjunto de los resultados de investigación obtenidos en el ámbito de la cobertura léxica, de la frecuencia léxica y de la selección léxica en materiales de enseñanza. Además, pretende concienciar sobre la escasez de investigaciones centradas en la enseñanza/aprendizaje de vocabulario en español y propone mejoras en el ámbito de la enseñanza de ELE.

Numerosos autores argumentan que la comprensión se produce de forma satisfactoria cuando el vocabulario conocido cubre entre el 95-98 % del texto (Laufer, 1989; Nation, 2006; Laufer y Ravenhorst-Kalovski, 2010; Schmitt, Jiang y Grabe, 2011). En efecto, una cobertura del 98 % asegura la comprensión total de un texto oral o escrito y una cobertura del 95 % garantiza una comprensión adecuada siempre y cuando sea posible la utilización de soportes externos.

Para poder cubrir el 98 % de un texto, Nation (2006) argumenta que es necesario conocer 9000 familias en la modalidad escrita y 6000-7000 en la modalidad oral. Para alcanzar un 95 % necesitamos conocer 4000 familias para la modalidad escrita y 3000 para el lenguaje oral. Si comparamos ambos umbrales, podemos observar que una diferencia de un 3 % en la cobertura supone un aumento drástico en la cantidad de palabras que un hablante debe conocer. Además, las primeras 1000 palabras representan alrededor del 80 % del léxico utilizado en inglés y español, mientras que las segundas y terceras 1000 palabras más frecuentes solo ofrecen un 7 % y 4 % de cobertura respectivamente. Estos resultados ponen en evidencia la existencia de palabras de alta frecuencia que son extremadamente útiles para los aprendientes de L2. Tradicionalmente comprendía las primeras 2000 familias, sin embargo, Schmitt y Schmitt (2014) proponen aumentar esta cifra a las 3000 familias ya que estas proporcionan un 95 % de cobertura léxica en el lenguaje espontáneo y aparecen en una amplia variedad de textos orales y escritos.

Desde un punto de vista pedagógico, los libros de texto el habla del instructor son un soporte esencial en la enseñanza de vocabulario; por lo tanto, estos deberían potenciar el aprendizaje de estas palabras y garantizar que se presentan con una frecuencia adecuada. Sin embargo, diversos estudios (Davies, 2005; Matsuoka y Hirsh, 2010; Martini, 2012) demuestran que los manuales de enseñanza no ofrecen a los estudiantes las herramientas necesarias para aprender vocabulario. Consecuentemente, es habitual que haya aprendientes que no dominen las 3000 palabras más frecuentes tras varios años de instrucción, lo que sugiere que es necesario introducir estrategias pedagógicas complementarias que potencien la adquisición de vocabulario.

Ahora bien, la mayoría de los trabajos que hacen referencia al conocimiento y el tratamiento del léxico se han desarrollado predominantemente en inglés. Hasta la fecha, no existen estudios que analicen el volumen léxico que conocen los estudiantes de español como lengua extranjera. Además, los únicos estudios que han tratado la cobertura léxica y el tratamiento del léxico en los manuales están desactualizados y no representan la realidad actual. Consecuentemente, es necesario crear una agenda de investigación en la que (1) se propongan más estudios sobre la cobertura léxica del español en los que se tenga en cuenta de forma equitativa todas las variedades del español, (2) se creen nuevos corpus de frecuencia, (3) se lleven a cabo estudios que analicen el volumen léxico que conocen los

estudiantes de español como L2 y que ofrezcan una visión general sobre la evolución léxica de un estudiante a lo largo de su etapa de aprendizaje y (4) se realicen más estudios relacionados con el tratamiento del vocabulario en los manuales de texto, ya que cuantos más artículos aborden esta problemática, más probabilidades habrá de que las editoriales mejoren los materiales disponibles. Además, los cursos de formación de profesorado deberían enfatizar estos aspectos para que los docentes sean más conscientes del vocabulario que tienen que potenciar en el aula.

Referencias bibliográficas

Alsaif, Abdullah. 2011. *Investigating vocabulary input and explaining vocabulary uptake among EFL learners in Saudi Arabia*. PhD dissertation, Swansea University, Swansea, UK.

Bauer, Laurie y Paul Nation. 1993. Word families. *International Journal of Lexicography* 6, 4: 253-279.

Brown, Ronan, Rob Waring y Sangrawee Donkaewbua. 2008. Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language* 20, 2: 136-163.

Cobb, Tom y Marlise Horst. 2011. Does Word Coach coach words? *CALICO Journal* 28, 3: 639-661.

Cubillos, Jorge. H. 2014. Spanish textbooks in the US: enduring traditions and emerging trends. *Journal of Spanish Language Teaching* 1, 2: 205-225.

Davies, Mark. 2002. *Corpus del Español*. <http://www.corpusdelespanol.org>

Davies, Mark. 2005. Vocabulary range and text coverage: Insights from the forthcoming Routledge Frequency Dictionary of Spanish. *Selected proceedings of the 7th Hispanic Linguistics Symposium*, ed. David Eddington, 106-115. Somerville, MA: Cascadilla Proceedings Project.

Davies, Mark y Timothy Face. 2006. Vocabulary coverage in Spanish Textbooks: How representative is it? *Selected proceedings of the 9th Hispanic Linguistics Symposium*, ed. Nuria Sagarra y Alemadía Jacqueline Toribio, 132-143. Somerville, MA: Cascadilla Proceedings Project.

Davies, Mark y Kathy Hayward Davies. 2017. *A frequency dictionary of Spanish: Core vocabulary for learners, Second Edition*. New York: Routledge.

Guerrettaz, Anne Marie y Bill Johnston. 2013. Materials in the classroom ecology. *The Modern Language Journal* 97, 3: 779-796.

Heatley, Alex, Paul Nation y Averil Coxhead. 2002. Range [Computer software]. Recuperado de <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>

Hilton, Heather. 2008. The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal* 36, 2: 153-166

Hirsh, David y Paul Nation. 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8, 2: 689-696.

Horst, Marlise, Tom Cobb y Paul Meara. 1998. Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language* 11, 2: 207-223.

Horst, Marlise. 2010. How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language* 22,1: 161-180.

Horst, Marlise. 2013. Mainstreaming second language vocabulary acquisition. *The Canadian Journal of Applied Linguistics* 16, 1: 171-188.

Hu, Marcella y Paul Nation. 2000. Vocabulary density and reading comprehension. *Reading in a Foreign Language* 23, 1: 403-430.

Instituto Cervantes. 2017. *El español en el mundo: Anuario del instituto Cervantes*. Madrid: Centro Virtual Cervantes.

http://www.cervantes.es/sobre_instituto_cervantes/prensa/2017/noticias/Presentación-Anuario-2017.htm

Issitt, John. 2004. Reflections on the study of textbooks. *History of Education* 33, 6: 683-696.

LaBelle, Jeffrey T. 2011. Selecting ELL textbooks: A content analysis of language-teaching models. *Bilingual Research Journal* 34, 1: 94-110.

Laufer, Batia. 1989. What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (eds.), *Special language: From humans thinking to thinking machines*, 316-323. Clevedon: Multilingual Matters.

Laufer, Batia. 1992. How Much Lexis Is Necessary for Reading Comprehension? Pierre Arnaud & Henri Béjoint (eds.), *Vocabulary and Applied Linguistics*, 126-132. Basingstoke: Macmillan.

Laufer, Batia. 2000. Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. *Selected Papers from AILA '99 Tokyo*, 47-62. Tokyo, Japan: Waseda University Press.

Laufer, Batia y Geke C. Ravenhorst-Kalovski. 2010. Lexical threshold revisited: Lexical coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22: 15-30.

Liu, Na y Paul Nation. 1985. Factors affecting guessing vocabulary in context. *RELJ Journal* 16: 33-42.

Macian, Janice L. 1986. An analysis of state adopted foreign language textbooks used in first- and third-year high school Spanish classes. *Foreign Language Annals* 19, 2: 103-118.

Mancilla-Martinez, Jeanette y Lesaux, Nonie. 2011. Early home language use and later vocabulary development. *Journal of Educational Psychology* 103, 3: 535-546.

Marcos Miguel, Nausica. 2015. Voluntad y viabilidad en la enseñanza de morfología léxica en el aula de español como L2. *Anexos de la Revista de Lexicografía* 37: 145-158.

Martini, Julianne. 2012. *High frequency vocabulary in a secondary Quebec ESL textbook corpus*. Master's thesis. Concordia University, Montreal, Canadá.

Matsuoka, Warren y David Hirsh. 2010. Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language* 22, 1: 56-70.

McLean, Stuart, Brandon Kramer y David Beglar. 2015. The creation and validation of a listening vocabulary levels test. *Language Teaching Research* 19, 6: 741-760.

- Meara, Paul. 1980. Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics Abstracts*, 13: 221-246.
- Meara, Paul, Patsy Lightbown y Randall Halter. 1997. Classrooms as lexical environments. *Language Teaching Research* 1: 28-47.
- Nassaji, Hossein. 2006. The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Modern Language Journal* 90, 3: 387-401.
- Nation, Paul. 1983. Testing and teaching vocabulary. *Guidelines* 5, 1: 12-25.
- Nation, Paul. 2001. *Learning Vocabulary in Another Language*. Cambridge: University Press.
- Nation, Paul. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63, 1: 59-82.
- Nation, Paul. 2014. How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language* 26, 2: 1-16.
- Pellicer-Sánchez, Ana y Norbert Schmitt. 2010. Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language* 22, 1: 31-55.
- Rieger, Caroline. 2003. Disfluencies and hesitation strategies in oral L2 tests. *Gothenburg Papers in Theoretical Linguistics: Proceedings of the 2003 Disfluency in Spontaneous Speech Workshop*, ed. Robert Eklund, 41-4. Goteborg: Goteborg University.
- Sánchez Gutiérrez, Claudia. 2013. *Priming morfológico y conciencia morfológica. Una investigación con Estudiantes Norteamericanos de E/LE*. Tesis doctoral, Universidad de Salamanca, España.
- Saville-Troike, Muriel. 1984. What really matters in second language learning for academic achievement? *TESOL Quarterly* 18: 199-219.
- Schmitt, Norbert, Diane Schmitt y Caroline Clapham. 2001. Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing* 18, 1: 55-88.
- Schmitt, Norbert y Cheryl Boyd Zimmerman. 2002. Derivative word forms: What do learners know? *TESOL Quarterly* 36: 145-171.
- Schmitt, Norbert, Xiangying Jiang y William Grabe. 2011. The percentage of words known in a text and reading comprehension. *Modern Language Journal* 95, 1: 26-43.
- Schmitt, Norbert y Diane Schmitt. 2014. A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching* 47, 4: 484-503.
- Schmitt, Norbert, Tom Cobb, Marlise Horst y Diane Schmitt. 2017. How much vocabulary is needed to use English? Replication of Van Zeeland & Schmitt (2012), Nation, (2006), and Cobb (2007). *Language Teaching* 50, 2: 212-226.
- Stæhr, Lars Stenius. 2008. Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal* 36: 139-152.
- Van Patten, Bill. 2015. Hispania white paper: Where are the experts? *Hispania*, 98, 1: 2-13.

Van Zeeland, Hilde y Norbert Schmitt. 2012. Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics* 34, 4: 457-479.

Webb, Stuart. 2007. The effects of repetition on vocabulary knowledge. *Applied Linguistics* 28, 1: 46-65.

Webb, Stuart y Michael P. H. Rodgers. 2009a. Vocabulary demands of television programs. *Language Learning* 59: 335-366.

Webb, Stuart y Michael P. H. Rodgers. 2009b. The lexical coverage of movies. *Applied Linguistics* 30: 407-427.

Webb, Stuart y Eve Kagimoto. 2010. Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics* 32, 3: 259-276.

West, Michael. 1953. *A general service list of English words*. London: Longman, Green.

Wilkins, D. 1972. *Linguistics in language teaching*. London: Arnold.