

Análisis del vocabulario en libros de texto para la enseñanza AICLE en 5º de Educación Primaria

Analysis of vocabulary in CLIL textbooks for 5th year of Primary Education.

(1)Verónica García-Bermejo

(1)Universidad de Extremadura

Correspondencia: Verónica García-Bermejo

Mail: vgarciasn@alumnos.unex.es

Recibido: 09-09-2015. Aceptado: 18-12-2015

Resumen.

En el presente estudio se analiza el vocabulario al que los alumnos están expuestos en dos libros de texto utilizados en AICLE (Aprendizaje Integrado de Contenidos y Lenguas Extranjeras), pertenecientes a las asignaturas de “Natural Science” y “Social Science” para 5º curso de educación primaria. El vocabulario de estos libros de texto se analiza desde una perspectiva de frecuencia, teniendo en cuenta también el tamaño de vocabulario necesario para comprender el texto en ambos libros en conjunto y en cada uno de ellos por separado. Para el estudio del vocabulario presente en ambos libros de texto se elaboró un corpus lingüístico y posteriormente se procedió a su análisis. Los resultados obtenidos en nuestro estudio parecen indicar que el libro de texto perteneciente a la asignatura de “Social Science” es más exigente en lo que respecta a vocabulario que la asignatura de “Natural Science”.

Palabras Clave: AICLE, primaria, frecuencia, cobertura léxica, libros de texto

Abstract.

In the present study, the vocabulary to which students are exposed in two different textbooks used in CLIL (Content and Language Integrated Learning) subjects will be analysed. The textbooks belong to ‘Natural Science’ and ‘Social Science’ subjects for 5th year of Primary Education. The analysis of the vocabulary is carried out from a frequency perspective, bearing also in mind the vocabulary size necessary to understand properly both textbooks and each one of them separately. In order to analyse the vocabulary, a linguistic corpus containing both textbooks was compiled. The results obtained in our analysis seem to indicate that the ‘Social Science’ textbook is more demanding concerning vocabulary than the ‘Natural Science’ textbook.

Keywords: CLIL, primary education, frequency, lexical coverage, textbooks

1. INTRODUCCIÓN

El Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE) es un enfoque educativo que se está extendiendo a lo largo de Europa en las últimas décadas (Pérez Cañado, 2012; Heras y Lasagabaster, 2014). En la actualidad existen numerosos centros bilingües así como secciones bilingües en centros tanto públicos como concertados o privados. En ambos, centros bilingües y en las secciones bilingües, el enfoque educativo AICLE es esencial. A pesar de que el objetivo principal en los contextos AICLE es el aprendizaje de contenido, la enseñanza de la lengua extranjera también se ve involucrada. Por ello, es necesario analizar los efectos que la enseñanza AICLE puede tener sobre el aprendizaje de la misma. Los libros de texto son una de las fuentes principales por la que los alumnos están expuestos al idioma, utilizándolos tanto en el aula como fuera de ella y estudiando el contenido principalmente a través de ellos. En el presente estudio nos centraremos en el vocabulario al que los alumnos están expuestos en los libros de texto utilizados en dos asignaturas en AICLE, adoptando una perspectiva de frecuencia.

Nuestro objetivo es analizar y comparar el vocabulario de dos libros de texto pertenecientes a distintas asignaturas pero destinados al mismo alumnado. Nos centraremos en la cobertura que cada banda de frecuencia tiene en ambos libros de texto juntos y en la que presenta en cada uno por separado. También se intentará establecer el vocabulario necesario para alcanzar una adecuada comprensión del texto en ambas asignaturas.

En el presente artículo comenzaremos con una breve revisión de la teoría publicada acerca de los principales aspectos relevantes para nuestro estudio, seguido del planteamiento de las preguntas de investigación a las que hemos pretendido dar respuesta y una explicación del método empleado en nuestro análisis. Posteriormente, presentaremos los resultados obtenidos en nuestro análisis y las conclusiones que se pueden establecer a partir de los mismos. Finalmente, se indicarán ciertas limitaciones de nuestro estudio y posibles futuras líneas de investigación.

2. COBERTURA LÉXICA DE UN TEXTO

Frecuencia

El concepto de frecuencia se ha analizado en los estudios de vocabulario desde aproximadamente la década de los noventa debido a su importancia en la enseñanza y aprendizaje de idiomas, así como en la elaboración de materiales.

En inglés existen aproximadamente 114.000 familias de palabras, excluyendo nombres propios (Goulden, Nation and Read, 1990). Esto implica un elevado número de palabras, imposibles de dominar tanto para un hablante nativo como para un aprendiz de la lengua.

Estos datos parecen indicar que los aprendices de inglés como lengua extranjera deberán aprender un elevado número de palabras. Sin embargo, no todas las palabras se utilizan con la misma frecuencia en un idioma. Existe un reducido número de palabras que debido a su elevada presencia en el idioma, cubre una gran parte del discurso oral y escrito en inglés. Por lo tanto, ese reducido grupo de palabras

frecuentes resultarían útiles de aprender para los alumnos puesto que les permite funcionar en el idioma (Nation, 2001). Diversos autores, además, relacionan la frecuencia con el aprendizaje, afirmando que las palabras más frecuentes en un idioma se aprenden antes que aquellas que no lo son (Hirsh y Nation, 1992; Canga Alonso, 2013; Ellis y Wulff, *en prensa*).

En la enseñanza, la relación entre frecuencia y aprendizaje implicaría la aplicación de un análisis de costo/beneficio a la hora de decidir si enseñar un ítem léxico o no (Nation, 2001; Schmitt y Schmitt, 2014). De acuerdo con el enfoque de costo/beneficio, profesores y creadores de materiales didácticos deberían analizar los distintos ítems léxicos con la finalidad de decidir si dedicar tiempo en el aula a enseñarlos o no. Si un ítem es muy frecuente, se le consideraría útil en el idioma y merecería dedicar tiempo en el aula a enseñarlo. Por el contrario, si un ítem es infrecuente en el idioma, se le consideraría poco útil y no merecería el dedicar tiempo en el aula a enseñarlo.

Wordlists

Con relación a la frecuencia con la que diferentes palabras aparecen en inglés, diversos autores han desarrollado a lo largo del tiempo diferentes listas de palabras de acuerdo con su frecuencia (West, 1953; Nation, 2004; Xue y Nation, 1984; Coxhead, 2000; Brezina y Gablasova, 2015). Estas listas pueden resultar de utilidad para los docentes y escritores de libros de texto a la hora de seleccionar el vocabulario al que los alumnos estarán expuestos en clase y en los libros de texto. La lista más utilizada de vocabulario frecuente es la desarrollada por Michael

West en 1953, *A General Service List of English Words*. Esta lista incluye aproximadamente las 2.000 familias de palabras más frecuentes en inglés. Nation (2001) la modificó, incluyendo ahora 1986 familias de palabras, para ser empleada en el programa RANGE, herramienta que se utilizará en el análisis llevado a cabo en el presente artículo.

Sin embargo, las listas de inglés general sirven de base para crear listas de palabras más especializadas, como la *Academic Wordlist*, desarrollada por Coxhead (2000) y que contiene 570 familias de palabras académicas.

Cobertura léxica

El concepto de cobertura léxica, o porcentaje de texto que el lector entiende, está fuertemente ligada a la comprensión del mismo. Por ello, dicho concepto ha sido ampliamente estudiado con la finalidad de establecer el umbral de cobertura léxica necesario para lograr una comprensión apropiada de un texto.

El vocabulario presente en un texto es uno de los factores esenciales que afectan a su comprensión (Terrazas Gallego y Agustín Llach, 2009; Hirsh y Nation, 1992), llegando Laufer (1996) a indicar que es el factor más interrelacionado con la comprensión. Diversos estudios han analizado la relación entre el conocimiento del vocabulario y el éxito en la comprensión lectora, teniendo como objetivo también encontrar el umbral recomendado de cobertura de texto para conseguir una adecuada comprensión lectora. La cobertura de texto se define como el porcentaje de palabras de un texto oral o escrito que un lector entiende (Laufer y Ravenhorst-Kalovski, 2010). Aunque

se han establecido diversos porcentajes a lo largo del tiempo, en el presente artículo tomaremos 95% de cobertura léxica como la apropiada para la comprensión de un texto, siguiendo a Laufer (1989), Liu y Nation (1985) y Alcaraz Mármol (2011).

Uno de los estudios más recientes es el que Nation llevó a cabo en 2006, quien desarrolló catorce listas de frecuencia basándose en el *British National Corpus*

(*BNC*) y analizó nueve corpus compuestos por diferentes textos orales y escritos, comprobando la cobertura que cada lista poseía en los textos. Los textos que Nation (2006) analizó pertenecían a novelas, periódicos, conversaciones diarias y personas participando en programas de radio (Schmitt y Schmitt, 2014). Es decir, principalmente inglés general. El patrón que encontró respecto a la cobertura es el siguiente:

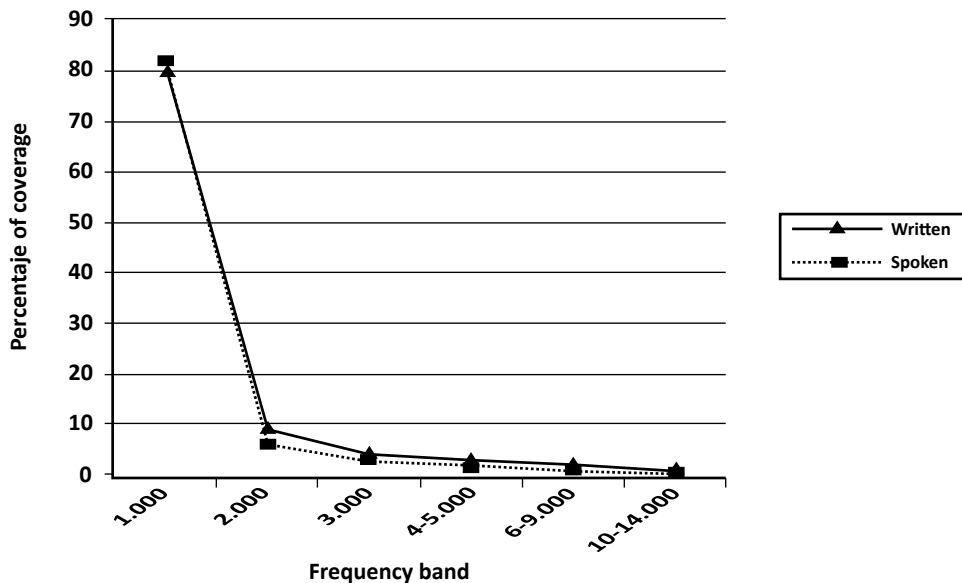


Figura 1.- Frecuencia y cobertura (Schmitt y Schmitt, 2014, p. 487)

En esta figura se puede observar el patrón de cobertura realizado por Schmitt y Schmitt (2014) a partir de los datos obtenidos por Nation (2006) en los textos analizados, conteniendo principalmente

inglés general. Como se puede observar en la figura, la cobertura de las bandas de frecuencia disminuye conforme el vocabulario se vuelve más infrecuente.

3. LIBROS DE TEXTO EN EL AULA

La importancia de los libros de texto en el aula de idiomas es destacada por Hutchinson y Torres (1994), quienes conciben el libro de texto como la guía que ayuda a los alumnos a organizar su aprendizaje tanto dentro como fuera del aula. Además, Häcker (2008) etiqueta los libros de texto como la principal fuente de léxico para alumnos que sólo utilizan el idioma dentro del aula y no leen libros fuera de la misma.

Sin embargo, el caso del inglés es particular, pues es sencillo encontrar fuentes de input en inglés fuera del aula, como películas y canciones, entre otros (Milton, 2009). Para aprendices de otros idiomas habría, no obstante, tres principales fuentes de input de dicho idioma: 1) el lenguaje del profesor, 2) los libros de texto y 3) el lenguaje usado en el aula por los alumnos. Aunque existe una mayor facilidad de encontrar fuentes de input en inglés fuera del aula que en otros idiomas, para muchos alumnos que aprenden inglés estas tres fuentes siguen siendo su único medio de contacto con el idioma (Alsaif and Milton, 2012). Por todo esto, no se debe ignorar la importancia de los libros de texto en el aula y su utilidad para alumnos así como para profesores (Alcaraz Mármol, 2011; Hutchinson and Torres, 1994).

Uno de los principales factores a tener en cuenta a la hora de desarrollar libros de texto es la selección del vocabulario que va a estar presente en los mismos. Tanto para O'Dell (1997) como para Milton (2009), la frecuencia debería ser el factor dominante a tener en cuenta para la selección de vocabulario.

Aunque todas estas consideraciones se hicieron con respecto a los libros utilizados en la enseñanza del inglés como lengua extranjera, serían, en cierto modo, extrapolables a los libros de texto utilizados en AICLE, puesto que emplean un idioma extranjero para enseñar asignaturas teóricas.

La pregunta de investigación principal a las que se pretende dar respuesta en el presente estudio es la siguiente:

1) ¿Cuál es la cobertura de cada banda de frecuencia en los diferentes textos? ¿Qué tamaño de vocabulario es necesario para alcanzar una adecuada comprensión del texto? (PI1)

A su vez, esta pregunta se desglosa en las siguientes:

1.1) ¿Cuál es la cobertura de cada banda de frecuencia en el corpus completo, conteniendo ambos libros de texto? ¿Qué tamaño de vocabulario es necesario para alcanzar una adecuada comprensión del texto? (PI1.1)

1.2) ¿Cuál es la cobertura de cada banda de frecuencia en el libro de texto de la asignatura “Natural Science”? ¿Qué tamaño de vocabulario es necesario para alcanzar una adecuada comprensión del texto? (PI1.2)

1.3) ¿Cuál es la cobertura de cada banda de frecuencia en el libro de texto de la asignatura “Social Science”? ¿Qué tamaño de vocabulario es necesario para alcanzar una adecuada comprensión del texto? (PI1.3)

- 1.4) ¿Cuáles son las similitudes y diferencias entre ambas asignaturas en lo que respecta al vocabulario? (PI1.4)

4. MÉTODO

Libros de texto

Para el presente estudio se analizaron dos libros de texto pertenecientes a la misma editorial y dirigidos al mismo público, aunque de asignaturas diferentes. El curso al que ambos libros están dirigidos es quinto curso de Educación Primaria y están destinados a la enseñanza bilingüe de las asignaturas de Ciencias Naturales y Ciencias Sociales. Los libros de texto analizados son: *Natural Science ByME: Primary 5* (2014) y *Social Science ByME: Primary 5* (2014).

La elección de los libros de texto estuvo influida por la implantación de la LOMCE (Ley Orgánica de Mejora de la Calidad Educativa) durante el curso académico 2014/2015. Durante dicho curso académico la LOE (Ley Orgánica de Educación), la ley anterior, y la LOMCE han coexistido. Los cursos impares se regían por la LOMCE mientras que los cursos pares seguían la LOE. Uno de los principales cambios implantados por la LOMCE es la división de la asignatura “Conocimiento del Medio Natural, Social y Cultural” en dos asignaturas: “Ciencias Naturales” y “Ciencias Sociales”. En el caso de los alumnos pertenecientes a secciones o colegios bilingües, la asignatura “Natural and Social Science” se dividió

en “Natural Science” y “Social Science”. Debido a los cambios en la legislación educativa que han tenido lugar en España, las distintas editoriales debieron publicar nuevos libros de texto. Entre los ejemplares más utilizados en las asignaturas bilingües de distintos centros educativos, encontramos los dos libros utilizados para nuestro análisis. Además, probablemente serán los libros utilizados en, al menos, los próximos cuatro años.

Elaboración del corpus lingüístico

Para el presente estudio, se compiló un corpus conteniendo los dos libros de texto objeto de nuestro análisis. En el corpus se incluyeron todas las palabras presentes en ambos libros de texto. El corpus se compiló y transcribió en un procesador de texto (bloc de notas) y guardado en formato .txt. Posteriormente, cada lección fue guardada de forma individual en un documento. El corpus creado contiene en total 57.142 tokens o palabras. En este punto es importante diferenciar los conceptos tokens y types. Los tokens hacen referencia a las palabras presentes en un texto, incluidas las repeticiones de las mismas, mientras que los types se refieren a las palabras diferentes en un texto. Así, en el ejemplo “el niño escribe en el libro”, podemos encontrar seis tokens y cinco types, puesto que la palabra “el” aparece dos veces. En la siguiente tabla se puede observar el número de tokens, types y la type/token ratio del corpus completo, cada libro de texto por separado y los trimestres de cada libro de texto.

Tabla 1: Tokens y types en el corpus.

Parte del corpus	Tokens	Types	Type/Token ratio
Corpus completo	57.142	5.723	0,10
Natural Science	27.945	3.486	0,12
Social Science	29.197	3.691	0,13
Primer Trimestre	9.971	1.813	0,18
Natural Science			
Segundo Trimestre	7.978	1.476	0,18
Natural Science			
Tercer Trimestre	9.996	1.778	0,18
Natural Science			
Primer Trimestre	9.809	1.904	0,19
Social Science			
Segundo Trimestre	8.967	1.554	0,17
Social Science			
Tercer Trimestre	10.421	1.826	0,17
Social Science			

Herramientas

La herramienta utilizada para analizar el texto ha sido el programa RANGE, desarrollado por Paul Nation en 1994. Este programa clasifica las palabras de un texto en distintas listas de frecuencia. Además, proporciona la cobertura de cada lista de frecuencia en un texto, objeto principal de nuestro estudio. En el programa distintas listas están incluidas y pueden ser utilizadas, o bien se pueden emplear listas desarrolladas por cada usuario. Una vez elegido el texto y las listas que van a ser utilizadas en el análisis, el programa analiza el texto y crea un documento en

el que se indica la cobertura de texto de cada lista, las palabras pertenecientes a la lista que se encuentran en el texto y la frecuencia con la que aparece cada palabra. Las palabras del texto que no pertenecen a ninguna lista aparecen como “no encontradas” (“not found”).

La versión de RANGE utilizada para nuestro estudio fue la que contiene las listas basadas en los corpus de BNC (*British National Corpus*) y COCA (*Corpus of Contemporary American English*). Esta versión contiene treinta y cuatro listas cuyo contenido se puede encontrar en la información del programa. Para nuestro

análisis empleamos las veinticinco primeras listas puesto que las listas veintiséis y siguientes son aquellas que el programa proporciona para que los usuarios las modifiquen. Así, esas listas contienen caracteres sueltos con la finalidad de que los usuarios creen sus propias listas si así lo consideran.

Listas

A la hora de realizar el análisis de los libros de texto decidimos que la unidad de análisis fuesen palabras y no familia de palabras. Los alumnos a los que están dirigidos ambos libros de texto son aprendices de inglés como lengua extranjera y no están familiarizados con el uso de afixos para la formación de palabras en dicho idioma. Por tanto, decidimos analizar las palabras individualmente y la frecuencia con la que cada una de ellas aparecía en el texto.

La segunda decisión que se tuvo que tomar fue si utilizar las listas de palabras proporcionadas en el programa o no. Decidimos utilizar las veinticinco primeras listas, con las 25.000 palabras más frecuentes, y todas aquellas palabras presentes en el texto y no incluidas en ninguna de las veinticinco listas fueron catalogadas como “offlist words” (palabras fuera de lista). Entre estas “offlist words” se podían encontrar principalmente nombres propios, palabras en español, letras del alfabeto, palabras compuestas y palabras no incluidas en ninguna de las veinticinco listas. Sin embargo, consideramos que la recategorización de algunas de estas palabras sería conveniente. Las decisiones que tomamos en lo que respecta a la recategorización fueron las siguientes:

- Incluir las palabras compuestas en las listas. Algunas palabras compuestas como “notebook” aparecían frecuentemente en ambos libros de texto, por lo que consideramos apropiado incluir las palabras compuestas en las listas. Cada palabra fue analizada individualmente para decidir en qué lista incluirla.

- Las palabras separadas por un guión se consideraron palabras compuestas y fueron analizadas de igual modo a la hora de incluirlas en las listas.

- Incluir las letras del alfabeto en la primera lista, dentro de las 1.000 palabras más frecuentes. Esta decisión vino determinada por distintas razones. La primera de ellas es que las letras son muy frecuentes en ambos libros de texto, particularmente, alguna de ellas. Además, a la hora de analizar la palabra “a”, el programa RANGE no hace diferenciación en lo que respecta a las veces que la palabra aparece como letra del alfabeto o como palabra con significado léxico. Finalmente, ambos libros de texto están destinados a alumnos españoles aprendices de inglés como lengua extranjera. Ambos idiomas comparten el alfabeto, por lo que los alumnos están familiarizados con las letras del alfabeto.

- Separar las palabras en español de las palabras “offlist”. Así, creamos otro grupo de palabras que contenía todas las palabras en español presentes en el texto.

Por tanto, tras la recategorización llevada a cabo, en la lista de las “offlist words” se incluían nombres propios y palabras no incluidas en ninguna de las veinticinco listas. Además, existe otra lista presente llamada “spanish words”

conteniendo todas las palabras en español que aparecen en los textos. En lo que respecta a los nombres propios, diversos autores, entre los que se encuentra Nation (2001), deciden omitir los nombres propios de los textos que analizan (principalmente libros de lectura), argumentando que los lectores van conociendo los nombres conforme se va avanzando en el texto. Así, los nombres de los personajes o lugares en los que una historia tiene lugar van siendo conocidos conforme el lector avanza en la historia. Sin embargo, los textos que nosotros analizamos no son libros de lectura, sino textos académicos, y como tales, los lectores no se van familiarizando con los nombres propios progresivamente, sino que éstos son parte del contenido y, por consiguiente, necesarios para la comprensión del texto.

5. RESULTADOS

En esta sección analizaremos los resultados obtenidos en nuestro estudio desde una perspectiva de lingüística de corpus. Los resultados del corpus completo serán analizados, así como los de cada libro de texto por separado.

El gráfico de distribución de Nation

(2006) (Figura 1) será tomado como referencia para el análisis de nuestros resultados, comprobando las similitudes y diferencias entre su patrón de cobertura y los que se pueden desarrollar a partir de nuestro corpus.

Corpus completo (PII.1)

La siguiente tabla (tabla 2) muestra los resultados de cobertura de las distintas listas de frecuencia en nuestro corpus. Siguiendo a Schmitt y Schmitt (2014), las bandas de frecuencia comprendidas entre la cuarta y la octava, ambas incluidas, se agruparon en un conjunto llamado “mid-frequency” (palabras de media frecuencia). De igual modo, las bandas de frecuencia comprendidas entre la novena y la vigésimo quinta, ambas incluidas, fueron agrupadas en un conjunto llamado “low-frequency” (palabras de baja frecuencia). Además, las palabras que aparecen como ‘OFF’ son aquellas pertenecientes a la lista de palabras “offlist” (fuera de lista) creadas con las palabras que no aparecían en ninguna de las listas tras la recategorización llevada a cabo. Finalmente, las palabras que aparecen como “SP” son las palabras en español que aparecen en el texto.

Tabla 2: Cobertura de cada lista en el corpus completo

Listas de frecuencia	Cobertura de cada lista	Porcentaje acumulativo
1	70,47%	70,47%
2	10,78%	81,25%
3	6,11%	87,36%
Mid-freq	5,78%	93,14%

Low-freq	1,28%	94,42%
OFF	3,35%	97,77%
SP	2,23%	100%

Como se puede apreciar en la tabla 2, la cobertura de las listas disminuye con la frecuencia. La cobertura de la primera lista (1.000 palabras más frecuentes) es mucho más elevada que la cobertura de cualquiera de las otras listas.

Tomando una cobertura del texto del 95% como la apropiada para comprender un texto adecuadamente (Laufer, 1989; Liu and Nation, 1985), algún conocimiento de las ocho primeras bandas de frecuencia (y las palabras en español) sería necesario para alcanzar aproximadamente dicha cobertura. Las palabras en español cubrirían aproximadamente un 2% del texto. Tomaremos las palabras en español como palabras que facilitan la

comprensión del texto, o al menos no impiden la comprensión del mismo, puesto que se encuentran en la lengua materna de los alumnos. Finalmente, en lo que respecta a las palabras catalogadas como “fuera de lista”, cubrirían alrededor de un 3% del corpus.

En la siguiente figura (figura 2), se muestra el patrón de cobertura de cada lista de palabras en nuestro corpus, superpuesto al patrón encontrado por Nation (2006). Puesto que Nation otorgaba rangos a la cobertura de cada lista, escogimos los valores más bajos de los rangos y la diferencia fue asignada a las palabras “fuera de lista”.

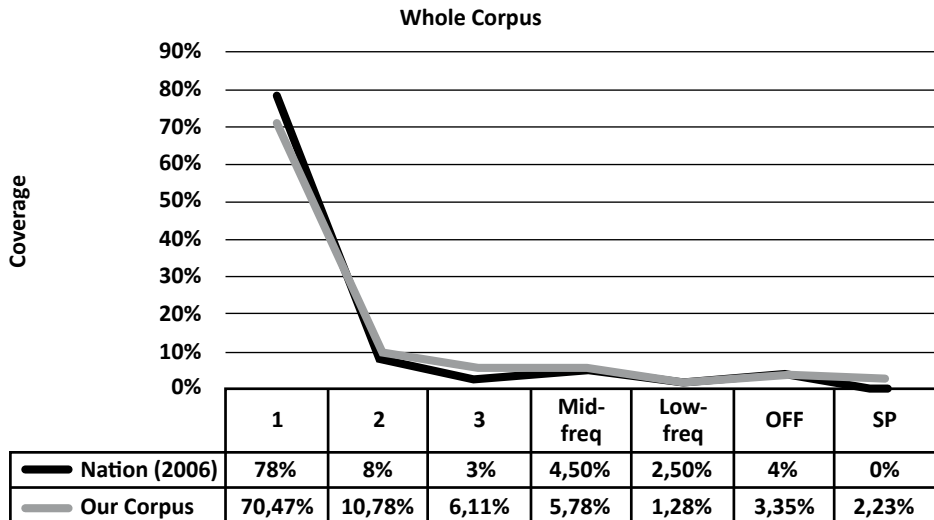


Figura 2.- Gráfico de cobertura corpus completo

Tal y como se puede observar en el gráfico, el patrón descendente en nuestro corpus es similar al encontrado por Nation (2006). La cobertura de la primera lista es mayor en los corpus analizados por Nation que en nuestro corpus. Dicha diferencia puede ser ocasionada por el análisis de textos relacionados con inglés general que realizó Nation mientras que

nuestro estudio se ha realizado sobre textos académicos.

Resultado de la asignatura “Natural Science” (PII.2)

La siguiente tabla (tabla 3) muestra la cobertura de cada banda de frecuencia en el libro de texto de “Natural Science”.

Tabla 3: Cobertura de cada lista en “Natural Science”

Listas de frecuencia	Cobertura de cada lista	Porcentaje acumulativo
1	73,46%	73,46%
2	10,96%	84,42%
3	6,13%	90,55%
Mid-freq	6,71%	97,26%
Low-freq	1,84%	99,10%
OFF	0,80%	99,90%
SP	0,10%	100%

En esta tabla se puede observar, como en el caso anterior, que la cobertura disminuye conforme el vocabulario se vuelve infrecuente. En el libro de texto de “Natural Science” estas palabras cubren aproximadamente un 1% del texto, lo que implica una de cada cien palabras. Por otra parte, la cobertura de la primera lista de frecuencia es mayor que en el corpus completo, aproximadamente un 3% más. Además, las listas pertenecientes a los grupos de media y baja frecuencia también cubren más texto que en el corpus completo. Finalmente, es destacable la diferencia entre la cobertura en el caso

del corpus completo y “Natural Science”, puesto que en este último caso dichas palabras cubren solo un 0,10% del texto.

En este libro de texto, con las ocho primeras bandas de frecuencia se alcanzaría una cobertura del 97% del texto. Tomando 95% como la cobertura apropiada del texto, ésta sería alcanzada aproximadamente con las seis primeras listas de palabras y las palabras en español.

En el siguiente gráfico (figura 3) se puede apreciar una comparación entre la cobertura en nuestro corpus y la encontrada por Nation (2006).

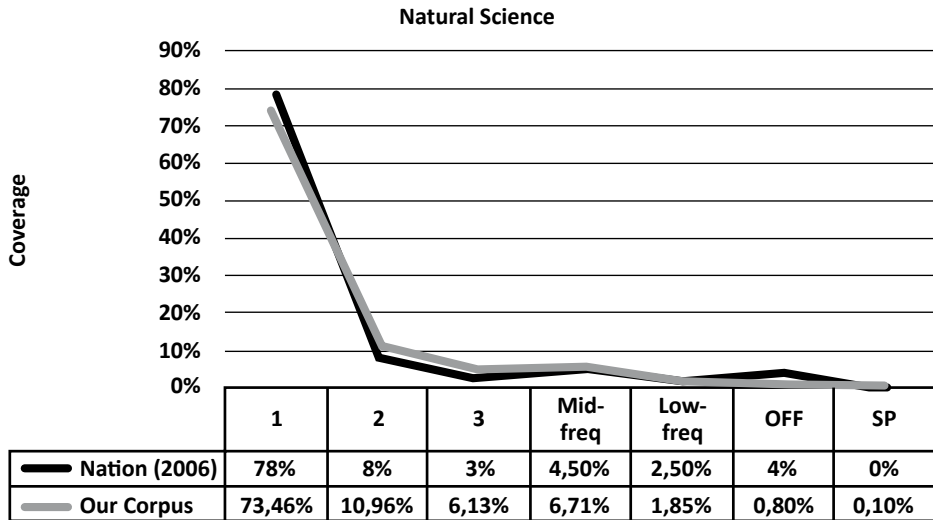


Figura 3.- Gráfico de la cobertura en “Natural Science”

En este gráfico se puede observar un patrón decreciente que continúa decreciendo cuando llega a las palabras fuera de lista, al contrario de lo que pasaba en el corpus completo. Además, la diferencia entre la cobertura de la primera lista en los corpus de Nation y nuestro corpus es menor que en el caso anterior.

Resultado de la asignatura “Social Science” (PII.3)

La siguiente tabla (tabla 4) muestra la cobertura de cada banda de frecuencia en el libro de texto de “Social Science”.

Tabla 4: Cobertura de cada lista en “Social Science”

Listas de frecuencia	Cobertura de cada lista	Porcentaje acumulativo
1	67,62%	67,62%
2	10,61%	78,23%
3	6,10%	84,33%
Mid-freq	4,88%	89,21%
Low-freq	0,73%	89,94%
OFF	5,82%	95,75%
SP	4,25%	100%

Como se puede observar en la tabla 3, en el libro de texto de la asignatura “Social Science” aumenta la cobertura léxica de las palabras fuera de lista y en español.

Centrándonos en la cobertura de la primera lista (1.000 palabras más frecuentes), encontramos que su cobertura disminuye con relación a “Natural Science” y a la cobertura de la misma en el corpus completo. En lo que respecta al vo-

cabulario necesario para alcanzar un 95% de cobertura del texto, se puede observar que sería necesario algún conocimiento de las veinticinco listas de frecuencia y las palabras en español.

Comparando los resultados obtenidos en nuestro libro de texto con los resultados obtenidos por Nation (2006) se puede elaborar el siguiente gráfico.

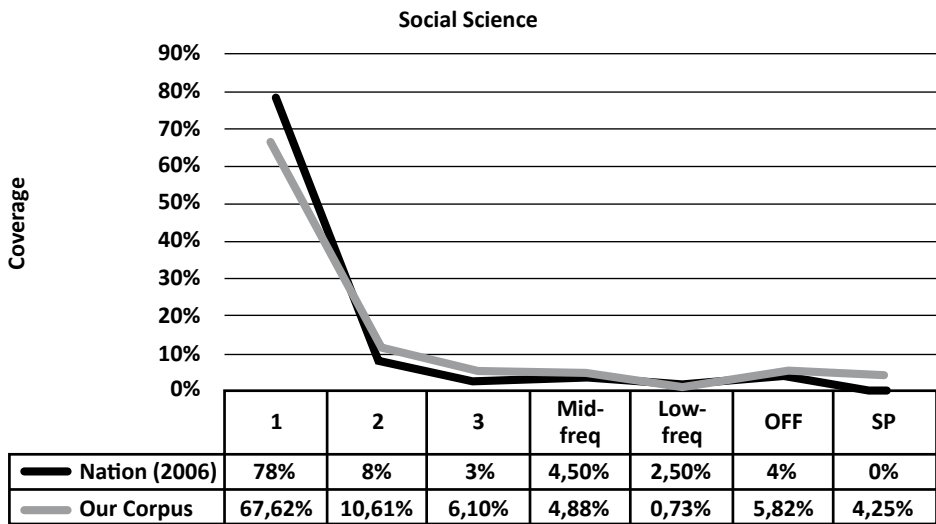


Figura 4.- Gráfico de la cobertura en “Social Science”

En el gráfico se puede observar que el patrón decreciente es similar al de Nation y al encontrado tanto en el corpus completo como en el sub-corpus de “Natural Science”.

Comparación “Natural Science” y “Social Science” (PII.4)

En el siguiente gráfico (figura 5) se muestra una comparación entre la cobertura de las distintas bandas de frecuencia en cada asignatura.

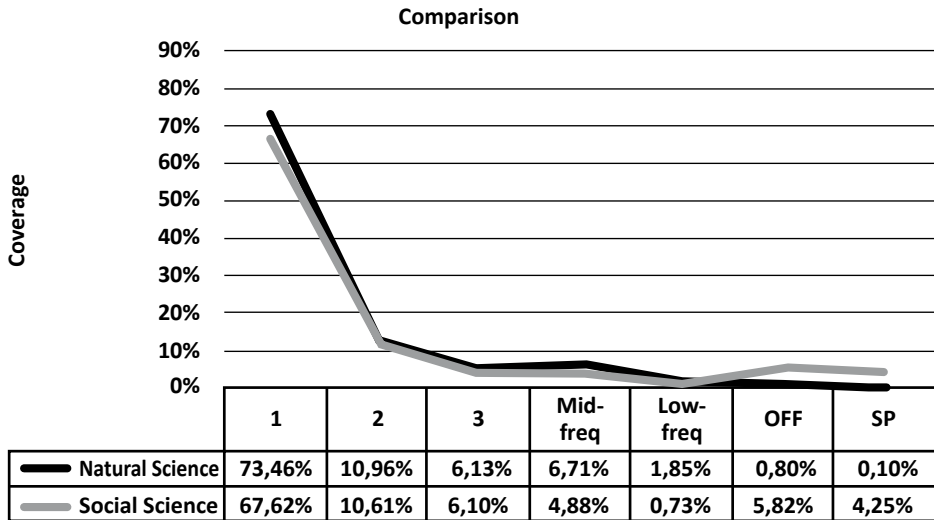


Figura 5.- Gráfico comparación “Natural Science” y “Social Science”

Como se puede observar en la figura 5, la principal diferencia entre ambos libros de texto reside en la cobertura de la primera lista (1.000 palabras más frecuentes) y la presencia de palabras fuera de lista y en español en el texto. La cobertura de la primera lista es aproximadamente un 6% mayor en “Natural Science” que en “Social Science” mientras que las palabras etiquetadas como “offlist” (principalmente nombres propios) cubren seis veces más texto en “Social Science” (5,82%) que en “Natural Science” (0,80%). Otra diferencia principal es la cantidad de vocabulario necesaria para alcanzar una cobertura del 95% del texto. Mientras que en la asignatura de “Natural Science” ese porcentaje sería alcanzado con aproximadamente las seis primeras bandas de frecuencia y las palabras en español (6.000 palabras más frecuentes), en la asignatura de “Social Science” las veinticinco bandas de fre-

cuencia y las palabras en español serían necesarias. En lo que respecta a la cobertura de las palabras pertenecientes a las listas de media frecuencia y baja frecuencia es mayor en “Natural Science” que en “Social Science” (aproximadamente un 3% mayor en total).

6. CONCLUSIONES

Tal y como se puede observar en los resultados obtenidos en nuestro estudio, la cobertura léxica disminuye conforme el vocabulario se vuelve infrecuente, así, la primera lista de frecuencia cubre mucho más texto que cualquiera de las otras listas. Por lo tanto, un reducido número de palabras cubren un gran porcentaje de los *tokens* presentes en nuestro corpus (aproximadamente siete de cada diez *tokens*), tal y como ocurre en el discurso general (Nation, 2001; Nation, 2006).

En lo que respecta a los resultados del corpus completo, nos parece necesario señalar que, aunque la máxima cobertura posible con las veinticinco bandas de frecuencia y las palabras en español es aproximadamente un 97%, un 94,5% de cobertura de texto se alcanzaría con las seis primeras bandas de frecuencia y las palabras en español. Por ello, hemos considerado las seis primeras bandas de frecuencia (6.000 palabras más frecuentes) como las palabras clave para entender los dos libros de texto.

Centrándonos ahora en el libro de texto de “Natural Science”, es notable el cambio en la cobertura de las palabras fuera de lista con respecto al corpus completo (cubriendo aproximadamente un 1% en “Natural Science” y un 4% en el corpus completo). Esta diferencia en la cobertura léxica se extiende a las palabras en español presentes (0,1% de cobertura léxica en “Natural Science” y 2,2% en el corpus completo). Sin embargo, en “Natural Science” la cobertura de la primera lista de frecuencia es mayor que en el corpus completo, lo que indica un menor porcentaje de nombres propios y palabras en español y una mayor presencia de las 1.000 palabras más frecuentes. En lo que respecta a las palabras de media y baja frecuencia, su cobertura es mayor en la asignatura de “Natural Science” que en el corpus completo y puesto que el vocabulario técnico suele estar incluido en las listas de palabras de media y baja frecuencia, se podría deducir que el libro de texto de “Natural Science” tiene más proporción de palabras técnicas. Finalmente, en este libro de texto se alcanzaría el 95% de cobertura léxica con las seis primeras listas y las palabras en español, lo que

coincide con nuestra consideración de las 6.000 palabras más frecuentes como las palabras clave para entender ambos libros.

En lo que respecta a la asignatura “Social Science”, en este libro de texto la cobertura de las palabras fuera de lista aumenta hasta alcanzar casi el doble de su cobertura en el corpus completo. La cobertura de la misma lista en la asignatura de “Natural Science” es 0,80%, por lo que en “Social Science” las palabras etiquetadas como “offlist” son seis veces más frecuentes que en “Natural Science”. Esta amplia diferencia puede deberse a la alta presencia de nombres propios (pertenecientes a la lista de palabras “offlist”) en las lecciones dedicadas a geografía e historia, contenidas en el libro de “Social Science”. En lo que respecta al porcentaje de palabras en español, este asciende aproximadamente a un 4%, lo que indica un elevado número de palabras en español, así como de nombres propios en dicho idioma. Es destacable el hecho de que la cobertura de las palabras fuera de lista en “Social Science” es aproximadamente igual que la de la tercera banda de frecuencia (3.000 palabras más frecuentes). Centrándonos en la primera banda de frecuencia y su cobertura en “Social Science”, encontramos que hay una diferencia (la cobertura es aproximadamente un 6% menor) que influirá en el vocabulario necesario para que los alumnos entiendan el libro de texto. Estos resultados sugieren que “Social Science” probablemente sea más exigente y difícil en lo que respecta a vocabulario para los alumnos.

En la comparación realizada entre los libros de texto de las asignaturas “Natural Science” y “Social Science”, encon-

tramos que las principales diferencias entre unos y otros es la cobertura de la primera lista, de las palabras fuera de lista y de las palabras en español. Puesto que en la lista de palabras fuera de lista y palabras en español encontramos principalmente nombres propios, podemos deducir que en “Social Science” encontramos un número elevado de nombres propios, seguramente debido a la presencia de unidades dedicadas a historia y geografía. Estas palabras requieren que los alumnos se familiaricen con ellas y las aprendan, en muchas ocasiones son presentadas de forma aislada (por ejemplo los nombres de países en los mapas). Además, a la hora de lograr un 95% de cobertura léxica, encontramos también una notable diferencia puesto que en “Natural Science” esa cobertura se lograría con las seis primeras listas de palabras y las palabras en español mientras que en “Social Science” sería necesario algún conocimiento de las veinticinco listas además de las palabras en español. Estas diferencias son relevantes y afectarán a la comprensión del texto por parte de los alumnos. Por otra parte, en la asignatura de “Natural Science” la cobertura de las palabras de media y baja frecuencia es mayor que en “Social Science”. Puesto que el vocabulario técnico se incluye en dichas listas, se puede deducir que la asignatura de “Natural Science” contiene más vocabulario técnico que la asignatura de “Social Science”. Finalmente, es interesante destacar la diferencia entre la cobertura de palabras en español, siendo mucho mayor en “Social Science” que en “Natural Science”. Sin embargo, siendo en su mayoría nombres propios las palabras en español, no es fácilmente dedu-

cible hasta qué punto pueden facilitar la comprensión o no.

Se suele relacionar la dificultad de un texto con la presencia de vocabulario técnico, pero en este caso particular los profesores conciben “Social Science” como la más difícil de las dos para enseñarla en inglés, mientras que es, por el contrario, la que menos vocabulario técnico contiene. Dicho ejemplo puede implicar la necesidad de tener en cuenta otros aspectos relacionados con la dificultad de una asignatura, no exclusivamente la cantidad de vocabulario técnico presente en la misma.

La diferencia en el vocabulario necesario para alcanzar un 95% de cobertura léxica estando ambos libros de texto destinados a los mismos alumnos, es bastante relevante e influye en la cantidad de vocabulario necesaria para entender cada libro de texto de una forma adecuada y, por lo tanto, en su complejidad.

Con todos los datos obtenidos en nuestro estudio se puede afirmar que los libros de texto de AICLE son exigentes para los alumnos. La distribución de la cobertura de cada banda de frecuencia en el texto es similar a la encontrada por Nation (2006) (figura 1). Sin embargo, es necesario tener en cuenta el hecho de que Nation analizó distintos corpus elaborados por y para hablantes nativos del idioma. Por lo tanto, ambos libros de texto siguen la distribución de textos hechos por y para hablantes nativos, y nosotros consideramos la distribución demasiado normal para tratarse de libros destinados a alumnos de diez u once años aprendices de inglés como lengua extranjera. De nuestros datos se podría deducir que los creadores de los libros de texto no han hecho un esfuerzo por adaptar el vocabu-

lario a las características de los alumnos, y creemos que incluir un vocabulario menos exigente sería aconsejable.

Finalmente, de nuestros resultados pueden deducirse algunas implicaciones metodológicas. A pesar de la complejidad de los libros de texto, los alumnos consiguen los objetivos tanto de contenido como lingüísticos, por lo que no se debe ignorar la importancia del papel del maestro en el aula. En el presente estudio se han analizado los libros de texto de forma aislada, pero se debe tener en cuenta que estos libros de texto son explicados a los alumnos por maestros, quienes a través de diferentes técnicas y métodos consiguen que los alumnos entiendan el contenido y el lenguaje. Sin embargo, las diferencias entre el libro de “Natural Science” y el de “Social Science”, dirigidos a los mismos alumnos, implicarán una carga extra para los docentes a la hora de trabajar el vocabulario y el contenido.

Teniendo en cuenta los resultados de la cobertura léxica de cada lista, sería aconsejable dedicar tiempo en el aula a enseñar las palabras más frecuentes en los libros de texto, puesto que son las que más presencia tienen en el texto. Nosotros hemos considerado las seis primeras listas como las clave para entender ambos textos, por lo que los alumnos deberían estar familiarizados con ellas para lograr una apropiada comprensión del texto. Además, la cooperación entre el maestro de AICLE y el especialista de inglés sería beneficiosa para ambas partes puesto que se podría trabajar en la asignatura de inglés el vocabulario y estructuras necesarios para el desarrollo de las unidades en las asignaturas AICLE.

7. LIMITACIONES Y POSIBLES FUTURAS LÍNEAS DE INVESTIGACIÓN

Nuestro estudio presenta ciertas limitaciones, siendo la más evidente y relevante el tamaño del corpus elaborado. Nuestro corpus contiene dos libros de texto y consideramos conveniente el tener una muestra mayor a la hora de obtener resultados en términos más generales. Al tratarse de libros de texto, nos encontramos con la limitación de no saber durante cuánto tiempo los libros de texto analizados se ajustarán a la ley educativa y se encontrarán en uso en los centros educativos.

En lo que respecta a las posibles futuras líneas de investigación, y centrándonos en el corpus elaborado por nosotros, no solo el vocabulario puede analizarse, sino también diversas estructuras gramaticales presentes en los libros de texto y su situación en el mismo (encabezados, actividades, contenido, etc.).

Asimismo, sería conveniente analizar una mayor variedad de libros de texto destinados a un alumnado similar con la finalidad de observar las similitudes y diferencias del vocabulario presentado así como establecer conclusiones más generales. De igual modo, consideramos una interesante línea de investigación el analizar la evolución del vocabulario presente en los libros de texto a lo largo de los distintos cursos de Educación Primaria. Finalmente, sería recomendable analizar la relación entre el vocabulario al que los alumnos están expuestos tanto en los libros de texto como en el aula y el vocabulario que los alumnos adquieren.

REFERENCIAS

- ALCARAZ MÁRMOL, G. (2011). Vocabulary input in EFL textbooks: Frequency levels. *Revista Española de Lingüística Aplicada*, 24, 9-28.
- ALSAIF, A. Y MILTON, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal*, 40(1), 21-33.
- BREZINA, V., Y GABLASOVA, D. (2015). Is there a core general vocabulary? Introducing the new General Service List. *Applied Linguistics*, 36(1), 1-22.
- CANGA ALONSO, A. (2013). The receptive vocabulary of Spanish 6th- grade primary-school students in CLIL instruction: A preliminary study. *Latin American Journal of Content and Language Integrated Learning*, 6(2), 22-41.
- COXHEAD, A. (2000). A new Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- ELLIS, N. Y WULFF, S. To appear in *Theories in Second Language Acquisition: An Introduction (2nd edition)*. In B. VanPatten and J. Williams.
- Goulden, R., Nation, P., y Read, J. (1990). How Large Can a Receptive Vocabulary Be? *Applied Linguistics*, 11, 341-363.
- HÄCKER, M. (2008). Eleven pets and 20 ways to express one's opinion: The vocabulary learner of German acquire at English secondary schools. *The Language Learning Journal*, 36(2), 215-226.
- HERAS, A., & LASAGABASTER, D. (2014). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70-88.
- HIRSH, D., Y NATION, P. (1992). What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- HUTCHINSON, T., Y TORRES, E. (1994). The textbook as agent of change. *ELT Journal*, 48(4), 315-328.
- LAUFER, B. (1989). What Percentage of Text-lexis is Essential for Comprehension? In C. Lauren, & M. Nordman (Eds.), *Special language: From Humans Thinking to Thinking Machines*. (pp. 316-324). Clevedon: Multilingual Matters.
- LAUFER, B. (1996). The Lexical Threshold of Second Language Reading Comprehension. In K. Sajavaara, & C. Fairweather (Eds.), *Approaches to Second Language Acquisition*. (pp. 55-62). Jyväskylä: University Printing House.
- LAUFER, B., Y RAVENHORST-KALOVSKI, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, 2(1), 15-30.
- LIU, N., Y NATION, I. S. P. (1985). Factors Affecting Guessing Vocabulary in Context. *RELC Journal*, 16, 33-42.
- MILTON, J. (2009). Vocabulary Acquisition and Classroom Input. In D. Singleton (Ed.), *Measuring Second Language Vocabulary Acquisition*. (pp. 193-217). Dublin: Multilingual Matters.
- NATION, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

- NATION, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a Second Language* (pp.3-14). Amsterdam: John Benjamins.
- NATION, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59-82.
- Natural Science ByME: Primary 5 (2014). Zaragoza: Macmillan and Edelvives.
- O'DELL, F. (1997). Incorporating vocabulary into the syllabus. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 258-278). Cambridge: Cambridge University Press.
- PÉREZ-CAÑADO, M. L. (2012). CLIL research in Europe: past, present, and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315-341.
- SCHMITT, N., Y SCHMITT, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503.
- Social Science ByME: Primary 5 (2014). Zaragoza: Macmillan and Edelvives.
- TERRAZAS GALLEGO, M., Y AGUSTÍN LLACH, M. P. (2009). Exploring the Increase of Receptive Vocabulary Knowledge in the Foreign Language: A Longitudinal Study. *International Journal of English Studies*, 9(1), 113-133.
- WEST, M. (1953). *A General Service List of English Words*. London: Longman, Green & Co.
- XUE, G., Y NATION, I. S. P. (1984). A university word list. *Language learning and communication*, 3(2), 215-229.

