

COHESIÓN Y VARIEDAD LÉXICA EN LOS *ABSTRACTS* CIENTÍFICO-TÉCNICOS

MARÍA LAURA ESTEBAN SEGURA
Universidad de Málaga

RESUMEN. En la presente comunicación se estudian los *abstracts* de revistas científico-técnicas del campo del hormigón. El estudio abarca las diferencias léxicas entre los citados *abstracts* y el cuerpo de los correspondientes artículos. Asimismo, se analizan los fenómenos de cohesión que se materializan en las estructuras plurilexémicas de las frases nominales. Para su realización, se ha compilado un corpus electrónico con revistas específicas y posteriormente se han analizado los datos tanto cuantitativos como cualitativos.

PALABRAS CLAVE. *Abstract*, cohesión, densidad léxica, lema, construcción plurilexémica.

ABSTRACT. The present paper looks at abstracts in scientific and technical journals, particularly in the specific field of cement and concrete materials. The study takes into consideration lexical differences between the abstracts and the text of the articles. At the same time, we analyse cohesion mechanisms which can be observed in the multi-lexemic structures of noun phrases. In order to carry out this, an electronic corpus containing specific journals has been compiled, and subsequently the data have been analysed from a quantitative and qualitative perspective.

KEY WORDS. *Abstract*, cohesion, lexical density, lemma, multi-lexemic construction.

1. INTRODUCCIÓN

El artículo de investigación juega un papel fundamental a la hora de transmitir nuevas ideas y conocimientos a la comunidad científica. Asociado a éste, se encuentra el *abstract*, que es una breve descripción o resumen del artículo en cuestión, cuya función es presentar las ideas fundamentales del mismo. Nuestro estudio analiza artículos científico-técnicos relacionados con la temática del hormigón para establecer las posibles diferencias léxicas entre los *abstracts* y el cuerpo de los correspondientes artículos. Asimismo, se analiza el mecanismo de «cohesión» léxica en las frases nominales plurilexémicas. Para su realización, se ha compilado un corpus electrónico con revistas específicas y posteriormente se han analizado los datos.

2. METODOLOGÍA

Para llevar a cabo la compilación del corpus, se recurrió a la web www.sciencedirect.com, especializada en revistas y artículos científico-técnicos. De todas las revistas dedicadas a la ingeniería, nos centramos en las correspondientes a la ingeniería

civil, y se delimitó el campo de estudio al caso concreto del hormigón, por lo que se seleccionaron las revistas *Cement and Concrete Composites (CCC)* y *Cement and Concrete Research (CCR)*. La información que se recoge en pantalla es la siguiente: las revistas se presentan en una serie de volúmenes, en los que están repartidos los distintos números de la misma. Junto a ellos aparece el número de páginas que componen los artículos en cuestión y la fecha de publicación. De dichas revistas se extrajeron los artículos para generar la base de datos con los términos léxicos relativos a nuestro campo. De *CCC* se tomaron los volúmenes 21 a 25, correspondientes a los años comprendidos entre 1999 y 2003; de *CCR*, los volúmenes 32 y 33, correspondientes a los años 2002 y 2003 respectivamente. Los artículos que se descargaron de Internet hicieron un total de 467. La biblioteca digital *ScienceDirect* permite distintas opciones a la hora de acceder a los artículos: se puede consultar el *abstract*, el *abstract* junto con las notas, o bien el artículo en *PDF*. También da la opción de consultar un resumen junto con el esquema general del artículo o el texto completo junto con los enlaces correspondientes. No todas estas posibilidades están disponibles para todos los artículos; en algunos de ellos sólo es posible la opción del texto en *PDF*; en otros, únicamente se puede acceder al *abstract*, al *abstract* y notas, etc.

Como norma general, los artículos siguen unos patrones establecidos que incluyen:

- a) Nombre de la revista, volumen, fecha y número de páginas. Ejemplo: *Cement & Concrete Composites* 23 (2001) 3-19; b) título del artículo; c) autor o autores del mismo. Normalmente se indica el autor al que se le puede enviar correspondencia con un asterisco al lado de su nombre, y a pie de página, la dirección del mismo; d) lugar (departamento, centro, etc., universidad, dirección completa); e) fecha de recepción y de aceptación; f) *abstract*, donde aparece un breve resumen de lo que se va a tratar en el artículo. Al final, se le añade las palabras clave (*keywords*), que son palabras de carácter técnico que utiliza el autor o autores; g) introducción; h) desarrollo, que puede constar de distintos apartados; i) conclusiones; j) agradecimientos/ reconocimientos; k) notas.

Algunos de estos datos, por ejemplo, el referente al lugar, podrían servir para estudios dialectales o de variedades de inglés (americano, australiano, británico, etc.). De igual modo, el esquema general, que está disponible en muchos artículos, permitiría analizar si todos los artículos siguen el mismo patrón.

De todos los datos, hemos seleccionado para nuestro estudio exclusivamente los *abstracts* y el cuerpo de los artículos, por lo que el resto de la información se ha desechado por el momento. Asimismo, se eliminaron gráficos, tablas, fórmulas, etc. Con esta información, se crearon dos *corpora* textuales: por un lado, los *abstracts* y, por el otro, el cuerpo de los artículos. Se obtuvieron los correspondientes listados de frecuencias mediante el programa informático *Wordsmith Tools 3.0* (M. SCOTT 1998), herramienta para manejar *corpora* electrónicos y que permite extraer listas de palabras y de concordancias. Una vez que se obtuvieron las listas, se llevó a cabo su jerarquización por clases de palabra: sustantivos, adjetivos, verbos, etc. No podemos omitir que en algunas ocasiones la palabra puede pertenecer a más de una clase, por lo que la hemos dejado en la de más frecuencia (por ejemplo, en el caso de *test*, que puede ser sustantivo y verbo, se ha incluido en la lista de verbos porque ahí se da con más frecuencia). Luego, se procedió al análisis propiamente dicho en el que se examinan las palabras más comunes y las construcciones nominales plurilexemáticas.

3. CONCEPTO DE COHESIÓN

Antes de entrar a definir el concepto de cohesión, nos parece importante ver el marco en el que ésta actúa, que no es otro que el texto. Así, el texto es una unidad lingüística difícil de delimitar porque no se trata de la suma de las oraciones que lo componen, sino que ha de entenderse en su conjunto y compuesto por un haz de relaciones y mutua dependencia entre sus partes (M. Díez Prados 2003). Según M.A.K. Halliday y R. Hasan (1976), el concepto de textura expresa de forma apropiada lo que es un texto distinguiéndolo de lo que no lo es. Para ellos, el requisito más importante entre oraciones es la cohesión, ya que los lazos cohesivos entre las mismas destacan más claramente debido a que son la única fuente de textura. Según W. Grabe y R.B. Kaplan (1996), la cohesión es un fenómeno que actúa en la estructura superficial del texto y que marca las relaciones que existen entre unidades oracionales o clausales en éste. Por su parte, R.A. Beaugrande y W.U. Dressler (1981) conciben la cohesión como un fenómeno de naturaleza textual, que explica las relaciones gramaticales que se establecen entre los componentes del texto en su estructura superficial.

En líneas generales, podemos decir que la cohesión es un rasgo semántico que puede expresar el significado mediante la gramática (cohesión gramatical) o el vocabulario (cohesión léxica). Aunque la cohesión se estudia a nivel oracional, en nuestro análisis nos hemos centrado en el nivel sintagmático por lo que cabría hablar mejor de densidad léxica, en la que se recogen las palabras de contenido y las palabras funcionales o estructurales. Por ello, se muestra más cohesión, según nuestro criterio, cuando las palabras tengan entre sí el menor número de enlaces posible. Por otra parte, la extensión de las frases nominales tiene unos límites que implican la utilización de verbos y/o preposiciones (D. Biber y V. Clark 2002).

4. ANÁLISIS

El estudio tiene una doble vertiente. Por una parte, se analizan las palabras más frecuentes en los *abstracts* y cuerpo de los artículos para ver posibles coincidencias, uso de vocabulario técnico, etc. Por la otra, para concretar nuestro estudio, se han analizado las construcciones plurilexemáticas que aparecen en el *abstract*, título y cuerpo del artículo, a fin de determinar dónde se da un mayor grado de cohesión.

4.1. FRECUENCIAS

Con este estudio se pretende observar, por un lado, los matices en la distribución de las palabras más frecuentes según aparezcan en el *abstract* o cuerpo del artículo¹ y, por el otro, determinar si dichas palabras se encuadran dentro del inventario del vocabulario general o técnico.

4.1.1. *Palabras y lemas en los abstracts y artículos.* En la tabla 1, se muestra la lista de palabras de los *abstracts* con sus respectivas frecuencias. Se han incluido las 25 palabras más comunes, ordenadas de forma decreciente.

¹ El corpus del *abstract* abarca unas 70.000 palabras y el del cuerpo de los artículos más de 1.500.000 de palabras.

Tabla 1

PALABRAS	FRECUENCIA
the	5.685
of	3.956
and	2.361
in	1.630
to	1.486
a ²	1.053
concrete	880
is	795
with	733
cement	651
was	552
on	522
by	520
that	516
were	515
this	498
as	444
strength	394
are	384
from	296
be	283
results	278
an	231
or	214
water	214

Tabla 2

LEMAS	FRECUENCIA
the	126.353
of	76.609
and	46.326
in	37.439
to	30.274
be	27.757
for	16.166
with	14.064
concrete	13.917
by	10.706
as	10.110
at	9.629
cement	9.158
on	8.743
this	8.509
use	7.429
from	7.057
have	6.586
warm	6.471
test	5.508
it	5.428
show	5.061
high	4.775
increase	4.529
an	4.250

Como se puede observar, las 25 palabras más usadas son estructurales o funcionales (determinantes, preposiciones, conjunciones y el verbo auxiliar *to be*). También aparecen palabras del lenguaje general (*strength*, *water* y *results*) y terminología técnica específica (*concrete* o *cement*).

En la tabla 2, se recogen los 25 lemas (entendiendo por *lema* cada una de las palabras o términos que se definen o traducen en un diccionario o enciclopedia) más frecuentes que aparecen en el cuerpo de los artículos, con sus correspondientes frecuencias. Según esto, bajo el lema 'have', se encuentran recogidas las cinco formas del verbo (infinitivo 'have', forma de 3ª persona del singular 'has', pasado 'had', participio pasado 'had', gerundio 'having'). En el caso de los sustantivos, estos podrán aparecer en singular o plural y, en el de los adjetivos, en sus distintos grados (positivo, comparativo y superlativo).

² En algunos casos *a* corresponde a la inicial de un tipo de ión o de un compuesto.

El caso de los lemas más frecuentes es similar al de la lista de palabras anterior: predominan las palabras funcionales o estructurales, como determinantes, preposiciones y conjunciones, aunque también aparecen términos del lenguaje general o palabras léxicas como *warm*, *test*, *increase*, etc. Asimismo, se encuentran vocablos específicos: *concrete*, *cement*, etc. Si se comparan las listas de palabras de los *abstracts* y de lemas del cuerpo de los artículos, se observa que las diez primeras palabras son, más o menos, las mismas.

4.1.2. *Sustantivos*. A continuación se exponen los lemas de los 25 sustantivos que aparecen con mayor frecuencia en los artículos, ordenados de forma descendente según su frecuencia de aparición en el texto.

Tabla 3

LEMAS	FRECUENCIA	LEMAS	FRECUENCIA
concrete	13.917	matrix	1.233
cement	9.158	zone	1.100
surface	3.255	bond	1.068
crack	3.019	weight	1.064
ratio	2.945	calcium	1.052
chloride	2.873	formation	1.052
steel	2.444	peak	1.037
hydration	2.375	distribution	1.036
sulfate	2.102	section	1.022
type	2.056	maximum	998
resistance	1.696	blower	928
fragment	1.399	reference	924
curve	1.371		

Aquí, se puede observar que predominan los términos específicos al ámbito técnico: *concrete*, *cement*, *chloride*, *sulfate*, etc., aunque también se encuentran términos del lenguaje más general: *ratio*, *distribution*, *section*, etc.

4.1.3. *Verbos*. De forma similar al caso de los sustantivos, en la tabla 4 se muestran los lemas de los verbos, ordenados por frecuencia de aparición en orden descendente y, a su lado, el número de apariciones en el texto.

Tabla 4

LEMAS	FRECUENCIA	LEMAS	FRECUENCIA
be	48.803	cure	2.495
use	7.429	obtain	2.315
have	6.586	load	2.295
test	5.508	study	2.144
show	5.061	give	1.851
increase	4.529	observe	1.783
mix	3.735	stress	1.782
aggregate	3.355	phase	1.765
effect	3.355	contain	1.750
sample	3.341	decrease	1.654
content	3.316	make	1.592
crack	3.019	amount	1.061
fly	2.707		

A tenor de lo anterior, se da un equilibrio entre verbos del lenguaje general (*be, use, contain, decrease, etc.*) y semi-técnicos, como es el caso de *cure*, que es una palabra del lenguaje general que se utiliza a nivel técnico.

4.1.4. *Adjetivos*. Al igual que en el caso de los sustantivos y de los verbos, en la tabla 5 se muestran los lemas de los 25 adjetivos más comunes en el texto, en orden decreciente, junto con el número de apariciones en el texto.

Tabla 5

LEMAS	FRECUENCIA	LEMAS	FRECUENCIA
warm	6.471	pozzolanic	770
high	4.775	great	687
low	2.993	new	647
temperate	2.419	late	615
experimental	1.797	physical	607
past	1.425	flexural	589
small	1.326	even	581
large	1.246	normal	575
dry	1.229	possible	555
total	1.092	wet	472
early	899	bilinear	448
less	870	steep	427
long	812		

Aquí predominan los adjetivos del inventario general (*warm, small, long, etc.*), aunque también hay presencia de adjetivos específicos al campo técnico (*pozzolanic, flexural o bilinear*).

4.1.5. *Palabras estructurales*. Por último, se muestran las frecuencias de las palabras estructurales: preposiciones, conjunciones, pronombres/determinantes y adverbios. Se han seleccionado los cinco más frecuentes, que aparecen en orden descendente, junto con el número de apariciones.

Tabla 6

PREPOSICIONES	FRECUENCIA	CONJUNCIONES	FRECUENCIA
of	76.609	and	46.326
in	37.439	as	10.110
to	30.274	or	3.867
for	16.166	than	3.515
with	14.064	when	2.226
PRONOM./DETER.	FRECUENCIA	ADVERBIOS	FRECUENCIA
the	126.353	not	3.305
this	8.509	also	2.875
it	5.428	however	1.958
which	4.137	only	1.625
these	3.041	where	1.517

4.2. CONSTRUCCIONES PLURILEXEMÁTICAS

Este tipo de construcciones aprovecha la potencialidad sintética de la lengua inglesa y condensa la máxima información de la forma más sucinta posible (B.L. DUBOIS 1985). Este autor afirma que este tipo de estructuras provoca la pérdida de dinamismo e información explícita y, como consecuencia, exige una mayor formación del lector respecto al área de especialidad. Otro punto de vista diferente es el de J.N. LEVI (1978), quien establece que los sintagmas nominales complejos se emplean preferentemente cuando el autor persigue la concisión, presentándose la información de forma condensada y, por ello, con un mayor efecto en el lector.

En nuestro corpus, las unidades plurilexemáticas predominan fundamentalmente en los *abstracts* y en los títulos, que son aquellas partes del artículo de investigación donde la concisión es un factor imperante debido a razones de espacio.

El componente obligatorio de este tipo de estructuras es el núcleo que, en función de su complejidad, puede ir acompañado de una premodificación y una postmodificación. Nosotros nos centraremos en la premodificación, ya que la postmodificación se realiza mediante otro tipo de mecanismos como, por ejemplo, los sintagmas preposicionales. Se ha utilizado la siguiente notación (J. CALLE y A. SÁEZ 2004):

I = intensificador

A = adjetivo, participio

N = nombre

x = 1, 2, 3, 4, etc.

que se distribuye secuencialmente según la siguiente fórmula:

$$\{x[\mathbf{I}] x[\mathbf{A}]\} x[\mathbf{N}] \mathbf{N}$$

En ella, **N** representa al nombre que constituye el núcleo (único elemento imprescindible), *x* al número de elementos posible de cada clase y el corchete indica que puede existir o no esa clase. Por su parte, la llave denota un mayor grado de cohesión entre los elementos que se encuentran dentro.

4.2.1. *Términos billexemáticos*. Son aquellos en los que el nombre o sustantivo puede aparecer premodificado por otro nombre o por un adjetivo. Veáanse las siguientes construcciones:

a) N N:

Cement *paste* «Pasta de cemento»

Chloride *diffusion* «Difusión de cloruro»

Lithium *compounds* «Compuestos de litio»

Metakaolin *concrete* «Hormigón con metacaolín»

Moisture *content* «Contenido de humedad»

b) A N:

Combined *water* «Agua combinada»

Compressive *strength* «Resistencia a compresión»

Expanded *clay* «Arcilla expandida»

Mineral *additions* «Aditivos minerales»

Pozzolanic *effect* «Efecto pozoalánico»

Muchas de estas construcciones acaban lexicalizándose, pasando a constituir entradas en los diccionarios especializados.

4.2.2. *Términos plurillexemáticos*. Están compuestos por el núcleo y por una premodificación que puede variar en longitud y que está formada básicamente por nombres y adjetivos. Los sintagmas que componen estas estructuras pueden mostrar distintos niveles de cohesión entre sí. La clasificación que se da es la siguiente:

a) {xN} N (x = 2)

Corrosion protection *effectiveness* «Eficacia de la protección a la corrosión»

X-ray absorption *measurement* «Medición por absorción de rayos X»

Force/displacement *relationship* «Relación Fuerza/Desplazamiento»

b) {xN} N (x = 3)

Mercury intrusion porosimetry *technique* «Técnica de porosimetría por intrusión de mercurio»

Strain gauge (SG) measurement *technique* «Técnica de medición de tensión»

Calcium aluminate cement *paste* «Pasta de cemento de aluminato de calcio»

c) {I A} A N

Fast-dry reinforced *concrete* «Hormigón reforzado de secado rápido»

Highly-developed curing *methods* «Métodos de fraguado muy desarrollados»

d) {xA} N (x = 2)

High pozzolanic *effect* «Alto efecto puzolánico»

Nondestructive rapid *test* «Ensayo rápido no destructivo»

Residual compressive *strength* «Fuerza residual compresiva»

e) {N A} N

Styrene acrylic *emulsion* «Emulsión de estireno acrílico»

Cement-based *materials* «Materiales compuestos de cemento»

Chloride-induced *corrosion* «Corrosión debida al ión cloro»

f) {A N} N

Maximum aggregate *size* «Tamaño máximo de aditivo»

Reinforced concrete *beam* «Vigueta de hormigón reforzado»

Smooth surface *finish* «Acabado de superficie suave»

g) {xA xN} N (x = 2)

Pure ordinary Portland cement (OPC) *concretes* «Cemento Portland»

5. CONCLUSIONES

En este apartado, se recogen las principales conclusiones a las que se ha llegado con nuestro estudio. En primer lugar, se ha constatado que existe una diferencia significativa entre el léxico del cuerpo del artículo y el de los *abstracts*. De este modo, donde mayor será el número de construcciones plurilexémicas es en los títulos, ya que estos han de resumir las ideas principales en el menor espacio posible, por lo que se usan estas estructuras, que se unen mediante conjunciones. Asimismo, aparecerán más construcciones de este tipo en los *abstracts* que en el cuerpo del artículo. En los *abstracts*, la relación de palabras estructurales es menor, por lo que hay más construcciones plurilexémicas y mayor densidad léxica. El artículo, por su parte, hace uso de una diversidad de recursos como, por ejemplo, el empleo de palabras estructurales. Si jerarquizamos la cohesión o densidad léxica de mayor a menor grado, el primer lugar lo ocuparía el título, seguido del *abstract* y, por último, tendríamos el cuerpo del artículo.

En un futuro estudio, pretendemos analizar el léxico perteneciente a otros materiales como, por ejemplo, el acero, para ver si se utilizan de igual modo las construcciones plurilexémicas, así como las mismas preposiciones, adjetivos, etc.

BIBLIOGRAFÍA

- BEAUGRANDE, R.A. DE y DRESSLER, W.U. (1981): *Introduction to Text Linguistics*, Londres, Longman.
- BIBER, DOUGLAS y CLARK, VICTORIA (2002): «Historical shifts in modification patterns with complex noun phrase structures: How long can you go without a verb?», en FANEGO, T., LÓPEZ-COUSO, M. y PÉREZ-GUERRA, J. (ed.): *English Historical Syntax and Morphology*, Ámsterdam, John Benjamins Publishing Company.
- CALLE MARTÍN, JAVIER y SÁEZ MELENDO, ANA BELÉN (2004): «Problemas de traducción de los sintagmas nominales complejos en el inglés científico-técnico», en ORTEJA ARJONILLA, EMILIO (dir.): *Panorama actual de la investigación en traducción e interpretación*, Granada, Atrio.
- DÍEZ PRADOS, MERCEDES (2003): *Coherencia y cohesión en textos escritos en inglés por alumnos de filología inglesa (estudio empírico)*, Alcalá de Henares, Servicio de publicaciones de la Universidad Alcalá de Henares.
- DUBOIS, B.L. (1985): «The Construction of Noun Phrases in Biomedical Slide Talks», en CORNU, A.M. et al. (eds.): *Beads or Bracelet? How Do We Approach LSP*, Oxford, Oxford University Press.
- FERNÁNDEZ POLO, FRANCISCO JAVIER (1999): *Traducción y retórica contrastiva. A propósito de la traducción de textos de divulgación científica del inglés al español*, Santiago de Compostela, Servicio de publicaciones de la Universidad de Santiago de Compostela.
- GÓMEZ MORÓN, REYES (2004): «La traducción de aspectos pragmáticos en los abstracts de artículos de investigación: breve estudio contrastivo inglés-español», en ORTEJA ARJONILLA, EMILIO (dir.): *Panorama actual de la investigación en traducción e interpretación*, Granada, Atrio.
- GRABE, WILLIAM y KAPLAN ROBERT B. (1996): *Theory and Practice of Writing*, Londres, Longman.
- HALLIDAY, MICHAEL ALEXANDER K. Y HASAN, RUQAIYA (1976): *Cohesion in English*, Londres, Longman.
- LEVI, JUDITH N. (1978): *The Syntax and Semantics of Complex Nominals*, Nueva York, Academic Press.
- SCOTT, MIKE (1996): *Wordsmith Tools 3.0*, Oxford, Oxford University Press.