

## EL LÉXICO COMO ELEMENTO ESTRUCTURADOR DE LA INFORMACIÓN SEMÁNTICA EN EL DISCURSO CIENTÍFICO

M<sup>a</sup> ÁNGELES GARCÍA DE SOLA  
Universidad de Granada

**RESUMEN.** *Dentro del contexto de IFE, el presente trabajo aplica el modelo de patrones léxicos propuesto por Michael Hoey en su libro *Patterns of Lexis in Text* (1991) al texto científico. Dicho modelo, diseñado para analizar la función organizativa del léxico en los textos no narrativos, se fundamenta en la construcción de matrices de reiteración léxica y demuestra que la presencia de interrelaciones cohesivas entre oraciones próximas y no próximas evidencia la organización de la información semántica del texto. En este sentido, el objetivo de nuestra investigación es observar si la cohesión léxica en la introducción del artículo de investigación permite la identificación de cuatro tipos de oraciones: central, marginal, principal y concluyente y si dichas oraciones posibilitan la confección de resúmenes que contengan la información fundamental del texto original.*

**PALABRAS CLAVE:** *léxico, cohesión, relaciones semánticas, tipos de oraciones, resumen.*

**ABSTRACT.** *Within the context of ESP, the present paper reports an application of the method of analysis proposed by Michael Hoey in his 1991 book *Patterns of Lexis in Text* to scientific texts. This model, designed to analyze the text-organising function of lexis in non-narrative text, is based on the construction of lexical repetition matrices and shows that the presence of cohesive relationships between adjacent and non-adjacent sentences signals the underlying semantic information of the written text. In this respect, our main aim is to investigate if lexical cohesion in research article introductions allows us to identify four types of sentences: central, marginal, topic-opening and topic-closing and if we can use them to create summaries of the content of the texts.*

**KEYWORDS:** *lexis, cohesion, semantic relations, types of sentences, summary.*

### 1. INTRODUCCIÓN

El estudio y análisis del léxico se está convirtiendo en uno de los focos principales de atención tanto de la lingüística, en su vertiente teórica, como de la lingüística aplica-

da. En esta misma línea, la tendencia actual dentro del área del Inglés para Fines Específicos se está dirigiendo a la elaboración de *corpus* específicos que permitan una descripción más detallada y fiable del léxico en los diferentes registros de especialidad. De esta forma, el análisis de los patrones léxicos que aparecen con una determinada frecuencia en el *corpus* va a poner de manifiesto el comportamiento de la lengua de especialidad y a mostrar el uso que hacen los propios especialistas de la terminología y la fraseología especializada de cada disciplina académica. Las aportaciones de estas investigaciones, no cabe duda, están siendo muy valiosas tanto desde el punto de vista teórico como práctico. De modo general, cabría hacer referencia a algunos de los investigadores que han elaborado *corpus* a partir de dominios específicos con la finalidad de obtener diseños de programas para IFE. Por ejemplo, Flowerdew (1993) utilizó la frecuencia de palabras y estructuras lingüísticas de una serie de materiales de Biología para decidir sobre el inglés que habría que enseñar a los estudiantes de ciencias y encontró que eran muy diferentes a las de un *corpus* general extenso. En España merece destacarse el trabajo de Curado Fuentes (2001), que analiza los comportamientos léxicos en varios campos de especialidad, y aporta interesantes aplicaciones metodológicas para su utilización en el aula y el de Gil Salom, Soler Monreal y Stuart (2001), que investigan las características del inglés científico-técnico. La elaboración de *corpus* a partir de materias específicas con tipos de texto similares abre, pues, muchas vías para el análisis del discurso, análisis de género, pragmática y sociolingüística.

Sin embargo, el léxico, desde la perspectiva de su función cohesiva en la estructuración del contenido del texto científico, no ha sido suficientemente analizado con el objeto de observar si éste organiza el texto mediante la creación de redes de repetición y si dicha repetición es la que nos hace percibir un texto como coherente. Es bien sabido que no existe un acuerdo general entre los lingüistas sobre si coherencia y cohesión constituyen dos conceptos independientes o estrechamente relacionados. Así, mientras Widdowson (1978), de Beaugrande y Dressler (1981) y Brown y Yule (1983) piensan que coherencia y cohesión constituyen dos aspectos separados, Hoey (1991) —en la línea de Winter (1979), Hasan (1984, 1985) y Phillips (1985)— opina que ambos aspectos están íntimamente relacionados. En este sentido, el objetivo del presente trabajo de investigación es observar si la cohesión léxica está relacionada con la coherencia, si es la presencia de repeticiones de unidades léxicas la que permite que un lector sea capaz de interpretar las proposiciones de una forma apropiada y captar la coherencia del texto. En concreto, esta investigación pretende verificar si es posible confeccionar resúmenes que contengan la información fundamental del texto original mediante cuatro formas:

- 1) Eliminando las oraciones marginales, entendiendo por éstas las que no establecen repeticiones léxicas con el resto de oraciones del texto y considerándolas, por lo tanto, no relevantes para el contenido fundamental del texto.
- 2) Combinando las oraciones centrales, entendiendo por éstas las que establecen más repeticiones léxicas y, a diferencia de las oraciones marginales, se les considera que son las que portan la información fundamental del texto.

- 3) Partiendo de la oración principal, entendiendo por ésta la que establece más repeticiones léxicas con las oraciones posteriores, y uniéndola con aquellas oraciones con las que establece conexiones.
- 4) Partiendo de la oración concluyente, entendiendo por ésta la que establece más repeticiones con las oraciones anteriores, y uniéndola con aquellas oraciones con las que establece conexiones.

Para ello, aplicaremos el modelo de análisis de patrones léxicos propuesto por Hoey (1991) en su libro *Patterns of Lexis in text*. Dicho modelo se fundamenta en la construcción de matrices de reiteración léxica, y demuestra que la presencia de redes de interrelaciones cohesivas evidencia la organización de la información semántica entre oraciones próximas y no próximas en un texto. Esta investigación, pues, no se va a limitar a la mera detección y contabilización de una serie de elementos léxicos presentes en el texto científico, sino que va más allá, pretenderá examinar cómo éstos se integran en las convenciones lingüísticas y socio-culturales del mismo. Las posibilidades prácticas desde el punto de vista metodológico y del análisis del discurso pueden ser valiosas y variadas y nos permitirá observar de una forma más precisa el comportamiento del léxico en el discurso científico.

## 2. DESCRIPCIÓN DEL *CORPUS*

El *corpus* que conforma este trabajo consta de veinte introducciones pertenecientes al artículo de investigación extraído de revistas especializadas en el campo de Química Analítica, escritos por especialistas y dirigidos a investigadores en la disciplina, cuyo objetivo comunicativo es trasladar a los foros científicos nacionales e internacionales los resultados de investigaciones puntuales. Todos ellos contienen la macroestructura *IMRD* (*introduction, methods, results y discussion*), establecida por Swales (1990) para diferenciar los artículos de investigación de otros géneros académicos y validada por posteriores investigaciones, como la efectuada por Lewin y Fine (1996) en el área de psicología y sociología, la realizada por Holmes (1997) sobre historia, política y sociología, la de Nwogu (1997) sobre las ciencias médicas o la de Posteguillo (1999) sobre los artículos de investigación en informática.

Debido a las características del *corpus* seleccionado y al tipo de análisis que pretendemos llevar a cabo, hemos recabado la ayuda de un especialista en Química Analítica, ya que sólo un lector capaz de comprender el contenido científico de los textos así como la lengua inglesa podrá relacionar las ideas nuevas aportadas por los textos con su conocimiento previo en el campo científico de su especialidad y captar la coherencia de los mismos. Dicho especialista realizará la lectura de los resúmenes resultantes de las introducciones seleccionadas para su análisis y confirmará si éstos contienen la información fundamental del texto original, dando así validez a nuestros resultados.

## 3. METODOLOGÍA

La mayoría de los recursos cohesivos que se cuentan como enlaces (*links*) son léxicos: repetición simple y compleja, paráfrasis simple y compleja, hiponimia y co-referencia. La repetición no léxica incluye deixis, sustitución y elipsis. Las siguientes tablas muestran los distintos tipos de enlaces junto con ejemplos extraídos del *corpus* analizado con las oraciones numeradas y los enlaces enmarcados para facilitar su identificación. (Para una exposición detallada y minuciosa de los distintos tipos de enlaces ver García de Sola 2002).

| TIPOS DE ENLACES LÉXICOS  | EJEMPLIFICACIÓN  |
|---|--|
| Repetición simple (rs): Dos unidades léxicas idénticas ( <i>yeast / yeast, oligosaccharides / oligosaccharides, glucose / glucose</i> ) o dos unidades similares cuya diferencia sea debida al funcionamiento gramatical de la lengua ( <i>ferment / fermented</i> ). | <p>10. <u>Yeast</u> can <u>ferment</u> a wide variety of sugars and <u>oligosaccharides</u> other than <u>glucose</u>.</p> <p>11. The D-hexoses and <u>oligosaccharides</u> <u>fermented</u> most often by <u>yeast</u> are <u>glucose</u>, <u>mannose</u>, <u>fructose</u>, <u>galactose</u>, <u>maltose</u>, <u>lactose</u>, <u>melibiose</u>, <u>trehalose</u> and <u>raffinose</u>.</p> <p>(del texto: <i>Bioconversion of solid food wastes to ethanol</i>).</p>                                  |
| Repetición compleja (rc): Dos unidades léxicas similares que compartan el mismo morfema, o dos unidades idénticas con diferente función gramatical ( <i>produced / products, starch / starchy</i> ).  | <p>5. Ethanol can be <u>produced</u> from many kinds of raw material that contains <u>starch</u>, sugar or cellulose.</p> <p>18. The objectives of this work were to study the effect of low - and high-temperature enzymes on hydrolysis of food wastes, to compare the fermentation of bakery products with mixed waste <u>products</u> and to study the cofermentation of cheese whey and <u>starchy</u> food wastes.</p> <p>(del texto: <i>Bioconversion of solid food wastes to ethanol</i>).</p> |

| TIPOS DE ENLACES LÉXICOS  | EJEMPLIFICACIÓN   |
|---|---|
| Paráfrasis simple (ps): Dos unidades léxicas de la misma categoría gramatical, siempre que puedan ser sustituidas sin que exista un cambio de significado ( <i>premature / pre-term, micronutrients / trace elements</i> ). | <p>6. Special attention must be paid to very low birth mass, <u>premature</u> infants because they are born with lower stores of essential <u>micronutrients</u>.</p> <p>7. <u>Trace elements</u> must be added to <u>pre-term</u> infants' formulas to satisfy their higher dietary requirements.</p> <p>(del texto <i>Speciation as an analytical aid in trace element research in infant nutrition</i>).</p> |

| TIPOS DE ENLACES LÉXICOS   | EJEMPLIFICACIÓN  |
|--|--|
| <p>Antonimia (a): Antónimos que no compartan el mismo morfema léxico (<i>higher / lower</i>).</p>  | <p><b>12.</b> <i>With the sole exception of selenium, the trace element intake of infants via formula is significantly <u>higher</u> than via breast milk.</i></p> <p><b>18.</b> <i>In spite of the significantly <u>lower</u> trace element intake of breast-fed infants, their serum concentrations of the essential elements Cu, Fe and Zn are comparable to those of formula-fed infants.</i></p> <p>(del texto <i>Speciation as an analytical aid in trace element research in infant nutrition</i>).</p> |
| <p>Paráfrasis compleja unidad intermedia sinonimia (pc): Cuando exista una unidad léxica intermedia que establezca una repetición compleja y a su vez una paráfrasis simple (<i>studied</i> sería la unidad intermedia que establece una repetición compleja con <i>studies</i> y a su vez una paráfrasis simple con <i>investigated</i>).</p> | <p><b>13.</b> <i>Several <u>studies</u> on ethanol production via fermentation and the effects of different factors on the fermentation have been published in the past decade.</i></p> <p><b>14.</b> <i>Utilization of cheese whey as the liquid portion of a fermenting corn mash has been <u>investigated</u> by Whalen et al.</i></p> <p>(del texto: <i>Bioconversion of solid food wastes to ethanol</i>).</p>  |

| TIPOS DE ENLACES LÉXICOS  | EJEMPLIFICACIÓN  |
|---|--|
| <p>Paráfrasis compleja unidad intermedia antonimia (pc): Cuando exista una unidad léxica intermedia que establezca una repetición compleja y a su vez una antonimia (<i>high</i> sería la unidad intermedia que establece una repetición compleja con <i>higher</i> y a su vez una antonimia con <i>low</i>).</p> | <p><b>7.</b> <i>In most of these reports, conventional packed columns, which give relatively <u>low</u> sensitivity and poor separation, were used.</i></p> <p><b>14.</b> <i>This means that 10-100 times <u>higher</u> sensitivity can be obtained by this method as compared with that of the previous methods.</i></p> <p>(del texto: <i>Determination of cyanide in whole blood by capillary gas chromatography with cryogenic oven trapping</i>).</p> |
| <p>Hiponimia / (hip): Cuando la unidad léxica más específica, hipónimo, precede a la más general, hiperónimo (<i>ethanol / alcohol</i>).</p>  | <p><b>4.</b> <i><u>Ethanol</u> production from agricultural products has been in practice for the past 80 years.</i></p> <p><b>7.</b> <i>Utilization of food processing wastes to produce fuel <u>alcohol</u> with an increased efficiency has been under investigation in our laboratory for the past few years.</i></p> <p>(del texto: <i>Bioconversion of solid food wastes to ethanol</i>).</p>  |

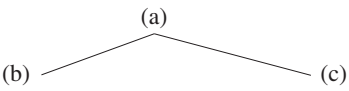
| TIPOS DE ENLACES LÉXICOS   | EJEMPLIFICACIÓN  |
|--|--|
| <p>Relación triangular (tr):</p>  <p>Si una unidad léxica (a) establece un enlace con las unidades léxicas (b) y (c), las unidades léxicas (b) y (c) establecerán igualmente una relación de repetición (<i>alcohol / ethanol</i>).</p> | <p><b>4.</b> <u>Ethanol</u> production from agricultural products has been in practice for the past 80 years.</p> <p><b>7.</b> Utilization of food processing wastes to produce fuel <u>alcohol</u> with an increased efficiency has been under investigation in our laboratory for the past few years.</p> <p><b>13.</b> Several studies on <u>ethanol</u> production via fermentation and the effects of different factors on the fermentation have been published in the past decade.</p> <p>(del texto: <i>Bioconversion of solid food wastes to ethanol</i>).</p> |

Tabla 1: Tipos de enlaces léxicos.

| TIPOS DE ENLACES NO LÉXICOS  | EJEMPLIFICACIÓN  |
|--|--|
| <p>Deixis (d):<br/>(<i>this</i> establece una relación de referencia con la oración anterior).</p>       | <p><b>3.</b> However, because of the difference in the environment of cell growth in the intact organism and the culture, great care must be taken in extrapolating the results of in vitro experiments to the reality in vivo.</p> <p><b>4.</b> <u>This</u> is particularly true for proteins, whose identity and abundance can vary greatly at different stages of cell development or expressing conditions.</p> <p>(del texto: <i>Nanoliter chemistry combined with mass</i>).</p> |
| <p>Sustitución (s):<br/>(<i>Their</i> establece una relación de referencia con <i>Whalen et al</i>).</p> | <p><b>14.</b> Utilization of cheese whey as the liquid portion of a fermenting corn mash has been investigated by <u>Whalen et al.</u></p> <p><b>15.</b> <u>Their</u> work involved the fermentation of lactose/corn mash by the use of a dual yeast inoculum (<i>Kluyveromyces marxianus</i> and distiller's yeast).</p> <p>(del texto: <i>Bioconversion of solid food wastes to ethanol</i>).</p>  |
| <p>Elipsis (e)<br/>(<i>In the field</i> constituiría la elipsis)</p>                                     | <p><b>6.</b> Currently, CO<sub>2</sub> concentrations are determined either by collecting air in flasks for analysis offsite or by continuous monitoring <u>in the field</u>.</p> <p><b>10.</b> Continuous monitoring is almost exclusively performed by NDIR.</p> <p>(del texto: <i>High-precision conductometric detector for the measurement of atmospheric carbon dioxide</i>).</p>  |

Tabla 2: Tipos de enlaces no-léxicos.

Una vez realizado el análisis exhaustivo y pormenorizado de todas las repeticiones que establecen las oraciones entre sí, éstas se presentan en forma de matriz: Una detalla las unidades que han presentado enlaces junto con el tipo de repetición (figura 1); la otra contabiliza el total de enlaces (figura 2).

A continuación presentamos las 5 primeras oraciones de la introducción del artículo de investigación *Speciation as an analytical aid in trace element research in infant nutrition* (cuadro 1) con las unidades léxicas que establecen enlaces, enmarcadas y sus matrices respectivas (figuras 1 y 2).

1. During the prenatal period, the fetus is supplied with minerals and trace elements via maternal circulation and controlled placental transfer. 2. After separation from the mother, the newborn has to develop its own functions and regulatory systems, including respiration, digestion and immune defenses. 3. Trace elements are involved in the form of metalloproteins and enzymes at all stages in the development of these processes. 4. Infancy is further characterized by an extremely high rate of synthesis of tissue cells, which leads to the infant's doubling its birth mass in a period of only 4 months. 5. The infant's trace element requirement is supplied not only by amounts transferred via the mother's milk in specific binding forms or by formula, but also from prenatal stores.

Cuadro 1: Introducción del artículo de investigación *Speciation as an analytical aid in trace element research in infant nutrition*.

|   |  |  |   |
|---|--|--|---|
| 1 |  |  |   |
| 2 | <b>pc.</b> maternal – mother ( <b>motherly</b> )   | 2  |   |
| 3 | <b>rs.</b> trace – trace<br><b>rs.</b> elements – elements   | <b>rc.</b> develop – development<br><b>d.</b> own functions ....<br>defenses – these | 3   |
| 4 |  | <b>psm.</b> newborn – infant   | 4   |
| 5 | <b>rs.</b> prenatal – prenatal<br><b>rs.</b> supplied – supplied<br><b>rs.</b> trace – trace<br><b>rs.</b> elements – element<br><b>rs.</b> via – via<br><b>pc.</b> maternal – mother ( <b>motherly</b> )<br><b>rc.</b> transfer – transferred | <b>rs.</b> mother – mother<br><b>psm.</b> newborn – infant                           | <b>rs.</b> trace – trace<br><b>rs.</b> elements – element |

Figura 1: Matriz de las 5 primeras oraciones que detalla las unidades léxicas que establecen enlaces.

Como se puede observar, la matriz de la figura 1 recoge todas los enlaces de repetición que las oraciones establecen entre sí –la oración 1 está enlazada con la 2 mediante una paráfrasis compleja entre *maternal* y *mother* (siendo la unidad intermedia

*motherly*), con la oración 3 mediante los enlaces de repetición simple de *trace* y *elements*, etc. –, mientras que la matriz de la figura 2 recoge el número de enlaces– uno, entre las oraciones 1 y 2; dos, entre las oraciones 1 y 3, etc.

|   |   |   |   |   |
|---|---|---|---|---|
|   | 1 |   |   |   |
| 2 | 1 | 2 |   |   |
| 3 | 2 | 2 | 3 |   |
| 4 | 0 | 1 | 0 | 4 |
| 5 | 7 | 2 | 2 | 1 |

Figura 2: Matriz numérica de enlaces de las 5 primeras oraciones.

Consideraremos que existe una conexión (*bond*) entre oraciones siempre que encontremos un número suficiente de enlaces entre las mismas. Dicho número no debe ser inferior a tres, debido a la gran probabilidad de que las oraciones del texto presenten dos enlaces. La razón es meramente práctica. Si reducimos este umbral, un número elevado de oraciones puede aparecer relacionado, lo cual no nos va a aportar datos relevantes sobre la cohesión. Phillips (1985) también llegó a la misma conclusión al tomar tres enlaces como mínimo para llegar a resultados significativos. De esta forma, si observamos la matriz numérica de la figura 2, sólo consideraremos que están conectadas las oraciones 1 y 5.

Existen circunstancias en que tendremos que tomar más de tres enlaces para establecer una conexión. Esto sucederá cuando encontremos que una proporción elevada de oraciones presenta tres o más enlaces. El criterio que hemos adoptado en nuestra investigación para fijar el número de enlaces necesario para considerar que existe una conexión entre oraciones ha sido calcular el tanto por ciento de los pares de oraciones que establecen como mínimo tres enlaces, que no debe ser superior a aproximadamente el 30%; si no es así, incrementaremos el número de enlaces hasta conseguir dicha proporción. Este 30% aproximado constituye el umbral que hemos establecido *a priori*, ya que se podría considerar, dentro del ámbito estadístico, como el que puede ofrecer datos más fiables. Por ejemplo, en el texto *Speciation as an analytical aid in trace element research in infant nutrition* encontramos que los pares de oraciones que presentan tres enlaces supone el 43'2% (ver tabla 3); por lo tanto, incrementamos a cuatro el número mínimo de enlaces para establecer una conexión.

| Número de enlaces | Pares de oraciones | %    |
|-------------------|--------------------|------|
| A partir de 3     | 100                | 43'2 |
| A partir de 4     | 61                 | 26'6 |
| Total de celdas   | 231                |      |

Tabla 3: Porcentaje de pares de oraciones que presentan a partir de tres o cuatro enlaces.



De la matriz numérica (figura 2) extraemos la siguiente tabla que muestra las conexiones que cada oración presenta con el resto de las oraciones del texto. El primer dato numérico entre paréntesis indica el número de conexiones que la oración establece con las anteriores; el segundo dato numérico entre paréntesis indica el número de conexiones que establece con las posteriores. Por ejemplo, la oración 5 establece 1 conexión con una oración anterior y 12 con posteriores. Los guiones que aparecen en las oraciones 1 y 22 indican que no es posible establecer conexiones con oraciones anteriores a la oración primera del texto, ni tampoco establecerlas con las oraciones posteriores a la última oración del texto. Los datos numéricos entre corchetes y en negrita contabilizan el total de conexiones de cada oración con el resto de oraciones del texto.

|                 |                |                |               |                |
|-----------------|----------------|----------------|---------------|----------------|
| 1. (-,2) [2]    | 2. (0,0) [0]   | 3. (0,0) [0]   | 4. (0,1) [1]  | 5. (1,12) [13] |
| 6. (0,0) [0]    | 7. (1,5) [6]   | 8. (2,8) [10]  | 9. (2,4) [6]  | 10. (2,5) [7]  |
| 11. (4,5) [9]   | 12. (7,6) [13] | 13. (2,1) [3]  | 14. (5,4) [9] | 15. (1,2) [3]  |
| 16. (2,2) [4]   | 17. (1,0) [1]  | 18. (9,2) [11] | 19. (1,0) [1] | 20. (8,2) [10] |
| 21. (10,1) [11] | 22. (4,-) [4]  |                |               |                |

Tabla 4: Número de conexiones que establecen el total de las oraciones del texto: *Speciation as an analytical aid in trace element research in infant nutrition.*

Este tipo de tabla nos ofrece una información fundamental para nuestra investigación, ya que muestra qué oraciones presentan un mayor número de conexiones –oraciones centrales– y cuáles establecen 0 conexiones –oraciones marginales–. Además, nos indica qué oración/es establece/n un mayor número de conexiones con las posteriores –oración principal– y cuáles establecen un mayor número de conexiones con las anteriores –oración concluyente–. En resumen, obtenemos una representación gráfica de lo que este trabajo se propone verificar.

#### 4. DISCUSIÓN DE LOS RESULTADOS

Este análisis nos ha permitido no sólo confeccionar resúmenes que contienen la información del texto original, sino también resultados interesantes con respecto a las características que han presentado las oraciones principales y concluyentes y los tipos de repetición utilizados. A continuación ofrecemos los datos más significativos teniendo en cuenta la totalidad del *corpus* analizado.

##### 4.1. RESÚMENES: CARACTERÍSTICAS

Un dato de gran interés es conocer qué porcentaje de oraciones de las introducciones analizadas constituye un resumen. Debido a la variedad de estos porcentajes, los agrupa-

mos en intervalos de 20. Así, en las siguientes tablas se agrupan los resúmenes que representan del 10% al 30%, del 30% al 50%, y del 50% al 70% de oraciones del texto original. En la tabla de los resúmenes confeccionados mediante la eliminación de las oraciones marginales, los intervalos oscilan entre el 50% y el 70%, el 70% y el 90%, y 90% y el 97%, siendo el 97% el porcentaje más elevado de oraciones que presenta un resumen.

En primer lugar, presentamos los valores absolutos y porcentajes de oraciones de los resúmenes confeccionados con las oraciones centrales, principal y concluyente:

| RESÚMENES ORACIONES CENTRALES, PRINCIPAL Y CONCLUYENTE |           |            |             |
|--|-----------|------------|-------------|
| Nº oraciones (%)                                       | Centrales | Principal  | Concluyente |
| 10-30  | 11 = 55%  | 3 = 15'7%  | 8 = 40%     |
| 30-50  | 7 = 35%   | 10 = 52'6% | 9 = 45%     |
| 50-70  | 2 = 10%   | 6 = 31'5%  | 3 = 15%     |
| Total  | 20 = 100% | 19 = 99'8% | 20 = 100%   |

Tabla 5: Valores absolutos y porcentajes de oraciones que constituyen resúmenes.

Con respecto a las oraciones centrales, la mayoría (55%) de los resúmenes confeccionados contiene entre el 10% y el 30% de las oraciones del texto original y los resúmenes que contienen más del 50% de oraciones constituyen la minoría (10%). Hoey (1991: 118) obtiene resultados más reducidos. Así, con las oraciones centrales, obtiene el 15% del texto original. En nuestro análisis, tan sólo un texto se aproxima a este porcentaje con un 16'6%.

Con respecto a la oración principal, la mayoría (52'6%) de los resúmenes contiene entre el 30% y el 50% de oraciones del texto. Hoey (1991: 128), de nuevo, con los resúmenes confeccionados con las oraciones que presentan conexiones con la oración principal, obtiene resúmenes más reducidos –concretamente, 17'5% y 22'5%–. En nuestro análisis sólo 3 textos han presentado un porcentaje no superior al 30%.

Con respecto a la oración concluyente, la mayoría (45%) de los resúmenes representa entre el 30% y 50% de oraciones del texto original. El resumen aportado por Hoey (1991: 146-147), confeccionado con las oraciones que presentan conexiones con la oración concluyente, representa el 17'5%. En las introducciones que hemos analizado todos los resúmenes han sobrepasado el 20% de oraciones del texto original.

Por último, en la siguiente tabla se ofrecen los valores absolutos y porcentajes de los resúmenes elaborados mediante la eliminación de las oraciones marginales:

| RESÚMENES ORACIONES MARGINALES |            |
|--------------------------------|------------|
| Nº oraciones (%)               | Marginales |
| 50-70                          | 4 = 20%    |
| 70-90                          | 9 = 45%    |
| 90-97                          | 7 = 35%    |
| Total                          | 20 = 100%  |

Tabla 6: Resúmenes de las oraciones marginales.

Hoey (1991: 105), con la elaboración de resúmenes mediante la eliminación de las oraciones marginales, obtiene resúmenes que representan el 80% de oraciones del texto original. Este porcentaje es similar al nuestro, ya que, aunque el 35% excede dicho porcentaje, el 45% de los resúmenes oscila entre el 70% y 90% de oraciones del texto original. Hoey (1991: 113) sugiere que en algunos textos la proporción de oraciones marginales con respecto a las no marginales puede llegar hasta el 50%. En nuestro análisis sólo el 20% de los resúmenes conseguidos se sitúa en el intervalo de entre el 50% y el 70% de oraciones, una vez eliminadas las marginales.

Arcay (1998) aplica el análisis de patrones léxicos de Hoey (1991) a textos escritos en español de Venezuela. Aunque su investigación se centra en textos escritos en lengua española, los resúmenes confeccionados mediante la combinación de oraciones centrales –Arcay no investiga la posibilidad de confeccionar resúmenes a partir de las conexiones que establece la oración principal y concluyente– presentan un porcentaje de oraciones del texto original muy similar al de Hoey (1991). La razón que justificaría el que nuestros resultados difieran de los obtenidos por estos dos investigadores podría ser que el texto analizado por Hoey (1991) pertenece a un manual académico que versa sobre filosofía política y el análisis de Arcay (1998) se centra en 4 ensayos académicos pertenecientes al área de lingüística, los cuales difieren de nuestros textos no sólo con respecto al área de conocimiento, sino también con respecto al objetivo comunicativo y a la comunidad discursiva a la que van dirigidos.

Con respecto al tipo de oraciones –centrales, principales y concluyentes– que ha presentado un resumen más reducido, las tablas anteriores muestran que los resúmenes más reducidos –del 10% al 30%– se han obtenido mediante las oraciones centrales (55%) seguidos por la oración concluyente (40%) y que la oración principal es la que aporta resúmenes más amplios.

Podemos afirmar, pues, que, a excepción de los resúmenes confeccionados con las oraciones marginales en que un 80% de los resúmenes contienen más del 70% de oraciones del texto original, la cohesión léxica nos ha permitido realizar varios resúmenes de un mismo texto que no superan dicho porcentaje. Por tanto, nos adherimos a la siguiente afirmación de Tanskanen (1995: 531): “cohesion devices at the surface level of

a text contribute to the coherence of the text by making its unity clearer and more noticeable to the reader or the hearer”.

#### 4.2. ORACIÓN PRINCIPAL: CARACTERÍSTICAS

La siguiente tabla muestra la presencia en las introducciones analizadas de una oración principal capaz de, mediante la repetición léxica, expandir su radio de acción y ofrecer un resumen. Como podemos observar, sólo una introducción de las analizadas no presenta una oración principal con dicha característica:

| ORACIÓN PRINCIPAL |                |                   |
|-------------------|----------------|-------------------|
| Resumen           | Genera resumen | No genera resumen |
| Nº introducciones | 19             | 1                 |
| %                 | 95             | 5                 |

Tabla 7: Oración principal que genera resumen.

A continuación presentamos las características comunes de todas aquellas oraciones principales que han ofrecido un resumen del texto original con respecto a dos parámetros: dato numérico, es decir, si son las que establecen el número máximo de conexiones con las oraciones posteriores y ubicación, es decir, si el escritor las sitúa como primera oración del párrafo:

| ORACIÓN PRINCIPAL |    | %   |
|-------------------|----|-----|
| Dato numérico     | 19 | 100 |
| Ubicación         | 5  | 25  |

Tabla 8: Oración principal: dato numérico y ubicación.

De un total de 19 oraciones principales (recuérdese que una introducción no presenta oración principal capaz de generar un resumen, ver tabla 11) el 100% establece el mayor número de conexiones con las oraciones posteriores.

Estos datos parecen confirmar la importancia de la oración principal, ya que va a ser la oración que mayor número de conexiones establezca con las posteriores y, por consiguiente, con la información fundamental.

Con respecto a la ubicación de la oración principal, observamos que sólo una cuarta parte de las oraciones principales están situadas al inicio del párrafo, por lo que podríamos señalar que los párrafos no necesariamente se inician con la oración principal, lo que coincide con las investigaciones llevadas a cabo por Reid (1996) y Allison, Varghese y Wu (1999).

#### 4.3. ORACIÓN CONCLUYENTE: CARACTERÍSTICAS

A continuación presentamos las características de la oración concluyente siguiendo los mismos parámetros utilizados en la descripción de la oración principal:

| ORACIÓN CONCLUYENTE |                |                   |
|---------------------|----------------|-------------------|
| Resumen             | Genera resumen | No genera resumen |
| Nº introducciones   | 20             | 0                 |
| %                   | 100            | 0                 |

Tabla 9: Oración concluyente que genera resumen.

Todas las introducciones analizadas han presentado una oración concluyente capaz de, mediante la reiteración léxica, expandir su radio de acción y ofrecernos un resumen del texto original.

| ORACIÓN CONCLUYENTE |    | %   |
|---------------------|----|-----|
| Dato numérico       | 20 | 100 |
| Ubicación           | 11 | 55  |

Tabla 10: Oración concluyente: dato numérico y ubicación.

Con un total de 20 oraciones concluyentes podemos ver en la tabla anterior que el 100% de ellas presentan el número mayor de conexiones con oraciones anteriores.

La importancia de la oración concluyente, al igual que la de la oración principal, parece, pues, manifestarse en el dato numérico. El escritor conecta las oraciones que contienen la información relevante con estos dos tipos de oraciones, por lo que podría pensarse que el escritor de este tipo de textos va guiando al lector hacia la información fundamental del texto.

Con respecto a la posición que la oración concluyente ocupa en el párrafo, el 55% de oraciones concluyentes están situadas al final del párrafo. Este hecho nos permite señalar que, frente a la ubicación de la oración principal en el párrafo, que como hemos visto no solía ocupar la posición inicial, parece existir una tendencia superior a situar la oración concluyente al final del párrafo.

#### 4.4. TIPOS DE REPETICIÓN

Por último, la siguiente tabla muestra los distintos tipos de repetición que han presentado las introducciones analizadas:

| rs  | rc   | ps   | a  | pc   | hip  | tr   | s    | co-ref | e    | d    |
|-----|------|------|----|------|------|------|------|--------|------|------|
| 70% | 8'7% | 8'9% | 1% | 1'8% | 3'8% | 4'7% | 0'1% | 0%     | 0'2% | 0'3% |

Tabla 11: Tipos de repetición.

El tipo de repetición más frecuentemente utilizado ha sido la repetición del mismo término (repetición simple y compleja) con un 78'7%. Este dato, y la poca frecuencia con que se emplea la sinonimia, coincide con las sugerencias aportadas por los manuales convencionales sobre la redacción de un texto científico en lengua inglesa (Day 1995; Cremmins 1996: 173), los cuales señalan que la repetición exacta de un término es preferible a la sinonimia para evitar ambigüedad y contradice a Yang (1995: 8), que aconseja justamente lo contrario; utilizar sinónimos para evitar la monotonía que provoca la repetición de un mismo término.

A este respecto, pensamos que es interesante citar la investigación llevada a cabo por Fernández Polo (1995). Dicho investigador contrasta dos textos extraídos de *Scientific American* con sus correspondientes versiones españolas y observa que en la versión inglesa el tipo de repetición que predomina es la repetición simple. Sin embargo en las traducciones al español, se tiende a la utilización de pronombres, sinónimos, hiponimia y elipsis. La razón es acomodar la versión española a las normas retóricas de esta lengua “in which lexical repetition is seen as one of the most conspicuous signs of poor style” (Fernández Polo 1995: 258), lo cual afecta a la percepción de la coherencia por parte del lector, ya que, según Fernández Polo, éste encuentra dificultades en asignar el referente correcto a las formas pronominales o establecer relaciones entre dos términos léxicos diferentes que se refieren a una misma identidad. Concluye señalando que un texto que, en principio, iba dirigido a una audiencia no especializada en el contenido científico se convierte, mediante la modificación de los patrones léxicos, en uno que requiere por parte del lector un conocimiento mucho más especializado y profundo que el de la versión original. Investigaciones de este tipo son sumamente interesantes, ya que aportan una valiosa información no sólo sobre las posibles diferencias de organización del discurso científico en distintas lenguas y distintos géneros, sino también cómo la modificación de los patrones léxicos puede afectar a la correcta interpretación de un texto y, por lo tanto, a la coherencia.

El segundo tipo de recurso cohesivo encontrado en los textos analizados es la paráfrasis simple. Estos resultados coinciden con los de Myers (1991), Tanskanen (1995) y Norman (2003), que, igualmente, encuentran bastantes menos sinónimos que repeticiones simples en los escritos científicos que analizan, siendo los sinónimos, aún así, la segunda categoría de recurso cohesivo utilizado.

Los recursos cohesivos que han presentado una frecuencia de uso bastante menor han sido los gramaticales, lo cual coincide con las afirmaciones de Hoey (1991: 74) de que es el léxico el que organiza la información textual. A este respecto habría que citar las investigaciones de Frawley y Smith (1989), que analizan textos pertenecientes a las

áreas relacionadas con periodismo, ficción, religión y ciencia. Su investigación se centra en la diferente utilización de la referencia como recurso cohesivo en dichos textos. Sus resultados muestran que la referencia exofórica –pronombres de 1ª y 2ª persona– y endofórica –pronombres de 3ª persona– se utilizan con bastante menos frecuencia en los textos científicos en comparación con el resto de los textos analizados, lo cual coincide con nuestros resultados.

Podemos afirmar, pues, que en las introducciones del artículo de investigación analizadas la cohesión se ha realizado fundamentalmente mediante la repetición del mismo término y que la utilización de unidades gramaticales ha sido prácticamente inexistente (0.6%).

## 5. CONCLUSIÓN

En este trabajo de investigación hemos podido constatar que un hecho objetivo –la interacción o combinación de unidades léxicas presentes en las introducciones del artículo de investigación analizadas– nos ha permitido realizar varios resúmenes de un mismo texto, cada uno desde una perspectiva del mismo, pero todos ellos igualmente válidos en tanto en cuanto recogen la información fundamental del texto original. Asimismo, podemos afirmar que la cohesión léxica fundamentalmente, y no la cohesión gramatical, es la que establece relaciones múltiples de repetición entre oraciones adyacentes y no adyacentes de un texto, organizando la información semántica del mismo y proveyéndolo de unidad y de coherencia.

Es obvio que consideramos que se necesitan investigaciones futuras que apliquen el modelo de análisis de patrones léxicos de Hoey (1991) a textos del mismo y distinto tipo (en cuanto a género y especialidad) para poder corroborar los resultados aquí obtenidos e investigar si la cohesión léxica puede constituir un parámetro sensible al tipo de texto en el que se establece. Igualmente, consideramos necesarios estudios que, como los de Arcay (1998), reafirmen la función universal del léxico a la hora de proporcionar coherencia a un texto, independientemente de la lengua en que esté escrito. Asimismo, son necesarias investigaciones que, como las de Fernández Polo (1995), verifiquen que la modificación de los patrones léxicos de un texto afecta a la correcta interpretación de la versión traducida del mismo y, por lo tanto, a la coherencia. En nuestra opinión, nos encontramos ante un vasto campo sumamente interesante en el que queda aún mucho por hacer.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- Allison, D., S. Varghese, y S. M. Wu. 1999. "Local coherence and its limits: a second look at second sentences". *Journal of Second Language Writing* 8: 77-97.
- Arcay Hands, E. 1998. "Patrones léxicos en textos académicos escritos en español de Venezuela". *Revista de Ciencias Humanas y Sociales* 14: 32-47.

- Brown, G. y G. Yule. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Cremmins, E. T. 1996 (1982). *The Art of Abstracting*. USA: Information Resources Press.
- Curado Fuentes, A. 2001. "Lexical behaviour in academic and technical corpora: implications for ESP development". *Language Learning & Technology* 5: 106-129.
- Day, R. A. 1995 (1992). *Scientific English: A Guide for Scientists and Other Professionals*. Phoenix: Oryx Press.
- de Beaugrande, R. y W. Dressler. 1981. *Introduction to Text Linguistics*. London: Longman.
- Fernández Polo, J. 1995. "Some discoursal aspects in the translation of popular science texts from English into Spanish". Organization in Discourse. *Proceedings from the Turku Conference, Anglicana Turkuensia*, 14. Eds. B. Wärvik, S-K. Tanskanen y R. Hiltunen. Turku: University of Turku. 257-264.
- Flowered, J. 1993, "Concordancing as a tool in course design". *System* 21: 213-229.
- Frawley, W. J. y R. N. Smith. 1989. "Patterns of textual cohesion in genre-specific discourse". *Humans and Machines: Fourth Delaware Symposium on Language Studies*. Ed. S. Williams. Norwood, NJ: Ablex Publishing Corporation. 145-154.
- García de Sola, M. A. 2002. "La función pragmática de la repetición". *A Life in Words: A Miscellany Celebrating Twenty-Five Years of Association between the English Department of Granada University and Mervyn Smale (1977-2002)*. Eds. M. Carretero, E. Hidalgo, N. McLaren, y G. Porte. Granada: Universidad de Granada. 181-198.
- Gil Salom, L., C. Soler Monreal, y K. Stuart. 2001. "ACIA: Corpus Lingüístico del inglés científico-técnico". *Discourse Analysis and Terminology in Language for Specific Purposes*. Ed. J. C. Palmer. Castellón: Servicio de Publicaciones Universitat Jaume I. 345-356.
- Hasan, R. 1984 "Coherence and cohesive harmony". *Understanding Reading Comprehension*. Ed. J Flood. Delaware: International Reading Association. 181-219.
- Hasan, R. 1985. "Part B". M.A.K. Halliday y R. Hasan. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Oxford: Oxford University Press. 52-126.
- Hoey, M. 1991. *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Holmes, R. 1997. "Genre analysis, and the social sciences: an investigation of the structure of research article discussion sections in three disciplines". *English for Specific Purposes* 16: 321-337.
- Lewin, B. A. y J. Fine. 1996. "The writing of research texts: genre analysis and its applications". *Theories, Models and Methodology in Writing Research*. Eds. G. Rijlaarsdam, H. van den Bergh y G. Couzijn. Amsterdam: Amsterdam University Press. 423-444.
- Myers, G. 1991. "Lexical cohesion and specialized knowledge in science and popular science texts". *Discourse Processes* 14: 1-26.
- Norman, G. 2003. "Consistent naming in scientific writing: sound advice or shibboleth?". *English for Specific Purposes* 22: 113-130.
- Nwogu, K. 1997. "The medical research paper: structure and functions". *English for Specific Purposes* 16: 119-138.



- Phillips, M. 1985. *Aspects of Text Structure: an investigation of the lexical organisation of text*. Amsterdam: North-Holland.
- Posteguillo, S. 1999. "The schematic structure of computer science research articles". *English for Specific Purposes* 18: 139-160.
- Reid, J. 1996. "US. academic readers, ESL writers, and second sentences". *Journal of Second Language Writing* 5: 129-161.
- Swales, J. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Tanskanen, S-K. 1995. "Lexical cohesion meets communicative conditions". *Organization in Discourse*. Proceedings from the Turku Conference, Anglicana Turkuensia, 14. Eds. B. Wärvik, S-K. Tanskanen y R. Hiltunen. Turku: University of Turku. 531-538.
- Widdowson, H. G. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.
- Winter, E. O. 1979. "Replacement as a fundamental function of the sentence in context". *Forum Linguisticum* 4: 95-133.
- Yang, J. T. 1995. *An Outline of Scientific Writing: For Researchers with English as a Foreign Language*. Singapore: World Scientific Publishing.

ARTÍCULOS DE INVESTIGACIÓN PERTENECIENTES AL CORPUS ANALIZADO

- Aguilar, M. I. et al. 1998. "RP-HPLC binding domains of proteins". *Analytical Chemistry* 70: 5010-5018.
- Birch, M. E. 1998. "Analysis of carbonaceous aerosols: interlaboratory comparison". *The Analyst* 123: 851-857.
- Brätter, P. et al. 1998. "Speciation as an analytical aid in trace element research in infant nutrition". *The Analyst* 123: 821-826.
- Gilbert, B. et al. 2003. Analysis of green copper pigments in illuminated manuscripts by micro-Raman spectroscopy". *The Analyst* 128: 1213-1217.
- Greenway, G. M. et al. 1992. "The determination of food colours by HPLC with on-line dialysis for sample preparation". *Food Chemistry* 43: 137-140.
- Hamano Nagaoka, M. et al. 2004. "Binding patterns of vanadium to transferrin in healthy human serum studied with HPLC/high resolution ICP-MS". *The Analyst* 129: 51-54.
- Ishii, A. et al. 1998. "Determination of cyanide in whole blood by capillary gas chromatography with cryogenic oven trapping". *Analytical Chemistry* 70: 4873-4876.
- Jothi V. K. et al. 1998. "Bioconversion of solid food wastes to ethanol". *The Analyst* 123: 497-502.
- Kubáe, P. et al. 2004. "Application of a contactless conductivity detector to the determination of inorganic ions in ion chromatography". *The Analyst* 129: 82-86.
- Kuck, L. R. et al. 1998. "High-precision conductometric detector for the measurement of atmospheric carbon dioxide". *Analytical Chemistry* 70: 4678-4682.

- Matthew J. et al. 2003. "Dithizone derivatives as sensitive water soluble chromogenic reagents for the ion chromatographic determination of inorganic and organo-mercury in aqueous matrices". *The Analyst* 128: 1209-1212.
- Mayer, B. et al. 2003. "A 50% n-octylmethyl, 50% diphenyl-polysiloxane as stationary phase with unique selectivity for gas chromatography". *The Analyst* 128: 1238-1242.
- McHugh, C. J. et al. 2004. "SERRS dyes". *The Analyst* 129: 69-72.
- Nagwa H. S. et al. 2004. "Analysis of erythromycin and tylosin in bovine muscle using disposable screen printed electrodes". *The Analyst* 129: 15-19.
- Opper, C. et al. 1995. "Analysis of serotonin in whole-blood samples - A novel fully automated method". *LC-GC INT* 8: 104-107.
- Salimi, A. et al. 2004. "Boron doped diamond electrode modified with iridium oxide for amperometric detection of ultra trace amounts of arsenic (III)". *The Analyst* 129: 9-14.
- Simó, R. et al. 1998. "Refinement of the borohydride reduction method for trace analysis of dissolved and particulate dimethyl sulfoxide in marine water samples". *Analytical Chemistry*, 70: 4864-4867.
- Tami L. et al. 2004. "Frequency-dependent electrical detection of protein binding events". *The Analyst* 129: 3-8.
- Whittal, R. M. et al. 1998. "Nanoliter chemistry combined with mass spectrometry for peptide mapping of proteins from single mammalian cell lysates". *Analytical Chemistry* 70: 5344-5347.
- Xinyu L. et al. 2004 "On-site monitoring of biogenic emissions from *Eucalyptus dunnii* leaves using membrane extraction with sorbent interface combined with a portable gas chromatograph system". *The Analyst* 129: 55-62.