

## ESTUDIOS SOBRE LA LENGUA DE INSTRUCCIÓN (LI): UN ÍNDICE DE FRECUENCIAS LÉXICAS LEMATIZADO A PARTIR DE UN MANUAL ESCOLAR\*

M<sup>a</sup> VICTORIA LÓPEZ PÉREZ\*\*  
*Universidad Pública de Navarra*

**ABSTRACT.** *The teaching of languages, in order to be successful, requires access to up-to-date samples of the different registers in which they are employed. In the case of Spanish as a second language, recently introduced formally in Spain, the linguistic and discursive synchronic description of the register of language used for instruction has yet to be carried out. This paper presents a word frequency list with 5.496 word families gathered using a school textbook as the linguistic base. This study initiates new research into the language of instruction vocabulary and seeks to contribute to a better understanding of this register in order to aid in the teaching of Spanish as a Second Language.*

**KEYWORDS:** *Spanish as a second language, language for instruction, vocabulary studies, word frequency list, word families, school textbook, Social Studies.*

**RESUMEN.** *Las enseñanzas de las lenguas necesitan contar con muestras de los diferentes registros en los que se actualizan y sobre los que incide su acción didáctica. En el caso del español como lengua segunda, docencia de reciente implantación en España, queda pendiente la descripción lingüística y textual, desde el punto de vista sincrónico, de una de las variedades funcionales en las que centra: la lengua de instrucción (LI). Este trabajo explora una vía de análisis del léxico de esta variedad llevando a cabo una ordenación de voces según criterios estadísticos de un manual escolar con ayuda de una herramienta informática, cuyo resultado es un índice que incluye 5.496 lemas. Nuestro estudio pretende servir de modelo a otros conductentes a un mejor conocimiento del léxico de la LI con miras a su tratamiento en aulas de E/L2.*

**PALABRAS CLAVE:** *Español como lengua segunda, lengua de instrucción (LI), vocabulario, estadística léxica, lematización, manual escolar, Ciencias Sociales.*

### 1. OBJETIVO DE ESTE TRABAJO

La reciente implantación del español como lengua segunda en nuestro sistema educativo pone de actualidad la *lengua de instrucción (LI)* en el ámbito investigador español.

Esta variedad funcional que identificamos con los usos de la lengua en situaciones de enseñanza-aprendizaje de los contenidos curriculares constituye uno de los principales objetivos del aprendizaje escolar: su consecución encamina a los estudiantes hacia el éxito académico<sup>1</sup>. El interés por ella en este trabajo tiene que ver con el afán de recabar información desde el punto de vista lingüístico sobre su léxico para fundamentar propuestas didácticas destinadas a alumnos en proceso de aprendizaje de la lengua española ubicados tanto en aulas específicas de E/L2 como en aulas ordinarias a las que estos se integran a corto plazo, ambas en centros donde la lengua de escolarización es el español.

La primera cuestión con la que nos enfrentamos a la hora de describir el léxico de la LI es cómo abordar el ingente número de voces que engloba teniendo en cuenta que incluye la lengua de las diferentes asignaturas del currículo escolar, en sus dos modalidades oral y escrita, y en los diferentes niveles escolares. Una forma de priorizar unas voces sobre otras con fines didácticos es fijarse en la representación que estas adquieren en un determinado corpus léxico, conocimiento al que se llega a partir de estudios estadísticos que revelan precisamente un aspecto de su uso: la frecuencia.

Este trabajo se plantea como objetivo acometer el ordenamiento de una parcela léxica de la LI según criterios de frecuencia valiéndonos de una herramienta informática que facilite el recuento. El corpus de voces elegido está formado por aquellas contenidas en un libro de texto en vigor correspondiente a una materia troncal de tercer curso de la E.S.O. que, por formar un todo en sí mismo y por su amplitud, lo consideramos suficientemente representativo para nuestro empeño. El resultado de tal ordenamiento es un *índice de frecuencias léxicas*, que, además de revelarnos información cuantitativa sobre las voces, muestra otros aspectos de estas –principalmente morfo-sintácticos–, debido al proceso de lematización al que las hemos sometido, esta vez de forma manual.

Además de conocer la representatividad que los vocablos adquieren en la muestra considerada con miras a su tratamiento didáctico, nuestro estudio pretende servir de modelo para elaborar otros listados a partir del resto de los manuales de las asignaturas curriculares que completen la configuración léxica de la LI en la actualidad, en sus diferentes cursos y asignaturas.

## 2. LÍMITES DEL TRABAJO

Nuestro estudio se limita a ofrecer un producto lexicográfico representativo de la variedad de lengua estudiada por el número de voces que maneja y por la coherencia de su factura. No entra, por tanto, en la interpretación teórica y aplicada de sus datos, tarea que queda pendiente. Con ello, salimos al paso de las críticas que surgen de la utilización exclusiva de la frecuencia como criterio de selección y de las deficiencias que presenta la aplicación directa de los índices de frecuencia en las aulas de lenguas maternas o segundas. Cervero y Picardo (2003: 39) dan cuenta de las principales carencias de las listas de voces ordenadas por su frecuencia: “no se ofrece información sobre el valor discursivo, pragmático o semántico de las unidades léxicas. No se indica el registro, ni las

acepciones, ni los co-textos y contextos de posible aparición, etc.”. Además de la falta de información sobre aspectos relacionados con el significado, los listados de palabras presentan problemas relacionados también con la forma de las voces. Entre ellos, según Izquierdo Gil (2002: 206), está el que se deriva de una incorrecta definición del concepto de unidad léxica. En la mayoría de los índices, también en el nuestro, las entradas de las voces corresponden a unidades monoverbales porque se parte de un concepto ortográfico más que semántico de unidad léxica, lo que provoca la ausencia de unidades pluriverbales o fraseológicas con un único significado y también muy frecuentes en la lengua. Otros problemas también relativos a la forma, señalados por Justicia (1995: 50), son los que presentan las palabras homógrafas y homónimas, las formas verbales y nominales de los mismos términos y la falta de profundidad analítica en el aspecto morfológico de las voces. A todos ellos, como queda explicado en § 5.3, nos hemos enfrentado en la delimitación de la unidades léxicas de nuestro estudio y algunos de ellos, concretamente los últimos señalados por Justicia, los hemos solventado con la lematización a que hemos sometido las voces.

A pesar de las objeciones señaladas y, como por otra parte reconocen estudios anteriormente mencionados (Cervero y Pichardo 2000:39; Izquierdo Gil 2002: 201), el cómputo estadístico de las voces, con todas las carencias que pueda presentar, se postula como un instrumento útil para sistematizar el conjunto de datos objeto de estudio, objetivo principal de nuestro trabajo. Jordà Lliteras y Pàmies inciden en esta idea:

Es evidente que una cuantificación de palabras no nos revela ni el sentido ni el entramado de relaciones de un uso determinado del lenguaje, sin embargo, esta misma cuantificación y su análisis estadísticos nos pueden proporcionar un conjunto de datos objetivos, representativos y fiables que nos permitan llevar a cabo una aproximación y estudio cualitativos (Jordà Lliteras y Pàmies 1999: 153).

### 3. PRECEDENTES DE ESTUDIOS ESTADÍSTICOS SOBRE EL LÉXICO DE LA LI

A diferencia de las disciplinas lingüísticas, que prácticamente la han ignorado, las disciplinas pedagógicas se han venido ocupando de la LI principalmente como discurso didáctico, sobre todo a partir de las teorías socioculturales y constructivistas del aprendizaje que ponen el énfasis en la expresión verbal de los contenidos<sup>2</sup>. Concretamente sobre su léxico encontramos dentro de la investigación educativa del español como lengua materna una serie de trabajos sobre el vocabulario de los escolares que parten de dos premisas: la relación que se establece entre el dominio léxico y el dominio de los conceptos y la actuación léxica de los estudiantes como indicador de su rendimiento académico. Podemos agrupar estos estudios en dos tipos según el objetivo fundamental que persigan: aquellos que observan el desarrollo del léxico a lo largo de las etapas educativas y los que miden el vocabulario de los alumnos con el fin de evaluar su competencia léxica en general y, en algunos casos, académica.

Respecto a los primeros, podemos constatar su apogeo en España durante los años 70. A esta época pertenecen los realizados por Díaz Castañón *et al.* (1977-78) y Armayor González (1977-1979), impulsados por los Institutos de Ciencias de la Educación de las universidades de Oviedo y Córdoba respectivamente. Los dos reparan en aspectos dialectológicos, así como tienen en cuenta variables sociológicas tales como la edad, el sexo, la clase social, el nivel de instrucción, etc. Más actualizado, pero en la línea de los anteriores, es *El desarrollo del vocabulario. Diccionario de Frecuencias* de Justicia (1995) que recoge 8.937 vocablos obtenidos tras el recuento de más de medio millón de ocurrencias a partir de muestras escritas de 3.402 escolares de 6 a 13 años. Entre los aspectos reseñables de esta obra destacamos los llamados “vocabularios diferenciales”, con un total de 4.292 palabras, en los que se registran los vocablos que aparecen exclusivamente en cada nivel de enseñanza y que no se repiten en niveles posteriores.

Estudios sobre el vocabulario de los escolares como los mencionados presentan las siguientes características:

- Ofrecen una lista de voces de dimensiones variables según las muestras y el método de recopilación empleados, reunidas a partir de fuentes orales y escritas, o una combinación de ambas, presentadas generalmente de forma alfabética y por frecuencias.
- Se centran en etapas escolares completas: educación infantil, etapa de Enseñanza obligatoria (hasta los 14 años) y/o Bachiller (de 14 a 17 años), según la organización del sistema escolar en el momento de elaboración de la obra.
- Se realizan con una manifiesta intención de servir de instrumento pedagógico, así como para su aplicación en diversos campos como el lingüístico, psicológico y social.
- Algunos de ellos, como los de De la Campa (1996) o Justicia (*op. cit.*), como hemos visto, pretenden servir de referencia para establecer una competencia léxica ideal.

La segunda modalidad de análisis del vocabulario escolar consiste en realizar mediciones de las voces que los alumnos conocen para determinar su competencia léxica en general y, en algunos casos, la académica. Estos cálculos se llevan a cabo utilizando diferentes referencias, a veces muy heterogéneas en sus fuentes. Hay casos en los que el vocabulario de los alumnos se compara, por ejemplo, con el vocabulario individual de un hablante nativo culto o con el que ofrece un diccionario. En otros se procede a contrastarlo con el resultado de recuentos de voces, basados fundamentalmente en la frecuencia, elaborados sobre un corpus de autores autorizados (por ejemplo, literarios) o a partir de encuestas que recogen, de forma casi siempre oral, pero también escrita, el vocabulario de un determinado grupo. De este tipo de estudios, aunque fuera del ámbito español y teniendo como fin su aplicación didáctica en cursos de L2, traemos a colación dos trabajos que han inspirado el nuestro. Ambos surgieron de un coloquio organizado por la Comisión Europea y el Servicio de Didáctica de la Lengua de la Uni-

versidad de Mons-Hainaut que tenía como objetivo reflexionar sobre la manera de facilitar la integración de los niños de los trabajadores inmigrantes en la escuela (Braun y Forges 1998). Se consideró que una vía de integración era el dominio de la lengua de enseñanza y, por ello, en los diferentes talleres del coloquio se expusieron iniciativas ya realizadas y otras por emprender, dirigidas todas ellas a mejorar la enseñanza de la LI. En uno de estos talleres se expusieron dos de las actividades centradas en el vocabulario llevadas a cabo en Italia. Acciaroli y Vertecchi (1998: 39-44) dan cuenta de la primera, que parte de la premisa de que las condiciones de desventaja en la que se encuentra una gran parte de los alumnos inmigrantes en el terreno escolar se deben en gran medida a un escaso conocimiento del vocabulario:

Il faut considérer qu'il existe une corrélation forte entre le niveau des compétences qui dépendent en large mesure du patrimoine lexical disponible, comme la capacité de compréhension à la lecture, et le niveau qu'on obtient dans l'apprentissage de différents secteurs disciplinaires. En d'autres termes, les élèves menant avec succès leur apprentissage sont ceux qui connaissent et utilisent une grande quantité de mots différents (Acciaroli y Vertecchi *ibid*: 42).

La estrategia que se adopta para remediar esta situación consiste en realizar un cálculo del número de voces que los alumnos inmigrantes conocen y compararla a continuación con el léxico de frecuencia de la lengua italiana elaborado por el Centro Electrónico de Cálculo de la Universidad de Pisa. Los datos para cuantificar el patrimonio léxico se obtienen a partir de un programa automatizado creado *ad hoc* por el Laboratorio Pedagógico Experimental de la Universidad de Roma Tre. Con todo ello se busca un diagnóstico previo e individualizado de cada alumno que oriente las futuras intervenciones didácticas.

En la segunda actividad, de la que da cuenta Tanini (1998: 59-64), se elaboró, en primer lugar, un diccionario de frecuencias con voces de manuales escolares y explicaciones orales de dos disciplinas (Matemáticas y Geografía) correspondientes a cursos de alumnos de 7 y 8 años. En un segundo momento, a partir de este diccionario, se confeccionaron unos cuestionarios que fueron administrados a los niños inmigrantes para medir su competencia léxica activa y pasiva y verificar de este modo su comprensión lingüística y el empleo de la lengua. Estas mismas encuestas se utilizaron con niños autóctonos para identificar dificultades lingüísticas. Se trata en ambos casos, como vemos, de modelos de diagnóstico de necesidades educativas de carácter léxico.

Por último, en este repaso por las obras que se han ocupado de léxico de la LI cabe mencionar una que consideramos precedente directo de nuestra investigación por el fin para el que se concibe, el plano de la lengua en el que se centra, el origen de la muestra y el tipo de estudio que acomete: *El vocabulario general de orientación científica y sus estratos* de García Hoz (1976). Esta obra incluye diferentes recuentos con fines pedagógicos para facilitar la enseñanza sistemática y organizada del vocabulario, entendido este como la expresión de los contenidos científicos, y contribuir con ello a la formación científica de los estudiantes. El corpus lingüístico sobre el que trabaja lo componen

libros de texto correspondientes a trece asignaturas de los niveles de Bachillerato y Preuniversitario del currículo en vigor en los años en los que la investigación tuvo lugar. En total se manejaron 2.302.360 ocurrencias correspondientes a 25.402 palabras diferentes, que son las que forman el *Vocabulario general de orientación científica* (García Hoz 1976: 26-168). A partir de esas voces se llevaron a cabo distintos agrupamientos según un criterio de frecuencia (absoluta y relativa) y *repartición (sic)*, que indica el número de contextos diferentes –materias en las que aparecen.

#### 4. JUSTIFICACIÓN DEL CORPUS

##### 4.1. *El libro de texto como muestra escrita de español LI*

La LI se realiza en el proceso de comunicación didáctica en ambas modalidades de discurso, oral y escrita, que suelen darse de forma complementaria en el tratamiento pedagógico de los temas en el contexto escolar. Ahora bien, si el discurso hablado ocupa una parte importante del tiempo en las clases –y en gran medida es desplegado por el profesor en su acción didáctica–, en nuestro sistema de enseñanza la lengua escrita tiene un peso mayor. Por un lado, es de destacar la contribución del lenguaje escrito a la ciencia y a la transmisión de los contenidos científicos. Por otro, y acercándonos al quehacer en las aulas, podemos decir que la lengua escrita actúa como modelo y soporte para el desarrollo de la comprensión lectora y la destreza escrita. Ambas destrezas son imprescindibles para adquirir el dominio de la LI (Guash Boyé 2000: 274). Es también en ellas, como indican Calsamiglia y Tusón (1999: 79), en las que se basa en gran medida la valoración del nivel de conocimientos de los alumnos por parte del profesor: “en el momento de sancionar la adquisición de conocimientos también es casi siempre el modo escrito en que se toma como referencia y objeto de evaluación”.

Entre los exponentes de discurso escrito en el aula es sin duda el libro de texto el que tiene un mayor protagonismo en las prácticas educativas. Tanto es así que, como han señalado Perales y Jiménez, constituye un elemento de primer orden como vehículo de conocimientos en el aprendizaje escolar:

Uno de los pilares básicos sobre los que se sustenta la acción docente, en cualquier nivel educativo, es el libro de texto. Resulta hoy por hoy incuestionable su poderosa influencia en el trabajo en el aula, tanto para los profesores como para los alumnos, constituyéndose en bastantes ocasiones como el referente exclusivo del saber científico (Perales y Jiménez 2002: 369)<sup>3</sup>.

Efectivamente, el manual de la asignatura suele actuar como principal fuente de información, y es además generador de las prácticas educativas. La contribución de los libros de texto a los saberes de los alumnos es asimismo esencial, ya que en gran medida contribuyen a modelar el aprendizaje del alumno, que adquiere el conocimiento de la materia a través de ellos.

El libro de texto en el que basamos nuestro estudio, *Ciencias Sociales, Geografía, LIMES 3*, Editorial Vicens Vives, año 2002, fue elegido tras realizar una encuesta a los profesores de Institutos públicos de Navarra en el año 2003 y constatar dos aspectos que confieren representatividad a la muestra: que era el de mayor implantación en esa asignatura en aulas ordinarias de 3º E.S.O y que realmente se utilizaba en las aulas.

#### 4.2. *La elección de la materia y el nivel escolar*

En la elección de un curso avanzado dentro de la Educación Secundaria Obligatoria, 3º de la ESO, se ha tenido en cuenta el progresivo aumento en la dificultad de los conceptos a lo largo de la escolaridad, a los que corresponde también una LI cada vez más elaborada y compleja. En concreto, en el curso elegido, se produce el mayor salto cualitativo de la etapa Secundaria obligatoria en cuanto a la complejidad conceptual e idiomática se refiere (IES Sancho III el Mayor de Tafalla 2003: 37-38). En este curso también se constata un descenso en los resultados académicos en los alumnos inmigrantes (Defensor del Pueblo 2003: vol. II, 237 y 273).

La decisión sobre qué asignatura había de formar el corpus, sin embargo, no resultó fácil ya que cualquiera de las que forman el currículo de Secundaria podía servirnos: todas son objeto de estudio obligatorio por los alumnos en edad escolar, con excepción de las lenguas extranjeras y las materias optativas. Varias son las razones que nos inclinaron por la materia de Ciencias Sociales. En primer lugar se pensó en una asignatura que permaneciese invariable en cursos venideros habida cuenta de los continuos cambios a los que el sistema educativo español está siendo sometido en los últimos años (leyes LOGSE, LOCE y LOE). En segundo lugar, se buscó una materia que estuviese presente en cursos inmediatamente anteriores al elegido, es decir, 1º y 2º de la ESO, con vistas a un futuro estudio sobre gradación léxica. Un tercer aspecto que se tuvo en cuenta es que los contenidos temáticos del área de las Ciencias Sociales se consideran especialmente adecuados para su tratamiento en programas de segundas lenguas (Snow *et al.* 1989: 212 y Met 1991: 289). Entre ellos citamos algunos presentes en el manual que hemos manejado: las características geopolíticas de los diferentes países, presentadas junto con las de la tierra de acogida, España, y los temas correspondientes a las actividades humanas y a las migraciones. Estos temas, por un lado, ayudan a los alumnos a familiarizarse con el nuevo entorno en que les toca vivir al tratar las características geográficas y culturales –instituciones, organización social, etc.– de la nueva comunidad y el nuevo país. Por otro lado, en el estudio de otras partes del mundo, los alumnos inmigrantes pueden aportar sus conocimientos como exponentes de sus respectivas culturas. Al lado de esto, su propia experiencia vital les ayuda a la contextualización de ciertos contenidos.

En definitiva, la cercanía de los alumnos a los temas, cognitiva y emocional, puede desarrollar elementos positivos en el aprendizaje en diversas direcciones. Areizaga Orube (2000: 153) dice al respecto: “tener un mayor conocimiento del mundo es importante para un mejor desarrollo y dominio de la lengua para la comunicación, incluida la comunicación que se da en el aula”. Efectivamente, la proximidad de los contenidos a la

experiencia del alumno favorece una dinámica de interacción en la clase y posibilita la participación del estudiante, redundando todo ello en un aumento de motivación de los alumnos por la materia (Ruiz Bikandi 2000: 86).

A la idoneidad de los contenidos de la asignatura de Geografía hay que añadir otro elemento que la hace especialmente apta para los niños inmigrantes y autóctonos: el de la finalidad que persigue esta área del conocimiento en el currículo de Secundaria. Según García Hoz (1996: 27 y ss.), esta finalidad no es otra que la educación para la ciudadanía, es decir, la formación para actuar en sociedad. Este mismo autor señala que la adquisición de unos conocimientos debe ir acompañada del desarrollo de unos hábitos para la vida social basados en unas actitudes, entre las que destaca la de respecto y tolerancia hacia todos los miembros de la sociedad. A este tipo de hábitos y actitudes se refiere precisamente Trujillo Sáez (2001) cuando nos habla de la competencia intercultural como componente básico de los currículos de lenguas extranjeras.

## 5. LA ELABORACIÓN DEL ÍNDICE DE FRECUENCIAS. CRITERIOS PARA LA AGRUPACIÓN Y EL RECUENTO DE LAS FORMAS LÉXICAS

### 5.1. *Aclaración terminológica*

Nos referimos al ordenamiento de voces que presentamos como *índice de frecuencias* con el sentido que le da Ávila Muñoz (1999: 62): “el índice de frecuencias de un léxico de una lengua está formado por un listado de palabras al que se le adjunta una determinada frecuencia de aparición en un corpus considerado”, porque expresa el criterio que se ha tomado para su elaboración. En obras lexicográficas encontramos otras denominaciones con significados semejantes: *diccionario*, *repertorio* y *listado de voces* –que nosotros utilizamos como variantes estilísticas– y también *glosario* o *vocabulario*, de los que Jordá LLiteras y Pàmies (1999: 155) dicen: “ofrece(n) todas las unidades del texto, presentadas en orden *lexicográfico* (alfabético) o en orden *lexicométrico* (según frecuencias y recurriendo al orden alfabético en los casos de igual frecuencia)”.

Es preciso aclarar además para una correcta interpretación de los diferentes cálculos realizados sobre las voces del corpus, y que se muestran en § 5.4, otros términos específicos de la estadística lingüística que designan conceptos relacionados con la confección y presentación de listas de palabras en general. Utilizamos el término *vocablo* –equivalente a *type* en el programa informático utilizado– para cada una de las “diferentes unidades que componen el vocabulario” (Justicia 1995: 11), lo que en ocasiones no es lo mismo que decir las diferentes entradas léxicas de un vocabulario. Siguiendo al mismo autor, denominamos *palabras –token–* a “todas y cada una de las unidades elementales, perfectamente distinguidas por la tipografía, que integran un texto”. Asimismo, Jordá LLiteras y Pàmies (1999: 154) definen *palabra* como “toda ocurrencia de un vocablo cualquiera” en un texto. Consecuentemente, el número de *tokens* de un texto es la suma de las ocurrencias de los diferentes *types* (*distinct words*). Comparando unos y



otros obtenemos la *Type/Token Ratio* (TTR), es decir, la proporción entre *type* o palabras diferentes y *tokens* o número de palabras totales de un texto que determina la *variedad léxica* de este.

Finalmente, tomamos el término *lema* en el sentido en que lo hacen Jordá LLiteras y Pàmies (1999: 154), como “reunión de los vocablos que cuentan con una misma raíz y con un significado equivalente para los fines del recuento (p.e.: categoría de verbos personales = amo, amaba, amé)”, lo que en nuestro caso se corresponde con las entradas léxicas. Por *lematización* entendemos “el proceso mediante el cual representamos bajo una forma léxica única todas las posibilidades paradigmáticas de un término” (Ávila Muñoz 1999: 56).

## 5.2. Procedimiento seguido para la confección del índice

En primer lugar se procedió al tratamiento óptico del libro de texto por medio de un escáner y, con ayuda de una herramienta de reconocimiento de caracteres (OCR), a su almacenamiento en documentos *Word* (.doc). En ellos se recogieron todos los caracteres de texto existentes, incluyendo pies de foto, leyendas de mapas, cuadros y tablas estadísticas o de datos, esquemas conceptuales, fórmulas, etc. El único texto no incluido en nuestra lista es el que aparece dentro de las fotografías o dibujos del libro (en vallas publicitarias, logotipos de marcas comerciales y semejantes), por estimar casual y, por ello, irrelevante su presencia en el manual.

Una vez los documentos que forman parte del libro elegido en el ordenador, se procedió a una primera limpieza y corrección de estos haciendo uso de la función “ortografía y gramática” del programa de tratamiento de textos de *Microsoft Word*, comprobando, en el manual que sirvió de referencia, cada caso que esta herramienta daba como dudoso o erróneo, y corrigiéndolo si era necesario. Hecho esto se obtiene lo que se podría llamar “texto en crudo”, esto es, una reproducción casi exacta<sup>4</sup> y apta para el tratamiento informático de los caracteres alfanuméricos existentes en los manuales de origen.

A partir de estos documentos se confeccionaron una primera lista de palabras, para lo cual se tuvieron que convertir previamente en documentos de texto (.txt), pues así lo exigía el programa informático utilizado. Se trata, en concreto, de un programa específico desarrollado por Mike Scott para Oxford University Press llamado *WordSmith Tools* (versión 4.0)<sup>5</sup>, que permite realizar tanto listados de palabras a partir de textos dados como mostrar las concordancias de cualquier palabra en estos, e incluso la extracción de las palabras clave de un texto. De esta lista original se quitó todo aquello que no fueran estrictamente palabras ya que el programa permite eliminar por defecto los caracteres numéricos, pero no así otros que son alfabéticos pero no lingüísticos, como las series ordenadas tipo *a), b), c)*, etc., los números romanos o caracteres no reconocidos que se asimilan al símbolo # y aparecen en la lista. Una vez obtenida la lista base con la que poder trabajar se procedió seguida y simultáneamente tanto al proceso de lematización de los vocablos según los criterios apuntados más abajo como a la desambiguación de los mismos, teniendo como resultado la lista final en el formato propio de *WordList* para

*WordSmith Tools* (.lst). Posteriormente pasamos esta lista, que constituye ya el *índice* que perseguíamos, a documento .doc para una presentación más clara.

Por último, en el *índice*, del que ofrecemos en el ANEXO las primeras entradas hasta la frecuencia 34 por razones de espacio, las voces se muestran en orden lexicométrico, de mayor a menor frecuencia. En cada entrada o artículo del repertorio ofrecemos cuatro campos que, respectivamente, muestran la palabra en cuestión (entrada léxica o artículo), el número de veces que aparece en el texto (bajo el epígrafe “frecuencia”), el porcentaje que ese número representa con respecto al total de palabras –*tokens*– y, finalmente, los lemas correspondientes a cada entrada, si los hay.

### 5.3. *Criterios de lematización*

Ofreceremos a continuación los criterios de lematización seguidos para la realización del índice, aludiendo asimismo a los problemas o deficiencias que este pueda presentar, y que en nuestro descargo debemos decir que no han encontrado todavía solución definitiva<sup>6</sup>.

Las dificultades surgen desde el mismo momento en que debemos elegir la unidad básica del recuento, pues la elección de la *palabra* como tal, si bien es lo habitual, no es un criterio definitivo. Por los resquicios de la definición al uso de *palabra* (‘unidad gráfica o conjunto de signos separados de los grupos vecinos por espacios en blanco o algún signo de puntuación’ –Hockett (1958 *apud* Justicia 1995: 13)– o ‘unidad semántica mínima dotada de *autonomía* en el sentido de operatividad comunicativa’ –Ávila Muñoz (1999: 53)– se nos cuelean, en castellano, grupos como *al* o *del*, formas verbales compuestas con verbo auxiliar (*¿ha* y *comido* se cuentan como dos palabras o como una sola?), formas pronominales átonas escritas junto a las verbales (*cógeme*, *díselo*) y otras, a pesar de que se nos avise normalmente de que se toma como modelo el criterio adoptado por la RAE para la delimitación de las entradas de su *Diccionario*. Fuera de esto, el proceso de lematización también presenta sus controversias teóricas: ¿debemos agrupar *niño* y *niña*, por tratarse de sustantivos, como una o como dos palabras? ¿En el caso de que las consideremos dos palabras diferentes, qué haremos con *alto* y *alta*? O bien, ¿agruparemos conjuntamente las formas de singular y de plural cuando sus sentidos correspondientes sean habitualmente distintos, como en el caso de *actividad* (laboral, física...) y *actividades* (los ejercicios de la lección)? Los hay que deciden, por ejemplo, contar tantos *type* como variantes formales encuentran, “basándose en que no existe suficiente base empírica que permita afirmar que la frecuencia de una determinada palabra pueda ser considerada la suma de las distintas presentaciones de la misma (masculino, femenino, etc.)”<sup>8</sup>. Otro problema se presenta a la hora de tratar lo que se suele denominar como *locuciones* o *expresiones léxicas pluriverbales* como *a lo largo de*, *en ese sentido*, *por lo tanto* o *sin embargo*.

Fuera de estas cuestiones nos encontramos en la práctica con los casos de los homógrafos u homónimos (por ejemplo, *casa* sustantivo o forma verbal de *casar*), pues la moderna recolección de datos por medios informáticos no distingue entre ambos usos,

y se necesita comprobar pacientemente las concordancias de cada palabra para separarlos. Siendo esto, evidentemente, lo aconsejable, hay recuentos que no toman en consideración asunto tan importante<sup>9</sup>, por lo que debemos estar prevenidos al respecto a la hora de comparar repertorios. Añádanse a ellos los homónimos sintácticos (por ejemplo, diversos usos de *que* como nexos o conjunción) y morfológicos (*decía*, 1ª o 3ª persona). E incluso las erratas tipográficas y los solecismos de cualquier tipo tienen su comentario y su dificultad teórica a la hora de elaborar los recuentos. En estos casos decidimos como norma general corregir por nuestra cuenta tales errores dado el carácter formal del vocabulario con el que trabajamos.

Finalmente, hemos aplicado las siguientes premisas de agrupación, que son por otro lado, las más comunes:

- Separamos los homónimos de las palabras de contenido léxico (sustantivos, adjetivos, verbos, adverbios), aunque no distinguimos siempre expresamente las voces según categorías gramaticales, ya que, tratándose de un recuento orientado a la enseñanza de español como L2, consideramos preferible trabajar posteriormente los diferentes usos de una misma forma en contexto, esto es, a partir de las listas de concordancias, de lo que algún ejemplo damos más abajo.
- Agrupamos tanto masculinos y femeninos como singulares y plurales de los adjetivos, así como singulares y plurales de los sustantivos y, obviamente, las diferentes formas simples de la conjugación verbal. No obstante, nos ha parecido conveniente ofrecer las ocurrencias de cada vocablo, de lo que se pueden extraer conclusiones relevantes. Como quiera que hemos decidido, siguiendo a otros listados, considerar los participios aparte, nos quedarán “sueltos” diversos usos de los verbos *haber* y *ser* como auxiliares. Somos conscientes de que tal criterio puede afectar a la parte al recuento final y a la exactitud de los resultados estadísticos, aunque también es cierto que esto sucederá en una proporción insignificante, y que no afecta a nuestro objetivo de ofrecer las palabras de los textos organizadas por frecuencias para extraer las pertinentes consideraciones didácticas.
- No aparecen como lemas o entradas léxicas sino formas que realmente aparecen en el texto. Así pues, no consignamos el singular *competidor* si no aparece más que *competidores* en el texto. En este sentido, priman en la lematización los masculinos sobre los femeninos y los singulares sobre los plurales en sustantivos y adjetivos. En el caso de los verbos, prima el infinitivo, después las formas del indicativo sobre las de los demás modos, los tiempos de presente sobre los de pasado y futuro, las formas de singular sobre las del plural y las personas según su orden.
- En cierta contradicción con lo anterior, no agrupamos las diferentes formas del artículo determinado ni de los pronombres personales del singular, ni distinguimos la homonimia entre pronombres y artículos (por ejemplo usos de *la*), simplemente porque no son el objeto preferente de nuestro estudio.

- Es lógico, creemos, que no delimitemos ni agrupemos las aludidas “expresiones pluriverbales”, asunto que creemos todavía se encuentra en una fase de discusión en cuanto al concepto mismo del término y su alcance.
- Sin embargo, al tratar en futuros trabajos las concordancias de las voces más frecuentes, tendremos la oportunidad de observar las *colocaciones*, es decir, aquellas ocurrencias que muestran una tendencia a aparecer asociadas con otras. Por ejemplo, la voz *comercial* aparece unida a *balanza* en 16 de las 52 ocurrencias de esta palabra.
- Hemos mantenido las siglas como unidades de recuento, como si fueran palabras, pues así las van a encontrar los estudiantes inmigrantes en el libro correspondiente.
- Por el contrario, ya hemos indicado que hemos hecho una limpieza previa a la presentación de las tablas de todos aquellos signos gráficos que nos parecen irrelevantes o no significativos en un recuento como este, a saber: las cifras, tanto arábigas como romanas, las letras que señalan orden en una enumeración o tabla (del tipo *a.-*, *b.-*, *c.-*, etc.), y otros (saltos de página, órdenes de formato...) que el ordenador no reconoce y agrupa con el signo #.

#### 5.4. *El índice en cifras*

Las cifras manejadas a lo largo del proceso de elaboración del índice de frecuencias de la asignatura de Ciencias Sociales III de 3<sup>o</sup> de la ESO, junto con las finales correspondientes al número total de entradas léxicas son las siguientes:

<i>Tokens (running words)</i> o palabras totales	112.477
<i>Tokens</i> o palabras utilizadas	108.637
<i>Types (distinct words)</i> o vocablos	8.896
<i>Type/token ratio</i> o ratio vocablo/ palabra (variación léxica)	8.19
Lemas (entradas léxicas del repertorio)	5.496

## 6. CONCLUSIONES

Este artículo surge de la necesidad de partir de un conocimiento riguroso y actualizado del léxico de la LI en el que basar las propuestas didácticas dirigidas a alumnos en proceso de adquisición de la lengua española inmersos en la Educación obligatoria en centros monolingües de nuestro país. La inexistencia de datos empíricos sobre las voces que forman la LI en estos días y el amplio número que estas puedan suponer nos ha llevado a considerar adecuado el criterio cuantitativo, por frecuencia, para un primer acercamiento al tema. Así se ha procedido a ordenar una muestra de voces formada por las de un libro de texto de la asignatura Ciencias Sociales de 3<sup>o</sup> de la E.S.O por el número

de veces que en ella aparecen. El resultado es un *índice de frecuencias léxicas lematizado* que, con las limitaciones que pueda presentar (§ 2), se nos revela como una importante fuente de datos sobre el uso de las palabras, además de servir como modelo para posteriores trabajos.

La manera en la que se ha procedido al ordenamiento de los vocablos, con medios informáticos y manuales, puede ser tomada en cuenta para la confección de otros índices que recojan el vocabulario de otras asignaturas, en sus respectivos niveles, con el objetivo de llegar a una estimación real del número de voces que en nuestros días componen la LI. El programa informático empleado arroja también otros datos, como la *variación léxica*, que orienta sobre el grado de complejidad léxica de cada materia (§ 5.4).

Estos hallazgos resultan de gran utilidad no solo a la Didáctica del Español como L2, sino a la Didáctica de las respectivas asignaturas. La *lengua de instrucción*, recordemos, da soporte verbal a los distintos contenidos curriculares y compete, por lo tanto, a las diferentes áreas del saber y a las disciplinas que se ocupan de su enseñanza, tanto como lengua materna para los alumnos autóctonos como L2 para alumnos inmigrantes. Conocer, por tanto, el número de voces que componen cada materia, daría lugar a múltiples aplicaciones, como establecer objetivos concretos respecto a la competencia léxica que deben adquirir los alumnos en cada asignatura y curso académico o servir de criterio para evaluar el conocimiento léxico del que parten. Este tipo de estudios nos sitúa además en línea con otros que, con el mismo fin, constituir un instrumento didáctico con miras al desarrollo de la competencia léxica de los escolares nativos y no nativos, se están llevando a cabo actualmente en países como Francia e Italia (§ 3). También se enlaza con anteriores trabajos en el ámbito español sobre el léxico de los escolares, trabajos que se han quedado desfasados no por la validez del método sino por la transformación a nivel lingüístico y textual que la LI ha sufrido en los últimos años debido a la evolución de los sistemas educativos en España.

Otra de las utilidades de nuestro índice tiene que ver con la representatividad que en él adquieren ciertos grupos de voces. Nos referimos a la posibilidad de delimitar a partir de él conjuntos léxicos con fines didácticos como el formado por las palabras más frecuentes, que proporcionan información acerca de sobre qué vocablos incidir para rentabilizar la enseñanza-aprendizaje, además de constituir el *vocabulario temático* de la asignatura. En el índice podemos acotar asimismo vocabularios basados en la frecuencia esenciales para planificar el aprendizaje del vocabulario del español como L2, como el *vocabulario académico* o el *vocabulario técnico* (cf. López Pérez 2006; 2007b; 2008b).

Los agrupamientos por lemas que muestra el índice pueden dar lugar a interesantes reflexiones no solo para la Didáctica del Español como L2 sino para la Lingüística Descriptiva. El comentario que suscitan las distintas formas verbales con sus correspondientes frecuencias reunidas en torno a un lema, o el hecho de que determinadas formas nominales evidencien una desproporción entre sus usos en singular o plural no son sino una de las innumerables reflexiones que se puedan producir.

Por último, es posible también extraer del índice conclusiones que exceden lo puramente lingüístico: el vocabulario del libro de texto de Ciencias Sociales evidencia una ideología concreta en el discurso escolar susceptible de ser analizada, por ejemplo, por la Teoría de la Educación. Queda claro, por tanto, que estudios como el presentado en este artículo, que ofrecen datos léxicos sistematizados por frecuencia y lematizados, se prestan a diferentes y provechosas lecturas desde distintas disciplinas y pueden contribuir definitivamente a un mejor conocimiento del léxico de la LI, conocimiento que debe guiar su enseñanza en los diferentes contextos en el que esta se ve implicada: como lengua segunda, como lengua materna, en aulas específicas de E/L2 y en aulas ordinarias del resto de las materias.

## NOTAS

\* STUDIES ON ACADEMIC LANGUAGE IN SCHOOL SETTINGS: A WORD FREQUENCY LIST FROM A TEXTBOOK.

\*\* Correspondence to: M<sup>a</sup> Victoria López Pérez, Av. Bayona, 39, 6<sup>o</sup> B, 31011 Pamplona, Spain. E-mail: victoria.lopez@unavarra.es.

1. Vid. López Pérez (2007a) para los objetivos del E/L2 en contextos escolares monolingües.
2. En López Pérez (2008a) se encuentra un breve repaso de los distintos estudios sobre la LI acometidos desde estas disciplinas.
3. En este mismo sentido se pronuncian Mateos Jiménez (1995:117) y Campanario (2003:162). Martínez Losada y García Barros (2003: 243-244) nos ofrecen una actualizada bibliografía sobre cuestiones relacionadas con el empleo del libro de texto en las prácticas educativas.
4. A pesar de haber revisado página a página y párrafo a párrafo el texto obtenido, siempre existe un margen de error por lecturas defectuosas que hayan pasado inadvertidas tanto al programa de corrección como a nosotros mismos.
5. Vid. detalles en <<http://www.liv.ac.uk/~ms2928/wordsmith/screenshots/>>. La eficacia de este programa queda avalada por la cantidad y calidad de obras de investigación que lo han usado, cuya lista se puede consultar en <[http://www.lexically.net/wordsmith/corpus\\_linguistics\\_links/papers\\_using\\_wordsmith.htm](http://www.lexically.net/wordsmith/corpus_linguistics_links/papers_using_wordsmith.htm)>. De ellos, podemos reseñar en el ámbito de la lengua española diversos trabajos conjuntos de A. Gómez González-Jover y Ch. Vargas Sierra, de la Universidad de Alicante, sobre diccionarios especializados bilingües.
6. Sobre este asunto, cf. las consideraciones de Ávila Muñoz (1999: 51-71).
7. Ávila Muñoz (*op. cit.*: 61), por ejemplo, incluye estas formas en las entradas de *ser* o *haber*, especificando en su caso si se trata de usos como auxiliar. Desgraciadamente, lo habitual es no entrar en consideraciones al respecto.
8. Alameda y Cuetos (1995 *apud* Ávila Muñoz 1999: 59). Léase a este respecto, las detalladas consideraciones que este último autor hace en el epígrafe “Heterónimos” (*ibid.*: 59-68).
9. Como el *Français fondamentale* (Gougenheim *et al.* 1964 *apud* Ávila Muñoz *op. cit.*: 58) por ejemplo.

## REFERENCIAS

Acciaroli, L. y B. Vertecchi. 1998. “Stratégies d’individualisation de l’enseignement pour lutter contre le désavantage linguistique”. *Enseigner et apprendre la langue de l’école. Vers une culture de la réussite pour tous*. Eds. A. Braun y G. Forges. Paris-Bruxelles: De Boeck y Larcier. 39-44.

- Areizaga Orube, E. 2000. "El enfoque comunicativo. Propuestas didácticas". *Didáctica de la segunda lengua en educación infantil y primaria*. Ed. U. Ruiz Bikandi. Madrid: Síntesis. 137-162.
- Armayer González, H. 1977/1979. *Determinación del Vocabulario Técnico Escolar en el alumno de 8º curso de E.G.B. y su incorporación al B.U.P.* Córdoba: Universidad de Córdoba-Servicio de Publicaciones del ICE.
- Ávila Muñoz, M. 1999. *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Málaga: Universidad de Málaga.
- Braun, A. y G. Forges, eds. 1998. *Enseigner et apprendre la langue de l'école. Vers une culture de la réussite pour tous*. Paris-Bruselas: De Boeck y Larcier.
- Calsamiglia, H. y A. Tusón. 1999. *Las cosas del decir. Manual de análisis del discurso*. Barcelona: Ariel Lingüística.
- Campanario, S.M. 2003. "De la necesidad, virtud: cómo aprovechar los errores y las imprecisiones de los libros de texto para enseñar Física". *Enseñanza de las Ciencias* 21(1): 161-172.
- Cervero, M.J. y F. Pichardo. 2000. *Aprender y enseñar vocabulario*. Madrid: Edelsa.
- De la Campa, H. et al. 1996. *Diccionario inverso del vocabulario escolar básico. Implicaciones didácticas*. Granada: Fundación Educación y Futuro.
- Defensor del Pueblo. 2003. *La escolarización del alumnado inmigrante en España: análisis descriptivo y estudio empírico*. Vols. I y II. [Documento de Internet disponible en <http://www.defensordelpueblo.es/informes2.asp>].
- Díaz Castañón, C. et al. 1977-78. *Vocabulario básico del español y sus aplicaciones a la enseñanza*. Vols. I y II. Oviedo: Universidad de Oviedo.
- García Hoz, V. 1976. *El vocabulario general de orientación científica y sus estratos. (Bases para una enseñanza vertebrada)*. Madrid: Instituto de Pedagogía-CSIC.
- García Hoz, V. 1996. "Enseñanza y formación sociales en Educación Secundaria". *Enseñanza de las Ciencias Sociales en Secundaria*. Eds. V. García Hoz et al. Madrid: Rialp. 17-45.
- Guash Boyé, O. 2000. "La expresión escrita". *Didáctica de la segunda lengua en educación infantil y primaria*. Ed. U. Ruiz Bikandi. Madrid: Síntesis. 273-300.
- IES "Sancho III el Mayor de Tafalla (Navarra). 2003. *Memoria del Aula de inmersión lingüística*. Documento inédito.
- Izquierdo Gil, M.C. 2002. "Los corpóra electrónicos en la enseñanza del vocabulario español LE: la utilidad pedagógica de las concordancias". *Tecnologías de la información y de las comunicaciones de la enseñanza de ELE. Actas del XII Congreso Internacional de ASELE*. Ed. A.M. Gimeno Sanz. Valencia: Universidad Politécnica de Valencia. 201-211.
- Jordá LLiteras, G.M. y C. Pàmies. 1999. "Términos y conceptos específicos en el análisis lexicométrico". *Lengua para fines específicos IV: investigación y enseñanza*. Ed. S. Barrueco. Alcalá: Universidad de Alcalá. 153-157.
- Justicia, F. 1995. *El desarrollo del vocabulario. Diccionario de Frecuencias*. Granada: Universidad de Granada.

- López Pérez, M. V. 2006. "Cuestiones relativas al léxico de la lengua de instrucción en el marco de la enseñanza del Español/L2 a inmigrantes". *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del XV Congreso Internacional de ASELE*. Eds. M<sup>a</sup> A. Castillo Carballo et al. Sevilla: Universidad de Sevilla-Secretariado de publicaciones. 534-541.
- López Pérez, M.V. 2007a. "Los objetivos de la enseñanza de L2 en contextos escolares monolingües". *Hyperfeira. Arts & Literature Internacional Journal* 9. [Documento de Internet disponible en <http://www.sinc.sunysb.edu/Publish/hiper/num9/Articulos/lopez.html>]
- López Pérez, M.V. 2007b. "El vocabulario de las asignaturas del currículo escolar: análisis de datos léxicos para su enseñanza en aulas de E/L2". *Las destrezas orales en la Enseñanza del Español L2-LE. Actas del XVII Congreso Internacional de ASELE*. Vol. II. Ed. E. Balmaseda. 779-794.
- López Pérez, M.V. 2008a. "La lengua de instrucción (LI): aproximación teórica y perspectivas de análisis en el marco de la enseñanza del español/L2". *RESLA* 20: 85-104.
- López Pérez, M.V. 2008b. "La enseñanza del vocabulario de la lengua de instrucción (LI) en E/L2 en contextos escolares: selección de voces y programación". *Actas del XXIV Congreso Internacional AESLA. Aprendizaje de lenguas, uso del lenguaje y modelación cognitiva: perspectivas aplicadas entre disciplinas*. Eds. R. Mairal et al. Madrid: UNED. 919-936.
- Martínez Bonafé, J. 1992. "Siete cuestiones y una propuesta". *Cuadernos de Pedagogía* 203: 8-13.
- Martínez de Sousa, J. 1995. *Diccionario de lexicografía práctica*. Barcelona: Vox-Bibliograf.
- Martínez Losada, C. y S. García Barros. 2003. "Las actividades de Primaria y ESO incluidas en libros escolares, ¿Qué objetivo persiguen? ¿Qué procedimientos enseñan?". *Enseñanza de las Ciencias* 21 (2): 243-264.
- Mateos Jiménez, A. 1995. "Los libros de texto en la enseñanza de las Ciencias Naturales: líneas generales para un uso crítico de los mismos". *Docencia e investigación. Revista de la Escuela Universitaria de Magisterio de Toledo*. Enero-Junio: 177-122.
- Met, M. 1991. "Learning Language through Content. Learning Content through Language". *Foreign Language Annals* 24 (4): 281-295.
- Muñoz López, B. 2002. "La enseñanza del español como Segunda Lengua para alumnos inmigrantes: análisis de la influencia de los factores afectivos y del contexto de aprendizaje". *Aspectos Didácticos de la Lengua y Literatura* 12. Zaragoza: Instituto de Ciencias de la Educación, Universidad de Zaragoza. 15-53.
- Perales, F.J. y J. de D. Jiménez. 2002. "Las ilustraciones en la enseñanza-aprendizaje de las ciencias. Análisis de libros de texto". *Enseñanza de las Ciencias* 20 (3): 369-386.
- Ruiz Bikandi, U. 2000. "La adquisición de la segunda lengua". Ed. U. Ruiz Bikandi, *Didáctica de la segunda lengua en educación infantil y primaria*. Madrid: Síntesis. 67-100.



- Schmitt, N. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Snow, M.A. *et al.* 1989. "A conceptual framework for the integration of language and content in Second/Foreign Language Instruction". *TESOL Quarterly* 23: 201-217.
- Tanini, M. 1998. "ATELIER: Les méthodes pédagogiques". *Enseigner et apprendre la langue de l'école. Vers une culture de la réussite pour tous*. Eds. A. Braun y G. Forges. París-Bruxelles: De Boeck y Larcier. 59-72.
- Terrádez Gurrea, M. 2001. *Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo*. Valencia: Publicaciones de la Universidad de Valencia.
- Trujillo Sáez, F. 2001. "Objetivos en la enseñanza de lenguas extranjeras: de la competencia lingüística a la competencia intercultural". *Inmigración, interculturalidad y convivencia*. Coords. F. Herrera Clavero *et al.* Ceuta: Instituto de Estudios Ceutíes. 407-418.

ANEXO: MUESTRA DEL ÍNDICE DE FRECUENCIAS LEMATIZADO DEL LIBRO DE TEXTO *CIENCIAS SOCIALES 3º ESO*

ENTRADA	FREC.	%	LEMATIZACIÓN
DE	7.761	6.90	
LA	4.753	4.23	
Y	3.281	2.92	
EL	3.205	2.85	
EN	3.015	2.68	
LOS	2.530	2.25	
LAS	2.296	2.04	
QUE	2.166	1.93	
UN	1.917	1.70	un [950] una[808] uno[70] unos[48] unas[41]
SER	1.766	1.57	ser[152] es[862] son[554] era[40] siendo[12] sea[28] sean[23] fue[20] será[18] eran[14] serán[11] fueron[7] sería[7]...
A	1.648	1.47	
SE	1.615	1.44	
DEL	1.372	1.22	
POR	1.028	0.91	
PAÍS	819	0.73	país[207] países[612]
QUÉ	812	0.72	
CON	747	0.66	
MÁS	725	0.64	
PARA	696	0.62	
ESTE	666	0.59	este[238] estos[151] estas[146] esta[131]
HABER	654	0.58	haber[5] ha[253] han[180] hay[144] había[11] has[15] haya[13] he[6] habrá[6] hemos[5] habían[4] hubo[4] hayan[3]...
SU	603	0.54	sus[397] sus[206]
POBLACIÓN	543	0.48	población[537] poblaciones[6]
O	531	0.47	o[529] ó[2]
CIUDAD	441	0.39	ciudad[193] ciudades[248]
COMO	436	0.39	
GRAN	430	0.38	gran[191] grande[9] grandes[230]
ACTIVIDAD	375	0.33	actividad[78] actividades[297]
ESPAÑA	372	0.33	
PODER (v.)	360	0.32	poder[17] pueden[131] puede[114] puedes[21] pueden[15] podría[11] podría[10] podrían[7] podría[s] [5] podrán[4]...
TENER	358	0.32	tener[49] tiene[123] tienen[110] tienes[14] tenemos[13] tendrá[8] tengal[7] teniendo[7] tenían[6] tendrán[5] tendrán[4]...
AL	357	0.32	
EUROPEO	353	0.31	uropeo[60] europeo[244] europeos[26] europeos[23]

ENTRADA	FREC.	%	LEMATIZACIÓN
ENTRE	344	0.31	
NO	288	0.26	
ZONA	275	0.24	zona[78] zonas[197]
OTRO	264	0.23	otro[44] otra[25] otras[8] otros[114]
ESTAR	247	0.22	estar[14] está[12] estás[1] estamos[2] están[104] estarían[1] esté[2] estén[1]
AÑO	240	0.21	año[107] años[133]
SECTOR	235	0.21	sector[166] sectores[69]
MUY	231	0.21	
PRODUCTO	221	0.20	producto[46] prod.[1] productos[174]
MUNDO	219	0.19	
CUÁL	218	0.19	cuál[89] cuáles[129]
INDUSTRIA	217	0.19	industria[118] industrias[99]
UNIÓN	217	0.19	
TODO	212	0.19	todo[68] toda[22] todas[33] todos[89]
AGUA	207	0.18	agua[144] aguas[63]
ESTADO (s.Gob.)	206	0.18	Estado [90] Estados [106] (de ellos, 58 Estados Unidos) EE. [16] (siempre con UU.)
ALGUNO	205	0.18	alguno[7] algún[29] alguna[26] algunas[56] algunos[87]
URBANO	202	0.18	urbano[64] urbana[86] urbanas[3] urbanos[21]
IMPORTANTE	201	0.18	importante[111] importantes[90]
TRANSPORTE	200	0.18	transporte[152] transportes[48]
INDUSTRIAL	193	0.17	industrial[118] industriales[75]
MUCHO (adj.)	192	0.17	mucho[21] mucha[32] muchas[57] muchos[81] muchísimo[1]
LUGAR	188	0.17	lugar[106] lugares[82]
MAYOR	186	0.17	mayor[156] mayores[30]
SERVICIO	186	0.17	servicio[25] servicios[161]
PERSONA	185	0.16	persona[18] personas[167]
TIPO	184	0.16	tipo[135] tipos[49]
ECONÓMICO	181	0.16	económico[76] económicas[46] económicos[59]
LO	180	0.16	
CADA	169	0.15	
ENERGÍA	168	0.15	energía[163] energías[5]
AGRICULTURA	167	0.15	
DESARROLLO	167	0.15	
ESPAÑOL (adj.)	165	0.15	español [53] española [74] españoles [12] españolas [27]
RECURSO	161	0.14	recurso[12] recursos[149]

ENTRADA	FREC.	%	LEMATIZACIÓN
COMUNIDAD	160	0.14	comunidad[85] comunidades[75]
CÓMO	156	0.14	
HACER	156	0.14	hacer[40] hace[48] hace[1] hacen[27] hacerlo[3] hacerse[1] hacersele[1] hacer[2] haciendo[9] haga[1] hagamos[1]...
PARTE	152	0.14	parte[142] partes[10]
TAMBIÉN	152	0.14	
PAISAJE	150	0.13	paisaje[86] paisajes[64]
PRODUCCIÓN	150	0.13	producción[149] producciones[1]
SOBRE	150	0.13	
PRINCIPAL	148	0.13	principal[20] principales[128]
FORMA (s.)	146	0.13	forma[99] formas[47]
TIERRA	142	0.13	tierra[86] tierras[56]
EJEMPLO	140	0.12	ejemplo[110] ejemplos[30]
DONDE	138	0.12	
MUNDIAL	136	0.12	mundial[127] mundiales[9]
HABITANTE	132	0.12	habitante[20] habitantes[82] hab[30]
PERO	131	0.12	
TRABAJO	131	0.12	trabajo[105] trabajos[26]
RELIEVE	128	0.11	relieve[94] relieves[34]
COSTA	127	0.11	costa[55] costas[72]
EMPRESA	126	0.11	empresa[34] empresas[92]
EXPLICAR	126	0.11	explicar[9] explicita[99] explican[8] explicatos[1] explicad[1] expliquen[2] explicarían[1]
CLIMA	123	0.11	clima[93] climas[30]
CRECIMIENTO	123	0.11	
EUROPA	123	0.11	
MAPA	123	0.11	mapa[117] mapas[6]
RÍO	123	0.11	río[38] ríos[1] ríos[3] ríos[81]
TRABAJAR	123	0.11	trabajar[11] trabaja[80] trabajaba[1] trabajan[4] trabajo[1] trabajando[2] trabajan[24]
DOS	122	0.11	
NUEVO	120	0.11	nuevo[14] nuevas[39] nueva[35] nuevos[32]
PRODUCIR	116	0.10	producir[26] produce[38] producen[29] producimos[1] produciría[1] produciria[1] producirse[6]...
VIDA	115	0.10	
MAR	114	0.10	mar[92] mares[22]
POCO	114	0.10	poco[76] poca[12] pocas[9] pocos[17]
PROBLEMA	114	0.10	problema[38] problemas[76]
SOCIAL	114	0.10	social[67] sociales[47]

ENTRADA	FREC.	%	LEMATIZACIÓN
PESCA (s.)	113	0.10	
ETC	113	0.10	
NATURAL	112	0.10	natural[64] naturales[48]
UNIDO	111	0.10	unido[26] unida[3] unidas[7] unidos[59] UU[16]
HASTA	110	0.10	
TU	110	0.10	tu[92] tus[18]
E	108	0.10	
ALTO	107	0.10	alto[20] alta[54] altas[20] altos[13]
BAJO (adj.)	105	0.09	bajo[24] baja[25] bajas[23] bajos[33]
COMERCIAL	105	0.09	comercial[52] comerciales[53]
DEBE	104	0.09	debe[39] deben[41] debéis[1] debemos[4] deberá[4] deberán[3] debería[3] deberían[2] debes[4] debía[1]...
RED	104	0.09	red[77] redes[27]
DESARROLLADO	102	0.09	desarrollado[12] desarrolladas[12] desarrollada[1] desarrollados[77]
EXISTIR	102	0.09	existir[1] existían[1] existía[3] existen[65] existe[27] existían[2] existía[3]
NÚMERO	102	0.09	
YA	102	0.09	
AGRARIO	100	0.09	agrario[29] agrarias[21] agraria[29] agrarios[21]
CREO	101	0.09	creo[1] cree[2] crees[88] creéis[1] creen[6] crea [2] crean [1]
MEDITERRÁNEO (adj.)	100	0.09	mediterráneo[57] (de ellas, 7 Mar Mediterráneo) mediterránea[25] mediterráneos[8] mediterráneas[10]
DIFERENCIA (s.)	99	0.09	diferencia [34] diferencias [65]
VEZ	99	0.09	vez[70] veces[29]
MILLÓN	98	0.09	millón[5] millones[92] mill.[1]
REGIÓN	98	0.09	región[39] regiones[59]
ENCONTRAR	96	0.09	encontrar[10] encuentran[56] encontramos[5] encontrabal[1] encontraréis[1] encontrarse[3] encuentra[17]...
SISTEMA	96	0.09	sistema[70] sistemas[26]
TURISMO	95	0.08	
CUANDO	94	0.08	
DESDE	94	0.08	
NIVEL	94	0.08	nivel[78] niveles[16]
SI	94	0.08	
SÍNTESIS	94	0.08	
MERCADO	93	0.08	mercado[88] mercados[5]
REALIZAR	93	0.08	realizar[19] realizan[25] realizarla[1] realizarías[1] realizas[2] realizó[2] realizad[4] realicen[3]...
SIN	93	0.08	
SIGLO	92	0.08	siglo[84] siglos[8]

ENTRADA	FREC.	%	LEMATIZACIÓN
ISLA	91	0.08	isla[16] islas[75]
CARACTERÍSTICA	90	0.08	característica[9] características[81]
FUNCIÓN	90	0.08	función[50] funciones[40]
MUJER	90	0.08	mujer[39] mujeres[51]
FORMAR	89	0.08	formar[4] forman[40] formando[11] formad[3] formarse[1] formaron[2] forme[1] formó[1]
DECIR	88	0.08	decir[61] di[7] decía[2] decían[1] decimos[4] dice[11] dicen[1] diciéndolo[1]
CAMBIO	88	0.08	cambio[51] cambios[37]
ELEVADO	88	0.08	elevado[22] elevadas[29] elevada[21] elevados[16]
INTERIOR	88	0.08	interior[73] interiores[15]
MENOS	88	0.08	menos[87] meno[1]
CULTIVO	87	0.08	cultivo[39] cultivos[48]
VIVIR	87	0.08	vivir[20] viva[2] vivimos[4] viviendas[1] vivían[2] vivía[4] vivirá[2] vives[10] viven[19] vive[20] vivan[2]
CONTINENTE	86	0.08	continente[62] continentes[24]
NORTE	86	0.08	
POLÍTICA (s.)	86	0.08	política[62] políticas[24] (siempre actuación, ex.gr. <i>política medioambiental</i> )
ACTIVO	83	0.07	activo[3] activos[6] activas[4] activa[70]
COMERCIO	83	0.07	comercio[82] comercios[1]
ESPACIO	82	0.07	espacio[54] espacios[28]
GOBIERNO	82	0.07	gobierno[72] gobiernos[10]
CENTRO	81	0.07	centro[43] centros[38]
CORDILLERA	80	0.07	cordillera[33] cordilleras[47]
DURANTE	80	0.07	
POBRE	80	0.07	pobre[9] pobres[71]
DAR	79	0.07	dar[8] da[27] daban[1] daban[1] damos[2] dan[23] dando[5] daría[1] darles[1] darse[4] dé[2] dieron[1] dio[3]
TERCIARIO	79	0.07	terciario[62] terciaria[1] terciarias[14] terciarios[2]
ÁREA	78	0.07	área[29] áreas[49]
RICO	78	0.07	rico[2] ricas[10] rica[1] ricos[65]
TEMPERATURA	78	0.07	temperatura[19] temperaturas[59]
ADEMÁS	77	0.07	
FACTOR	77	0.07	factor[6] factores[71]
MATERIA	77	0.07	material[22] materias[55]
ECONÓMICA	76	0.07	
SEGÚN	76	0.07	
CIUDADANO (s.)	75	0.07	ciudadano[4] ciudadanos[55] ciudadanas[15] ciudadana [1]
PORQUE	75	0.07	

ENTRADA	FREC.	%	LEMATIZACIÓN
TASA	75	0.07	tasa[61] tasas[14]
TERRITORIO	75	0.07	territorio[65] territorios[10]
NECESARIO	74	0.07	necesario[35] necesaria[9] necesarias[9] necesarios[21]
PERMITIR	74	0.07	permitir[2] permite[33] permiten[23] permito[2] permitir[1] permitiendo[1] permitirán[2] permitirán...
FUENTE	73	0.07	fuentes[21] fuentes[52]
HUMANO	72	0.06	humano[33] humana[12] humanas[5] humanos[22]
OBSERVAR	72	0.06	observar[3] observa[63] observas[6]
SUR	72	0.06	
ACTUALMENTE	71	0.06	
ATLÁNTICO	71	0.06	atlántico[50] atlánticas[4] atlántica[17]
DIFERENTE	71	0.06	diferente[1] diferentes[70]
INFORMACIÓN	71	0.06	información[68] informaciones[3]
PRIMA	71	0.06	prima[16] primas[55]
PROCESO	71	0.06	proceso[62] procesos[9]
TURÍSTICO	71	0.06	turístico[2] turística[8] turísticas[19] turístico[21] turísticos[21]
DISTRIBUCIÓN	70	0.06	
CASO	69	0.06	caso[52] casos[17]
GANADERÍA	69	0.06	
LLEGAR	69	0.06	llegar[21] llega[20] llegará[2] llegando[1] llegán[13] llegad[1] llegaran[2] llegas[2] llego[1] llegue[1]...
IR	67	0.06	ir[12] fbamos[1] irías[1] irse[1] va[17] fueron[7] fue[3] vamos[5] van[16] vas[1] vaya[2] voy[1]
OBTENER	67	0.06	obtener[40] obtendría[1] obtendrá[1] obtenerla[1] obtengan[2] obtiene[11] obtienen[11]
PRESENTAR	67	0.06	presentar[5] presenta[33] presentan[24] presentamos[1] presentarlo[1] presentaron[1] presentarse[2]
COMUNICACIÓN	66	0.06	comunicación[35] comunicaciones[31]
DÓNDE	66	0.06	
MISMO	66	0.06	mismo[43] mismas[4] misma[12] mismos[7]
TRES	66	0.06	
CONJUNTO	65	0.06	conjunto[59] conjuntos[6]
MONTAÑA	65	0.06	montaña[33] montañas[32]
UTILIZAR	65	0.06	utilizar[10] utilizamos[1] utilizando[1] utilizan[25] utiliza[19] utilicen[2] utilizaban[1] utilizarías[1]...
FORMADO	64	0.06	formado[28] formada[26] formadas[6] formados[4]
CENTRAL (adj.)	62	0.06	central [54] centrales [8]
GRUPO	62	0.06	grupo[18] grupos[44]
JAPÓN	62	0.06	
NATALIDAD	62	0.06	
PARTIR	62	0.06	partir[59] partiendo[11] partimos[1] partieron[11]

ENTRADA	FREC.	%	LEMATIZACIÓN
SOCIEDAD	62	0.06	sociedad[49] sociedades[13]
TOTAL	62	0.06	total[61] totales[11]
EXPLOTACIÓN	61	0.05	explotación[49] explotaciones[12]
RELACIÓN	61	0.05	relación[38] relaciones[23]
CANTIDAD	60	0.05	cantidad[39] cantidades[21]
PETRÓLEO	60	0.05	
AUNQUE	59	0.05	
AUTÓNOMA	59	0.05	autónoma[18] autónomas[41]
CONCENTRAR	59	0.05	concentrar[2] concentrar[26] concentra[22] concentre[3] concentrándose[1] concentraron[1] concentrarse[3]
PENINSULA	59	0.05	península[51] penínsulas[8]
PODER (s.)	59	0.05	poder[51] poderes[8]
PRECIPITACIÓN	59	0.05	precipitación[2] precipitaciones[57]
SIGUIENTE	59	0.05	siguiente[13] siguientes[46]
SITUAR	59	0.05	situar[1] situa[21] situado[1] situán[30] situálos[1] situaran[1] situaron[1] situarse[2] situó[1]
TIEMPO	59	0.05	tiempo[54] tiempos[5]
CONSUMO	58	0.05	consumo[57] consumos[11]
NECESIDAD	58	0.05	necesidad[19] necesidades[39]
POLÍTICO (adj.)	58	0.05	político[14] política[16] políticos[15] políticas[13]
SABER	58	0.05	saber[21] sabe[3] sabes[18] saben[1] sabemos[3] sabrías[12]
ASÍ	57	0.05	
SITUADO	57	0.05	situado[6] situados[17] situadas[18] situada[16]
ACTUAL	56	0.05	actual[36] actuales[20]
EJE	56	0.05	eje[26] ejes[30]
ELLO	56	0.05	
MENOR	56	0.05	menor[44] menores[12]
AUMENTAR	55	0.05	aumentar[28] aumentan[5] aumentará[2] aumentarán[1] aumentaría[1] aumente[1] aumentó[3]
CAMPO	55	0.05	campo[33] campos[22]
DEBIDO	55	0.05	
MUNICIPIO	55	0.05	municipio[15] municipios[40]
UE	55	0.05	
ECONOMÍA	54	0.05	economía[49] economías[5]
ÁFRICA	53	0.05	
IMAGEN	53	0.05	imagen[30] imágenes[23]
LARGO	53	0.05	largo[42] largos[1] largal[5] largos[5]
MEDIO (s.instr.)	53	0.05	medio[7] medios[46]



ENTRADA	FREC.	%	LEMATIZACIÓN
SÓLO	53	0.05	
TECNOLOGÍA	53	0.05	tecnología[25] tecnologías[28]
ANTIGUO	52	0.05	antiguo[10] antiguas[12] antigua[7] antiguos[23]
BIENES	52	0.05	
DISPONER	52	0.05	disponer[15] dispondrán[1] disponen[20] disponemos[1] dispone[9] dispones[1] disponga[1] dispongan[1] disponían[3]
ELLAS	52	0.05	
LOCALIZARSE	52	0.05	localizarse[1] localizarla[2] localizan[22] localizalo[4] localizaban[1] localiza[22]
MESETA	52	0.05	meseta[45] mesetas[7]
ORIGEN	52	0.05	origen[50] orígenes[2]
AIRE	51	0.05	aire[47] aires[4]
BUSCAR	51	0.05	buscar[7] buscando[2] buscan[9] busca[33]
CAUSA	51	0.05	causa[11] causas[40]
LITORAL	51	0.05	litoral[43] litorales[8]
MANO	51	0.05	mano[45] manos[6]
PROPIO	51	0.05	propio[17] propios[9] propia[12] propias[13]
RURAL	51	0.05	rural[39] rurales[12]
SITUACIÓN	51	0.05	situación[37] situaciones[14]
TRABAJADOR	51	0.05	trabajador[4] trabajadora[1] trabajadoras[1] trabajadores[45]
MANTENER	50	0.04	mantener[16] mantenerlos[2] mantengan[1] mantenían[1] manteniendo[1] mantiene[12] mantienen[14]
MORTALIDAD	50	0.04	
ORGANIZACIÓN	50	0.04	organización[40] organizaciones[10]
RECIBIR	50	0.04	recibir[6] recibiendo[1] recibían[1] reciben[20] recibe[21] recibirán[1]
AMÉRICA	49	0.04	
CONSECUENCIA	49	0.04	consecuencial[23] consecuencias[26]
DESTACAR	49	0.04	destacar[6] destaca[19] destacan[19] destacarías[1] destacas[1] destacamos[2] destacaba[1]
MANERA	49	0.04	
MEDIO (s.lug.)	49	0.04	medio[48] medios[11]
MIEMBRO	49	0.04	miembro[4] miembros[45]
NUMEROSO	49	0.04	numeroso[1] numerosas[16] numerosos[29] numerosa[3]
OCUPAR	49	0.04	ocupar[2] ocupa[31] ocupaban[3] ocupando[1] ocupan[10] ocuparse[1] ocupe[1]
PEQUEÑO	49	0.04	pequeño[8] pequeñas[18] pequeña[10] pequeños[13]
DIVERSOS	48	0.04	diversos[17] diversas[31]
INMIGRANTE	48	0.04	inmigrante[3] inmigrantes[45]
LLUVIA	48	0.04	lluvia[16] lluvias[32]
MERCANCÍA	48	0.04	mercancía[9] mercancías[39]

ENTRADA	FREC.	%	LEMATIZACIÓN
PRIMARIO	48	0.04	primario[39] primaria[6] primarios[3]
SIDO	48	0.04	
VEGETACIÓN	48	0.04	
AGRÍCOLA	47	0.04	agrícola[19] agrícolas[28]
CUENTA	47	0.04	
GLOBALIZACIÓN	47	0.04	
IMPORTANCIA	47	0.04	parlamento[46] parlamentos[1]
PARLAMENTO	47	0.04	posible[37] posibles[10]
POSIBLE	47	0.04	tú[2] te[44] ti[1]
TÚ	47	0.04	valor[35] valores[12]
VALOR	47	0.04	
BIENESTAR	46	0.04	
FLUJO	46	0.04	flujo[8] flujos[38]
MINERAL	46	0.04	mineral[20] minerales[26]
OBRA	46	0.04	obra[43] obras[3]
OFRECER	46	0.04	ofrecer[8] ofrecen[14] ofrecemos[1] ofrece[18] ofrecerlos[1] ofrecían[1] ofrecían[1] ofrecen[2]
PRIMER	46	0.04	primer[12] primera[18] primeras[8] primeros[8]
VOCABULARIO	46	0.04	
CONDICIÓN	45	0.04	condición[3] condiciones[42]
CONSTRUCCIÓN	45	0.04	construcción[44] construcciones[1]
NUESTRO	45	0.04	nuestro[21] nuestras[4] nuestra[14] nuestros[6]
OBJETIVO	45	0.04	objetivo[20] objetivos[25]
RECORDAR	45	0.04	recordar[14] recordamos[3] recordaremos[1] recordad[1] recordaba[1] recorda[13] recuerdan[1] recuerdas[11]
BÁSICO	44	0.04	básico[6] básicas[17] básica[6] básicos[15]
CONOCER	44	0.04	conocer[26] conocen[1] conocéis[1] conoce[6] conocerla[1] conoces[6] conocía[2] conocían[1]
ESE	44	0.04	ese[20] esa[9] esas[8] esos[7]
HACIA	44	0.04	
INFLUENCIA	44	0.04	
USO	44	0.04	uso[42] usos[2]
ALTITUD	43	0.04	altitud[36] altitudes[7]
ANIMAL	43	0.04	animal[12] animales[31]
DEPRESIÓN	43	0.04	depresión[26] depresiones[17]
ESTABLECER	43	0.04	establecer[13] establece[14] establecen[9] establecerse[1] establecían[1] establecieron[1] estableció[4]
FUTURO	43	0.04	
PLANETA	43	0.04	

ENTRADA	FREC.	%	LEMATIZACIÓN
REDUCIR	43	0.04	reducir[22] reducen[8] reduce[3] reducimos[1] reducirán[1] reducirlo[1] reducirlos[1] reducirse[2]...
ÚLTIMO	43	0.04	último[8] última[2] últimas[12] últimos[21]
ALCANZAR	42	0.04	alcanzar[14] alcanzó[3] alcanzando[1] alcanzan[7] alcanza[9] alcanzaba[1] alcanzará[1] alcanzarlos[1] alcanzaron[4]...
CONCENTRACIÓN	42	0.04	concentración[35] concentraciones[7]
CONSIDERAR	42	0.04	considerar[3] considera[12] consideran[8] consideras[5] considerarse[2] consideramos[1] consideráis[2] considerara[1]...
KILÓMETRO	42	0.04	kilómetro[3] kilómetros[11] km[28]
PRECIO	42	0.04	precio[29] precios[13]
SUPERFICIE	42	0.04	superficie[36] superficies[6]
SUPONER	42	0.04	suponer[3] suponen[8] suponga[2] supongan[2] suponía[1] suponían[1] supone[19] supondrían[1] supondrá[1] supuso[4]
VER	42	0.04	ver[10] ve[6] vea[1] veamos[3] veían[1] veis[1] vemos[3] ven[11] verá[1] verás[2] verse[2] viendo[1]
DESARROLLAR	41	0.04	desarrollar[7] desarrollan[10] desarrollando[2] desarrollarán[1] desarrollal[10] desarrollaron[2] desarrollarse[4]...
DISTINTA	41	0.04	distinta[3] distintas[15] distintos[23]
EBRO	41	0.04	
HOY	41	0.04	
MIGRACIÓN	41	0.04	migración[7] migraciones[34]
PROVINCIA	41	0.04	provincia[13] provincias[28]
SECUNDARIO	41	0.04	secundario[29] secundarias[5] secundaria[3] secundarios[4]
SERVIR	41	0.04	servir[5] servirá[1] serviros[1] sirva[1] sirvan[1] sirve[16] sirven[14] sirvieron[1] sirvió[1]
COMÚN	40	0.04	común[31] comunes[9]
ELEMENTO	40	0.04	elemento[3] elementos[37]
LLANURA	40	0.04	llanura[13] llanuras[27]
PARO	40	0.04	
PUNTO	40	0.04	punto[18] puntos[22]
TANTO (adv.corr.)	40	0.04	(tanto ... como)
TÉCNICA (s.abs.)	40	0.04	técnica[6] técnicas[34]
TRADICIONAL	40	0.04	tradicional[17] tradicionales[23]
UNIDAD	40	0.04	unidad[14] unidades[26]
AMBIENTAL	39	0.03	ambiental[20] ambientales[19]
BALANZA	39	0.03	
DENSIDAD	39	0.03	densidad[33] densidades[6]
DERECHO (s.jur.)	39	0.03	derecho [20] derechos [19]
INTERNACIONAL	39	0.03	internacional[21] internacionales[18]
RIQUEZA	39	0.03	
SUELO	39	0.03	suelo[24] suelos[15]
TEMA	39	0.03	tema[36] temas[3]

ENTRADA	FREC.	%	LEMATIZACIÓN
TRAMA	39	0.03	trama[37] tramas[2]
VERANO	39	0.03	verano[30] veranos[9]
APARECER	38	0.03	aparecer[1] aparecen[21] aparece[14] aparecieron[1] aparezcan[1]
ASIA	38	0.03	
CAPITAL (s.)	38	0.03	capital [19] capitales [19]
EMBARGO	38	0.03	
ÉSTE	38	0.03	éste [13] ésta[10] éstas[3] éstos[12]
EURO	38	0.03	euro[12] euros[26]
GENERAR	38	0.03	generar[5] generándose[1] generan[15] generen[1] generamos[1] genera[15]
MEDIANTE	38	0.03	
PLANO (s.)	38	0.03	plano[33] planos[5]
RESIDUO	38	0.03	residuo[1] residuos[37]
CARACTERIZA	37	0.03	caracteriza[23] caracterizan[13] caracterizara[1]
CRECER	37	0.03	crecer[7] crecen[12] crece[12] crecía[1] creciendo[4] crezca[1]
DESIGUALDAD	37	0.03	desigualdad[4] desigualdades[33]
ESPERANZA	37	0.03	esperanza[36] esperanzas[1]
FAMILIA	37	0.03	familia[23] familias[14]
INTERCAMBIO	37	0.03	intercambio[9] intercambios[28]
PÚBLICO (adj.)	37	0.03	público[13] pública[8] públicas[7] públicos[9]
CONTAR (con)	36	0.03	contar[2] cuenta [18] contar[1] cuentan[15]
EDAD	36	0.03	edad[28] edades[8]
ESPECIE	36	0.03	especie[3] especies[33]
MADRID	36	0.03	
MARINA	36	0.03	marina[8] marinas[10] marino[4] marinos[14]
TRATAR	36	0.03	tratar[5] trata[24] tratad[1] tratando[1] tratan[2] tratasen[2] trató[1]
AUMENTO	35	0.03	
CONTAMINACIÓN	35	0.03	
ESCASO	35	0.03	escaso[5] escasas[12] escasa[11] escasos[7]
EVOLUCIÓN	35	0.03	
EXTERIOR	35	0.03	exterior[29] exteriores[6]
FÍSICO	35	0.03	físico[17] física[1] físicas[3] físicos[14]
IBÉRICO	35	0.03	ibérico[10] ibérica[25]
INVIERNO	35	0.03	invierno[21] inviernos[14]
NATURALEZA	35	0.03	
OCÉANO	35	0.03	océano[26] océanos[9]

ENTRADA	FREC.	%	LEMATIZACIÓN
ORIGINAR	35	0.03	originar[2] original[13] originando[2] originan[14] originaron[2] originarse[1] originó[1]
POBLADO (adj.)	35	0.03	poblado[8] poblada[2] poblados[10] pobladas[15]
PUES	35	0.03	
AFECTAR	34	0.03	afectar[2] afectan[7] afectando[2] afectal[19] afectarían[1] afectó[3]
CANARIAS	34	0.03	
CAPACIDAD	34	0.03	capacidad[33] capacidades[1]
CLASIFICAR	34	0.03	clasificar[9] clasifican[11] clasifica[11] clasificantas[1] clasificarse[2]