

4. Revisión del concepto de *cohesión*

Uno de los parámetros que determinan los segmentos en los que se puede fragmentar un determinado tipo textual es la cohesión (Hasan 1978). En este capítulo se citan los estudios principales sobre cohesión, sobre todo cohesión léxica e indagamos sobre si existe una correlación entre tipo textual y cohesión.

Tras mencionar muy brevemente la diferencia entre los términos *coherencia* y *cohesión*, repasaremos algunos de los modelos de análisis más conocidos en el ámbito anglosajón: Halliday y Hasan (1976), Hasan (1984), Winter (1974, 1979), Hoey (1983, 1991, 1995) y Morris y Hirst (1991). Posteriormente, comentaremos algunas de las aplicaciones de los estudios sobre cohesión, por una parte, para una mejor comprensión de la organización del texto y de los tipos de texto, y por otra, para el reconocimiento automático de la cohesión. También se verá qué dice la *Teoría Sentido Texto* sobre la cohesión.

4.1. Coherencia y cohesión

En torno a los años setenta, en el Reino Unido surge una gramática del texto (Quirk, Greenbaum, Leech and Svartvik, 1972) en la que adquiere importancia el concepto de *cohesión*. Las especulaciones sobre la cohesión cristalizaron en Halliday y Hasan (1976) y se extendieron fuera del Reino Unido. Y así, desde la lingüística textual alemana, se empezó a hablar de *coherencia* y *cohesión* como dos criterios fundamentales de textualidad (Beaugrande y Dressler 1981).

En esencia, la cohesión es la forma en que determinados elementos léxicos o gramaticales de una oración pueden conectar esa oración a elementos anteriores o posteriores dentro de un texto (Hoey 1991: 11). La cohesión se presenta de un modo explícito en el texto, es una propiedad objetiva del texto. Por el contrario, la coherencia es subjetiva porque es la unidad de significado que percibe el receptor de un mensaje al relacionar el texto con la realidad extralingüística. Ambos están relacionados en el sentido de que la cohesión ayuda a que el lector perciba la coherencia del texto.

4.2. Halliday y Hasan (1976)

Halliday y Hasan (1976) presenta una clasificación de los medios disponibles en inglés para crear cohesión en los textos. Para Halliday y Hasan, la organización del texto (*texture*) se construye a partir de relaciones entre elementos del texto tanto gramaticales como léxicas. A estas relaciones las denominan *enlaces cohesivos* (*cohesive ties*).

Los cinco tipos de enlaces cohesivos que identifican son la *conjunción*, *referencia*, *sustitución*, *elipsis* y *cohesión léxica*.

4.2.1. La conjunción

En esta categoría, se incluyen las conjunciones y otros *conectores* textuales que marcan de forma explícita relaciones semánticas (consecuencia, adición, oposición, etc.) entre oraciones o entre unidades superiores.

- (2) In the present study, loss of Rb protein expression was not a statistically significant prognostic variable as an individual factor in the cohort examined. **However**, our data suggested that [...]

(Cfr. Apéndice II: texto 1, pág. 2)

4.2.2. La referencia

Este recurso cohesivo no marca relaciones semánticas sino que es una relación semántica que se establece cuando elementos, principalmente pronominales y demostrativos, indican que la identidad de la entidad o acción a la que hacen referencia puede recuperarse del cotexto próximo. Hay dos tipos de referencia, anafórica y catafórica. En los ejemplos, los elementos relacionados mediante referencia aparecen en **negrita**.

Referencia anafórica:

- (3) **Inactivation of the retinoblastoma (Rb) gene** has been documented in various types of cancers, including lung cancer. **It** has also been associated with aggressive tumor behavior and poor clinical outcome in specific types of cancer.

(Cfr. Apéndice II: texto 1, pág. 1)

Referencia catafórica:

- (4) Survival was analyzed for the 89 patients who met **the following criteria**: (1) survived for more than 3 months after surgery; (2) did not die of causes other than lung cancer within 5 years after surgery; and (3) were followed for more than 3 years after surgery (for patients who remained alive).

(ibid, pág.2)

4.2.3. La sustitución

Relación gramatical por la que un elemento reemplaza a un término precedente cuyo referente no es la misma entidad, sino una entidad a la que el mismo término puede designar.

- (5) **A tougher substance** is needed. This **one** is too soft

(Kennedy y Bolitho 1984: 132)

4.2.4. La elipsis

Este recurso se emplea cuando se omite una frase o idea que está explícitamente expresada en la oración previa porque no se considera necesaria para la comprensión de la segunda oración.

- (6) The exploitation of energy will guarantee a more pleasant way of life. But for how long? (**will the exploitation of energy guarantee a more pleasant way of life?**)

(ibid)

4.2.5. La cohesión léxica

Bajo el nombre de cohesión léxica se incluyen una variedad de relaciones semánticas que pueden existir entre lexemas. La cohesión léxica es el tipo de enlace cohesivo más frecuente e importante, como se constata en el análisis realizado por Halliday y Hasan. Este punto de vista es compartido por Hoey (ibid: 10):

Lexical cohesion is the only type of cohesion that regularly forms multiple relationships (though occasionally reference does so too). If this is taken into account, lexical cohesion becomes the dominant mode of creating texture.

Halliday y Hasan identifican dos formas de cohesión léxica: la *reiteración* y la *colocación*. Esta terminología se modificará en Hasan (1984).

4.2.5.1. La reiteración

La *reiteración* es la forma en que un lexema puede evocar o traer a la memoria el sentido de un lexema anterior. Esta se consigue mediante la repetición literal de esa palabra o mediante la utilización de un sinónimo, sinónimo parcial, hipónimo o hiperónimo.

- (7) **Investigadores** de Baltimore acaban de usar en ratones una vacuna contra este tipo de **tumor** y la experiencia ha resultado exitosa. El estudio, publicado en la revista Cancer Research, revela que los **científicos** utilizaron una proteína, llamada E7, que juega un papel fundamental en la progresión del **cáncer** de cuello de útero.

(Cfr. Apéndice II, texto 2)

4.2.5.2. La colocación

La *colocación* aparece definida de un modo tan ambiguo que incluye relaciones tan variadas como las existentes entre *boy—girl*; *wet—dry*; *ill—doctor*; *try—succeed*.

Esta ambigüedad ha hecho que, en estudios posteriores (Hoey 1991), la colocación se entienda en el sentido que le da Firth (1957): "the company a word keeps". Este autor entiende la colocación como la relación que un lexema establece con otros lexemas que aparecen con una probabilidad significativa en su cotexto inmediato. Normalmente, desde la lingüística de corpus (Sinclair 1991), este cotexto abarca cuatro palabras a ambos lados del lexema en cuestión.

A diferencia de otras combinaciones habituales de palabras, según Tercedor (1999: 89), en las colocaciones participan dos elementos, de los cuales, la base o palabra clave mantiene su significado y goza de plena autonomía, de ahí que admitan cambios de orden y variación sintáctica. En la idea de autonomía de sus componentes también coincide Montero Martínez (2000) cuando afirma que el significado de las colocaciones puede inferirse de las partes.

Para ilustrar el concepto de colocación, mencionaremos una serie de ejemplos en inglés y en castellano. Stubbs (1996) analiza un corpus de textos periodísticos publicados en el Reino Unido durante el gobierno del partido conservador y destaca la frecuencia con que la palabra *nation* “coocurre”²⁶ con otros lexemas como *biggest, European, favoured, industrialized, largest, trading* y *wealthiest*. En este ejemplo, se destaca el papel que puede jugar la repetición de determinados patrones lingüísticos en la consolidación de una determinada ideología (Fowler 1996; Hatim 1997).

El ejemplo correspondiente al castellano ha sido sacado de un corpus de textos periodísticos sobre educación publicados en *El Mundo* y *El País* en los años 1995 y 1996 (López Rodríguez 1997). En este corpus, la palabra *enseñanza* aparece con mucha frecuencia en compañía de las palabras *pública, obligatoria, calidad, secundaria, primaria* y *centros*. Estos resultados fueron obtenidos con el programa informático *Wordsmith Tools*²⁷.

TABLA 5: Colocaciones de la palabra *enseñanza*²⁸

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	ENSEÑANZA	129	0	0	0	0	0	0	0	129	0	0	0	0	0
2	PÚBLICA	19	0	19	0	0	0	0	0	0	19	0	0	0	0
3	OBLIGATORIA	17	1	16	1	0	0	0	0	0	6	10	0	0	0
4	CALIDAD	17	13	4	1	0	8	4	0	0	0	2	0	0	2
5	SECUNDARIA	14	0	14	0	0	0	0	0	0	13	0	1	0	0
6	PRIMARIA	12	0	12	0	0	0	0	0	0	10	1	0	0	1
7	CENTROS	9	8	1	0	1	3	4	0	0	0	0	0	0	1
8	FEDERACIÓN	8	8	0	2	0	0	6	0	0	0	0	0	0	0
9	GENERAL	8	6	2	5	0	0	1	0	0	0	0	2	0	0
10	ESO	8	3	5	0	1	0	0	2	0	0	0	5	0	0
11	ENSEÑANZA	7	0	0	0	0	0	0	0	7	0	0	0	0	0
12	CCOO	6	1	5	0	0	1	0	0	0	0	4	1	0	0
13	SINDICATOS	5	4	1	1	1	1	1	0	0	0	0	0	0	1
14	TRABAJADORES	5	4	1	0	0	4	0	0	0	0	0	0	0	1
15	CASTELLANO	5	0	5	0	0	0	0	0	0	0	3	0	0	2

²⁶ El verbo *coocurrir* se utiliza con frecuencia en la lingüística de corpus y es un calco del inglés *co-occur*

²⁷ Programa creado por Mike Scott [<http://www.liv.ac.uk/~ms2928/homepage.html>] de la *University of Liverpool* y distribuido por Oxford University Press.

²⁸ Se incluyen sólo las palabras plenas en su contexto.

4.3. Hasan (1984)

En 1984, Hasan publica un estudio en el que explora la relación entre coherencia y cohesión en la literatura infantil e intenta medir la coherencia sobre la base de la comprensión de los relatos por parte de los lectores. Hasan (1984) también reformula las categorías de cohesión léxica propuestas en 1976 y las organiza en dos categorías: *general* e *inmediata* (*instantial*).

TABLA 6: Categorías de cohesión léxica

COHESIÓN LÉXICA GENERAL	COHESIÓN LÉXICA INMEDIATA
Repetición: <i>leave, leaving, left</i>	Equivalencia: <i>the <u>sailor</u> was their <u>daddy</u></i>
Sinonimia: <i>leave, depart</i>	Designación (<i>naming</i>): <i>the dog was called <u>Toto</u></i>
Antonimia: <i>leave, arrive</i>	Apariencia (<i>semblance</i>): <i>the <u>deck</u> was like a <u>pool</u></i>
Hiponimia: <i>travel, leave</i>	
Meronimia: <i>hand, finger</i>	

Hasan (1984: 202)

Después de analizar los enlaces cohesivos en relatos infantiles, Hasan descubre que no existe una relación entre el número de enlaces cohesivos y la coherencia que encuentran los lectores. Por eso, Hasan sostiene que la función de los enlaces cohesivos es formar una especie de *cadena* (*chains*) y que estas interactúan unas con otras. A esta interacción la denomina *armonía cohesiva*. Son las cadenas y no los enlaces cohesivos aislados los que crean cohesión.

Una cadena cohesiva se forma cuando un elemento cohesivo refiere hacia un elemento anterior, que a su vez también está relacionado cohesivamente con otro elemento previo. Existen dos tipos de cadenas: la de *identidad* y de *similitud*.

La cadena de identidad está compuesta de cadenas cohesivas que comparten los mismos referentes. En inglés es necesario hacer referencia mediante pronombres, reiteraciones o elementos equivalentes al tema del que se está tratando en sucesivas oraciones.

La cadena de similitud se da cuando distintas entidades están implicadas en procesos o descripciones paralelas y, por tanto, comparten relaciones cohesivas con palabras de significado parecido. Por ejemplo, si en una descripción tres entidades son descritas con el adjetivo *little*, se formará una cadena de similitud formada por las distintas instancias de *little*.

Como ilustración, presentamos un fragmento de uno de los cuentos estudiados por Hasan (Hoey 1991: 15).

- (8) 1 Once upon a time there was a little girl
 2 and she went out for a walk
 3 and she saw a lovely little teddy bear
 4 and so she took it home
 5 and when she got it home she washed it
 6 and she had the teddy bear for many many weeks and years.

En este tipo de texto, las cadenas son muy obvias. Es posible reconocer tres cadenas de identidad y dos de similitud.

Cadenas de identidad

1 girl 2 she 3 she 4 she 5 she, she 6 she
 3 teddy bear 4 it 5 it, it 6 teddy bear
 4 home 5 home

Cadenas de similitud

2 went out 5 got ... home
 4 took 6 had

La conclusión a la que llega Hasan es que la presencia de cadenas aisladas no garantiza la coherencia del texto, pero sí la interacción entre cadenas. Esta conclusión será crucial en la metodología de Hoey, en la que los recursos cohesivos adquieren sentido sólo cuando están interrelacionados.

4.4. Winter (1974, 1979, 1986)

La investigación de Winter gravita en torno a dos ejes: la forma en que la gramática de las oraciones contribuye a su interpretación en contexto y la función de los enlaces cohesivos en la organización del texto.

La principal función de los enlaces cohesivos es la *repetición*, un recurso que llama la atención sobre determinada información del texto. La repetición puede adquirir distintas formas:

- a) La *sustitución*, que es similar a la *referencia* de Halliday y Hasan (1976)
- b) La *eliminación (deletion)*, que corresponde a la *elipsis*
- c) y la *repetición léxica*, que coincide con la *cohesión léxica*.

De todas formas, Winter (1979: 101) presta atención a la repetición que tiene un carácter sistemático. Es decir, su énfasis está en la repetición significativa de uno o varios de los elementos constituyentes de una *cláusula* en la estructura de otra. Esta repetición proporciona a cada cláusula una serie de elementos constantes que permiten reconocer cuál es la información nueva y qué relevancia tiene esta en el cotexto.

La repetición también hace necesarios una serie de cambios en la estructura de la nueva cláusula. Winter, siguiendo a Dixon (1965), Quirk *et al.* (1972), denomina a estos cambios *reemplazos (replacements)*. El reemplazo no se ha de entender como la sustitución de una palabra por otra sino la presentación de nueva información en un contexto que ya ha sido previamente presentado.

A continuación, se incluye un ejemplo tomado de la publicación *New Statesman* y la representación de cómo se producen estos reemplazos en Hoey (1983), que destaca qué elementos son constantes y cuáles variables.

- (9) Pressures built up on all sides: **his** father, a ‘moderately successful plumbing contractor’ (said Time) **demande**d performance. **His** mother, who left her husband in Florida and moved to Austin to be near her son, **demande**d love. Whitman could provide neither.

TABLA 7: Reemplazos y elementos constantes del ejemplo (9)

	His His	father mother	demanded demanded	performace love
REPETICIÓN REEMPLAZO	His	*****	demanded	*****
CONSTANTE VARIABLE	His	parent which sex	demanded	response by him what kind of response

Mediante la repetición y el reemplazo, se consigue la interpretación de pares de oraciones en el texto y Winter consigue explicar cómo la gramática y la cohesión interactúan en los textos. A partir de ahí, Winter especula sobre la relación entre las oraciones y la organización del texto.

En conjunto, el trabajo realizado por Winter contribuye a una mejor comprensión del papel que juega la cohesión. Los puntos clave de su aproximación aparecen resumidos en Hoey (1991: 20) y tendrán una gran repercusión en este discípulo de Winter.

- La cohesión ha de ser interpretada en el contexto de las oraciones en las que ocurre.
- Para estudiar la cohesión hay que ver cómo la repetición forma agrupaciones entre pares de oraciones.
- Es más importante ver la función de repetición de la cohesión que las clasificaciones de tipos de cohesión.
- La repetición tiene valor informativo y proporciona un marco para interpretar los elementos que han cambiado en cada nueva oración.
- Las relaciones de repetición que se establecen entre oraciones son normalmente múltiples y entre oraciones no contiguas.

4.5. Metodología de Hoey (1991, 1995)

Hoey (1983, 1991, 1995) continúa en la línea de Winter, Halliday y Hasan y propone una metodología de análisis textual basada en el estudio de la cohesión, sobre todo, de la cohesión léxica. Su metodología hace posible detectar la sistematicidad de las relaciones cohesivas de un texto y representar gráficamente cómo los rasgos cohesivos se combinan para organizar los textos y crear continuidad en los mismos. El objetivo último es demostrar que la coherencia, la cohesión y la organización (*texture*) del texto se crean principalmente por las relaciones léxicas entre las palabras y entre las oraciones²⁹.

A esta metodología se la conoce como *patterns of lexis* y ha destacado por su capacidad de evaluar la red compleja de relaciones léxicas que caracterizan los textos naturales.

En su revisión a la taxonomía de enlaces cohesivos de Halliday y Hasan, Hoey (1991: 7) destaca que la reiteración léxica y la colocación son principalmente formas de relación léxica, y sólo de un modo secundario, marcadores de relaciones textuales. En esto se diferencian de la conjunción, la referencia, la sustitución y la elipsis. Las relaciones léxicas creadas por reiteración léxica y la colocación contribuyen a la creación y a la organización o *textura* del texto, a la vez que este proporciona el contexto para la creación y la interpretación de relaciones léxicas.

La mayoría de estos enlaces cohesivos, a excepción de la conjunción, se basan en la repetición porque permiten que se mencione de nuevo lo que ya se ha mencionado. La repetición tiene como función mostrar las relaciones léxicas entre las palabras y las oraciones. Por este motivo, Hoey, al igual que Winter, estima necesario remodelar la clasificación de recursos cohesivos de Halliday y Hasan para dar cuenta de aquellos que se sustentan en la repetición.

²⁹ A la cohesión que se establece entre palabras, Sinclair (1993) la denomina *point-to-point cohesion*. Este tipo de cohesión tiene una naturaleza léxica, en contraposición con la cohesión entre oraciones o complejos de cláusulas, que, según Sinclair, tiene una naturaleza textual o estructural.

4.5.1. Cohesión y tipos de repetición

A continuación presentamos los recursos cohesivos que Hoey considera esenciales en el establecimiento de las relaciones léxicas de un texto. Las categorías propuestas se basan en la repetición.

4.5.1.1. Repetición léxica simple

Se produce cuando un lexema que ya ha aparecido en el texto se repite tal cual o con la única alteración de un morfema gramatical que no modifica las relaciones sintagmáticas. Hoey sólo aplica esta repetición a palabras plenas (*open-set lexical items*).

Presentamos tres fragmentos de oraciones que contienen las siguientes instancias de repetición léxica simple: *authors—authors*; *prognostic significance—prognostic significance* y *study—studies*.

- (10) [...] and the **authors** of this article have reported previously on the **prognostic significance** of both of them.
- (11) In the present **study**, the **authors** evaluated the **prognostic significance** of [...]
- (12) Larger prospective **studies** with Stage I NSCLC patients are necessary to confirm the current findings.

(Cfr. Apéndice II: texto 1, pág. 1)

4.5.1.2. Repetición léxica compleja

Ocurre cuando se repiten dos lexemas que comparten un morfema léxico, aunque son formalmente diferentes, o que son formalmente idénticos pero no tienen la misma función sintáctica. Por ejemplo, el sustantivo *humans* y el adjetivo *human*. Dos elementos forman una repetición de este tipo si se pueden parafrasear en el contexto en que aparecen de modo que la paráfrasis de uno incluya al otro. Y así, *human* es igual que *of humans*. Veamos ahora un ejemplo tomado de nuestro corpus.

- (13) [...] and p53 protein expression in the same **tumor** specimens, and [...]
- (14) [...] those with Rb-/ras+ **tumors** survived for a significantly shorter period

(ibid)

4.5.1.3. Paráfrasis simple

Es el tipo de repetición en la que una unidad léxica sustituye a otra en contexto sin que se produzca una pérdida o ganancia en especificidad y sin que se produzca ningún cambio de significado. Hasan denominaba a este tipo de repetición *sinonimia*. La paráfrasis simple puede ser *parcial*, si la sustitución sólo funciona en una dirección o *mutua*.

- (15) Aunque el 40% de los varones mayores de 50 años tiene **cáncer** de próstata, sólo el 8% de todos los **tumores** acabará produciendo un problema al **paciente**. Por eso, es necesario encontrar métodos y fórmulas que concreten mucho más el qué hacer con los **enfermos**.

(Cfr. Apéndice II, texto 3)

4.5.1.4. Paráfrasis compleja

Se da cuando una unidad léxica del texto se puede definir en función de otra unidad léxica del texto con la que no comparte ningún morfema y cuando se cumple una de las siguientes situaciones:

- Cuando entre las dos unidades léxicas existe una relación de antonimia. El término de antonimia engloba un gran número de casos que Hoey no ha sabido sistematizar. Según Lyons (1977), los distintos tipos de oposiciones léxicas relacionadas con la antonimia entran en alguna de estas categorías:

(a) *antónimos graduales: hot—cold*

(b) *antónimos binarios: dead—alive; same—different*

(c) *términos inversos (converses): lend—borrow; parent—child*

(d) *términos pertenecientes a sistemas de incompatibilidad múltiple (open-ended systems of multiple incompatibility):*

*Secretary of State—Minister—Minister of State—Parliamentary
Under-Secretary of State*

January—February—March ...

Esta falta de sistematicidad le lleva a eliminar la antonimia de su clasificación de recursos cohesivos (Hoey 1995), aunque añadiendo una categoría, la pertenencia a un conjunto cerrado de palabras plenas.

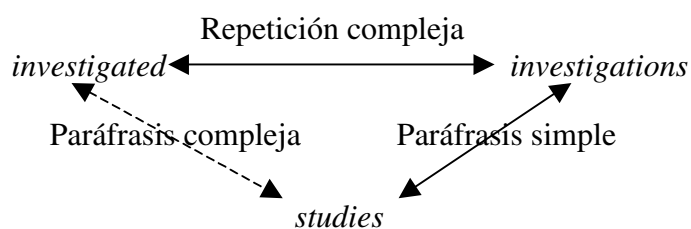
- Cuando el lexema que se analiza es una repetición compleja de otro elemento y, al mismo tiempo, una paráfrasis simple de un tercer elemento. En este caso se establece una paráfrasis compleja entre el segundo y el tercer elemento. Veamos un caso de paráfrasis compleja en el siguiente texto

(16) For this reason, induction chemotherapy is being **investigated** in patients expected to have a poor prognosis after standard surgery and radiotherapy. [...] Clinical **investigations** have reported an improvement of survival in stage III NSCLC after induction chemotherapy by using different combinations of cytotoxic drugs. [...] The combination of mitomycin, vindesine or vinblastine and cisplatin has produced encouraging results in several **studies** [...]

(Cfr. Apéndice II, texto 4)

A continuación presentamos un triángulo que ilustra cómo se genera una paráfrasis compleja a raíz de las relaciones de repetición que se establecen entre un elemento y otros dos.

DIAGRAMA 6: Triángulo de conexiones



En el diagrama se ve cómo, si una palabra (*investigations*) está en relación con otras dos (*investigated* y *studies*), se crea también un *enlace putativo* entre *investigated* y *studies*. El nombre que Hoey asigna a esta relación es el de *link triangle* o triángulo de conexiones.

- Otra situación en la que se habla de paráfrasis compleja es cuando uno de los elementos del triángulo de enlaces que funciona de mediador de relaciones no aparece y el lector infiere este elemento. Este tipo de paráfrasis se da, por ejemplo, entre los lexemas *investigadores* y *estudio*. El elemento implícito que sirve de mediador es la palabra *investigación*.

(17) **Investigadores** de Baltimore acaban de usar en ratones una vacuna contra este tipo de tumor y la experiencia ha resultado exitosa. El **estudio**, publicado en la revista [...]

(Cfr. Apéndice II, texto 2)

4.5.1.5. Repetición superordinada, hipo-nímica y correferencial

Se produce cuando el significado de un término está incluido en el significado del otro y sus referentes son idénticos. En esto se diferencia de los casos mencionados con anterioridad, en los que los lexemas implicados contenían la misma información.

Por ejemplo, la relación entre *genetic* y *biologic* o *cancer* y *lung carcinoma* (ibid, texto 1), o la repetición correferencial que se produce al mencionar *tumor maligno más frecuente en el hombre* como referencia al *cáncer de próstata* (texto 3). Creemos interesante resaltar que para automatizar este último tipo de repetición es necesaria información enciclopédica.

4.5.1.6. Otras formas de repetición que no son léxicas.

Bajo esta categoría, Hoey incluye la *sustitución* y la *elipsis*. La sustitución es una forma de repetición de elementos textuales que dependen absolutamente de otros elementos para ser interpretados semánticamente porque no tienen significado proposicional por sí solos. Es decir, en la sustitución se emplean *palabras forma* que hacen referencia no sólo a un elemento léxico del texto sino también a una realidad externa al texto. Hoey utiliza el término sustitución en un sentido más amplio que Halliday y Hasan, quienes denominaron a este tipo de repetición *reference*. La sustitución recurre a pronombres personales, pronombres y determinantes demostrativos y proformas como *one*, *do*, *so*, *(an)other* y *similar*.

Hoey también incluye como forma de repetición la elipsis. Sin embargo, en su método de análisis, Hoey presta atención principalmente a las formas de repetición léxica, que son más fáciles de reconocer por medios computacionales.

Para que quede más claro qué categoría asignar a cada tipo de repetición, Hoey (ibid: 58-60 y 69) elabora unos diagramas con preguntas (Véase Apéndice III).

En 1995, Hoey elimina de esta lista de elementos cohesivos la paráfrasis compleja (en la que se incluía la antonimia), la repetición superordinada e hiponímica y la elipsis, y añade una categoría relacionada con la colocación, la pertenencia a un conjunto ordenado de palabras plenas como *March—April*.

4.5.2. Conexiones, vínculos y redes

A cada repetición de un elemento léxico de una oración en otra oración, ya sea contigua o no, Hoey la denomina *link*. De ahora en adelante, utilizaremos el término *conexión* como traducción para *link*. Se supone que, al repetir elementos léxicos en otra oración (*conexión*), estamos estableciendo una relación entre los mismos y entre las oraciones implicadas.

Al proponer la palabra *conexión*, Hoey está marcando una diferencia con respecto al término *cohesive tie* de Halliday y Hasan. El término *conexión* es más restringido porque designa las relaciones cohesivas léxicas y no-léxicas entre los elementos de dos oraciones pero sin incluir las conjunciones, colocaciones y algunos tipos de referencia que el término *cohesive tie* abarca. Además, la palabra *conexión* no parece indicar la dirección que siguen las relaciones.

A continuación presentamos las dos primeras oraciones del capítulo introductorio de un libro de texto para universitarios sobre filosofía política tomado de Hoey (1991: 246)³⁰. Hemos marcado las conexiones entre elementos léxicos mediante flechas.

³⁰ En el Apéndice IV aparecen las dieciséis primeras oraciones del texto.

What is attempted in the following volume is to present to the reader a series of actual excerpts from the writings of the greatest political theorists of the past; selected and arranged so as to show the mutual coherence of various parts of an author's thought and his historical relation to his predecessors or successors; and accompanied by introductory notes and intervening comments designed to assist the understanding of the meaning and importance of the doctrine quoted. The book does not purport to be a history of political theory, with quotations interspersed to illustrate the history.

Estas conexiones adquieren importancia en la organización de los textos en tanto en cuanto aparecen en conjunción con otras conexiones (*clustering of links*). Por tanto, no interesan las conexiones aisladas entre oraciones sino los haces de conexiones que llegan a tejer una especie de cuerda entre dos oraciones.

Por este motivo, Hoey introduce un nuevo término para referirse a la relación estrecha entre dos oraciones que comparten más conexiones de lo que es normal para un determinado tipo textual. Este término es el de *bond*, que hemos traducido como *vínculo*, porque el significado de este último lexema es el de una unión significativa entre dos objetos³¹. Con el fin de delimitar aún más los términos *link* y *bond*, Hoey (ibid: 91) afirma: "Lexical items form *links*, and sentences sharing three or more links form *bonds*".

En textos que no son ni narrativos ni especializados, Hoey establece que el número mínimo de conexiones para establecer un *vínculo* es de tres. Este umbral cambia de unos textos a otros y no se puede establecer *a priori*. Y así, en algunos textos científicos hay tantas conexiones entre oraciones que es necesario fijar un umbral de seis oraciones. Es decir, en esos textos sólo se computan como oraciones vinculadas aquellas que comparten un mínimo de seis conexiones.

El conjunto de vínculos que se interrelacionan en un texto (*net*) juegan un papel esencial en la organización del texto en el sentido de que crean su textura (véase p. 76). En esta *red*, un lector es capaz de calificar dos oraciones entre las que existe un vínculo como portadoras de proposiciones mutuamente relevantes. También es capaz, previa alteración de algunos elementos cohesivos, como la sustitución de una proforma por su

³¹ Fernández Polo (1999: 194) traduce *links* como "lazos" y *bonds* como "uniones".

referente, de entender esas oraciones como coherentes y como si estuvieran contiguas en el texto.

4.5.3. Cómputo y representación gráfica de las conexiones y vínculos

En los apartados anteriores hemos visto qué tipos de conexiones se pueden establecer entre los elementos léxicos de un texto y cómo un determinado número de conexiones pueden formar un vínculo entre oraciones. Ahora, nos disponemos a presentar cómo se pueden computar estas conexiones y vínculos de una forma sistemática. En aras de la claridad, hemos presentado las indicaciones metodológicas de Hoey en forma de etapas.

(1) Identificación de repeticiones y elaboración de una matriz de repetición

En primer lugar, se reconocen los casos de repetición léxica. Para hacerlo de una forma consistente, Hoey propone que se computen las conexiones siempre como si fueran anafóricas. Es decir, al leer la segunda oración de un texto, se empiezan a establecer las conexiones con la primera oración.

También apunta que las conexiones se establecen no sólo con elementos de la oración precedente, sino con *todas* las unidades léxicas de las que son una repetición en otra oración. A continuación se identifican los casos de paráfrasis, sustitución y elipsis.

Para evitar errores en el cómputo, las repeticiones se representan en una *matriz de repetición*. Esta matriz permite registrar el número y el tipo de conexiones que se establecen entre elementos de forma que se puedan visualizar las conexiones de cada oración con otras, las *cadenas léxicas* del texto. Para comprender lo que es una matriz de repetición, sería conveniente consultar las tablas 1 y 2 del Apéndice V.

En la Tabla 1 del Apéndice V aparece la primera columna ampliada de la matriz de repetición correspondiente a las primeras oraciones del texto sobre filosofía política mencionado con anterioridad (véase Apéndice IV). En esta se perciben las conexiones basadas en repetición léxica de la primera oración con las quince siguientes, de ahí que sólo aparezcan palabras plenas.

La Tabla 2 representa la matriz de repetición completa e incluye además de los casos de repetición léxica, los de elipsis y referencia deíctica. En la matriz aparecen filas y columnas encabezadas por números que corresponden a cada una de las oraciones del

texto. En las filas, se presentan las conexiones de cada oración con las que le anteceden; en las columnas, aparecen las conexiones con oraciones posteriores. En cada casilla, hay pares de palabras entre los que se establece un tipo de conexión. La primera palabra del par pertenece a la oración precedente, indicado por el número de columna, mientras que la segunda pertenece a la oración con el número de la fila.

Así por ejemplo, la primera casilla de la tabla, representa las conexiones entre la primera oración (número de columna) y la segunda oración (número de fila). Los seis pares de palabras indican que hay seis conexiones entre la primera y la segunda oración. En el par (*volume-book*), que es un ejemplo de paráfrasis simple parcial, *volume* pertenece a la primera oración (primera columna) y *book*, a la segunda oración (segunda fila).

(2) Ajuste de las conexiones computadas

Siguiendo el principio de que un elemento sólo puede formar una conexión por oración³², en el caso de que un elemento repita dos elementos de una oración anterior o de que una oración contenga dos elementos que repiten un elemento anterior, se registra sólo una conexión

En estos casos hay que plantearse qué conexión incluir en el análisis. Hoey establece que existen conexiones que tienen más peso y que deben primar en el análisis. Las presentamos en orden decreciente.

- Repetición léxica simple
- Repetición léxica compleja
- Paráfrasis simple mutua
- Paráfrasis simple parcial
- Paráfrasis compleja por antonimia
- Otros tipos de paráfrasis compleja
- Sustitución
- Correferencia
- Elipsis

Hoey también elimina las conexiones internas de las oraciones y no las computa en el análisis. No estamos de acuerdo con esta decisión porque dejaría al margen casos de cohesión importantes como la colocación y la relación tan frecuente en textos

³² Esto no impide que un elemento establezca conexiones con más de una oración.

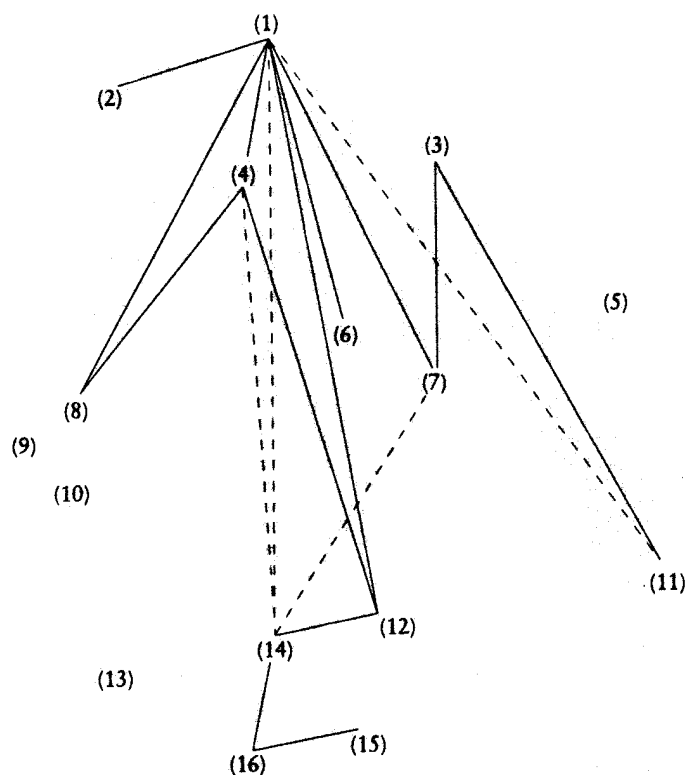
Observando la tabla de conexiones se establece qué umbral es apropiado para determinar un vínculo. Para los textos no-narrativos, Hoey lo establece en tres conexiones. La red de vínculos de un texto se puede representar gráficamente.

(5) *Representación de una red de vínculos: el diagrama topológico*

A partir de la tabla de conexiones se puede representar gráficamente la red de vínculos del texto. Hoey denomina a esta representación *diagrama topológico*. En un diagrama topológico, cada oración está representada por un número y cada trazo que la une con otra oración es un vínculo. Los trazos discontinuos indican que el vínculo no es claro. Sólo hay vínculos entre oraciones que tienen un número de conexiones por encima del umbral establecido.

El siguiente diagrama representa la red de *oraciones vinculadas* del texto objeto de análisis, que son las que aportan al texto coherencia. Esta red está tejida principalmente por patrones de repetición léxica, como vimos en la matriz de repetición.

DIAGRAMA 7: Diagrama topológico del texto del Apéndice IV



Los resultados de este diagrama se pueden representar como una tabla parecida a la del apartado (3), sólo que ahora estamos computando los vínculos entre oraciones. En esta, se le asignan a cada oración dos coordenadas. La primera se refiere al número de vínculos con las oraciones precedentes, y la segunda, al número de oraciones que le siguen. Cuando aparecen dos grupos de coordenadas, estamos ante vínculos formados por conexiones dudosas. Sumando las dos coordenadas, se obtiene el número total de vínculos de cada oración con las oraciones de este fragmento. Por ejemplo, la cuarta oración está vinculada con tres oraciones, una precedente y dos que le siguen.

TABLA 9: Tabla de vínculos del texto del Apéndice IV

1	(-, 6) (-, 8)	9	(0, 0)
2	(1, 0)	10	(0, 0)
3	(0, 2)	11	(1, 0) (2, 0)
4	(1, 2)	12	(2, 1)
5	(0, 0)	13	(0, 0)
6	(1, 0)	14	(1, 1) (4, 1)
7	(2, 0)	15	(0, 1)
8	(2, 0)	16	(2, 0)

Hoey (1995) presenta un formato de tablas de vínculos, adecuado para representar textos más extensos. En el eje de ordenadas aparecen todas las oraciones, y en el de abscisa, las oraciones precedentes (a la izquierda) y las oraciones que siguen (a la derecha). El umbral para formar un vínculo se estableció en tres conexiones. Especificamos entre paréntesis el número de conexiones igual o superior a tres. El texto del Apéndice IV quedaría del siguiente modo:

TABLA 10: Tabla de vínculos (adaptación de Hoey 1995: 77)

	1	2(6) 4(5) 6(3) 7(4) 8(5) 11? 12(3) 14?
1(6)	2	
	3	7(3) 11(3)
1(5)	4	8(4) 12(3) 14?
	5	
1(3)	6	
1(4) 3(3)	7	14?
1(5) 4(4)	8	
	9	
	10	
1? 3(3)	11	
1(3) 4(3)	12	14(3)
	13	
1? 4? 7? 12(3)	14	16(3)
	15	16(3)
14(3) 15(3)	16	

Antes de pasar a ver las aplicaciones de estos diagramas, se harán tres puntualizaciones. En primer lugar, Hoey ha diseñado esta metodología de análisis de conexiones, vínculos y redes de vínculos sobre la base de textos no-narrativos. En segundo lugar, cuando se produce una *coocurrencia* de repeticiones entre oraciones, estas oraciones quedan relacionadas más allá de las conexiones cohesivas individuales que existen entre las palabras. Por último, las conexiones y los vínculos se dan tanto entre oraciones contiguas como entre oraciones separadas y son el resultado de los patrones léxicos propios de cada texto.

4.6. Aplicaciones de la metodología de Hoey

Hoey (1991) apunta algunas de las aplicaciones que puede tener la identificación de *redes de vínculos* en la mejor comprensión de los textos. Antes de nada, es necesario analizar los diagramas topológicos correspondientes a un texto y la tabla con las coordenadas que indican el número de vínculos con oraciones previas y posteriores.

4.6.1. Fragmentación de textos y didáctica de las lenguas

Por una parte, se pueden identificar las oraciones centrales y las oraciones marginales de un texto. Las oraciones centrales son aquellas que contienen un número de vínculos

con otras oraciones mayor de lo que se considera normal para un tipo textual determinado. A raíz de esta identificación y tras suprimir las oraciones marginales, se pueden obtener resúmenes coherentes e inteligibles del texto original. Esto es posible porque, como dijimos en las páginas 116-117, si se yuxtaponen dos oraciones vinculadas, aunque estas no sean contiguas en el texto analizado, el lector las reconoce como mutuamente relevantes y coherentes entre sí.

Por otra parte, estas redes léxicas y la dirección que siguen los vínculos permiten identificar oraciones que inician o finalizan las distintas secciones en las que se divide un texto (*topic-opening, topic-closing sentences*).

El interés por la fragmentación de textos en función de la cohesión léxica ha sido continuado por algunos discípulos de Hoey como Berber Sardinha (1994a, 1994b, 1997a) y Renouf y Collier (1995), como veremos en el apartado 4.8.1. Berber Sardinha también ha explorado la distancia que existe entre las oraciones vinculadas (*bonding distance*). Esta distancia puede ser de muchas oraciones, e incluso de capítulos como indican Phillips (1985) y Hoey (1991). La opinión de Berber Sardinha (1995) al respecto es que la vinculación es más frecuente entre oraciones no contiguas y que, probablemente, esta distancia sea sensible al tipo textual.

Otra aplicación de la metodología de Hoey la encontramos en la didáctica de las lenguas y en los aspectos cognitivos relacionados con la lectura y la escritura. Sobre la hipótesis de que la comprensión lectora se basa en los vínculos oracionales que establece el lector, Hoey incide en la necesidad de explotar en el aula de idiomas, entre otras, una forma de lectura no-lineal que ayude al lector a descubrir los vínculos entre oraciones y a plasmarlos por escrito. Según este autor, la explotación de esta técnica posiblemente aumenta la comprensión, ya que las oraciones más vinculadas son las más relevantes para la interpretación de otras. Además, sostiene que esta técnica despierta la creatividad. En España, Díez Prados (1999) y Suau Jiménez (1998) han seguido esa línea de investigación.

4.6.2. Vinculación e intertextualidad (Hoey 1995)

Hoey (1995, 1996) explora la relación que existe entre el fenómeno de vinculación y la intertextualidad. La intertextualidad es un fenómeno que explica el hecho de que

podamos relacionar unos textos con otros y de que existan distintos *géneros*. Estos géneros no sólo sirven de signos de identidad de una determinada comunidad de discurso, sino que también le aportan vitalidad. Esta intuición de una posible relación entre *vinculación* y *género* ya está presente en Hoey (1991: 190).

Antes de adentrarse en el fenómeno de la intertextualidad, Hoey (1995) busca una base psicológica a la capacidad de identificar en los textos patrones léxicos y vínculos entre oraciones. Estos vínculos han sido identificados en textos en lengua inglesa, portuguesa, española, francesa y alemana. Su razonamiento está guiado por las siguientes preguntas:

- a) ¿El lector tiene realmente consciencia de los vínculos entre oraciones? ¿Son estos vínculos un constructo que sólo puede descubrir el analista?
- b) Si el lector realmente construye estos vínculos ¿qué nos dice este fenómeno sobre el modo en que procesamos el lenguaje en nuestra mente?

4.6.2.1. La percepción de vínculos por parte del lector

Aunque no se ha realizado un estudio experimental que demuestre que los lectores son conscientes de los vínculos que se establecen entre oraciones, Hoey aporta como argumento el hecho de que se hayan podido reconstruir fragmentos de la Biblia y de otros textos literarios antiguos a partir de fragmentos separados de texto. Los estudiosos de estos textos han sido capaces de establecer conexiones entre oraciones que no son contiguas e incluso que pertenecen a partes distintas del texto.

Otro posible indicio de la percepción de vínculos por parte del lector lo podemos encontrar en el ejercicio que Eugene Winter diseñó en 1970 para comprobar que los hablantes de inglés comparten una intuición sobre la construcción del discurso. Este ejercicio aparece descrito en Hoey (1983: 4-7). En el mismo, Winter presentaba a sus alumnos una serie de diez oraciones pertenecientes a un discurso, que habían sido desordenadas y les pedía que reconstruyeran el texto original en 20 minutos. La mayoría de los alumnos fueron capaces de reconstruir de una forma bastante aproximada el texto original. Hoey (1983: 5) afirma que los alumnos estaban utilizando un conocimiento instintivo que Winter (1976) llama *consenso*. Este consenso se basa en la habilidad de

establecer conexiones entre oraciones mediante distintos recursos como la repetición, la anáfora y la catáfora.

Por último, es probable que los lectores establezcan vínculos puesto que, al leer detenidamente un texto, volvemos varias veces a oraciones previas e intentamos relacionarlas con oraciones anteriores y posteriores. Este tipo de lectura no es lineal y permite que el lector detecte qué oraciones son relevantes para la interpretación de otras. Curiosamente, las oraciones relevantes de un texto son las que tienen más contenido en común, las que tienen más vínculos entre ellas.

4.6.2.2. Creación de vínculos y procesamiento del lenguaje

Para intentar responder cómo se crean los vínculos, conjugaremos la opinión de Hoey (1983, 1991, 1995) y la de estudiosos del procesamiento de la información y del lenguaje (Baddeley 1992; Bell 1991; Givón 1998a, 1998b; Faber 1998a).

Como ya se señaló en el apartado sobre el modelo cognitivo de Bell (1991), los procesos mentales para interpretar un texto no son lineales sino que son interactivos y suponen un movimiento tanto de avance como de retroceso a través del texto.

Cuando empezamos a leer un texto, nuestra memoria a corto plazo nos permite concentrar nuestra atención en un máximo de tres oraciones. En la memoria a corto plazo, hay un módulo denominado memoria de trabajo (*working memory*) que va a permitir realizar las operaciones mentales necesarias para procesar cada oración (Baddeley 1992).

Para analizar de forma simplificada los procesos mentales que pueden explicar el fenómeno de vinculación, consideraremos las primeras líneas del quinto texto del Apéndice II.

(18) ONCOLOGÍA

El triunfo del taxol

El fármaco pasa a primera línea en el tratamiento del cáncer de ovario

Ha dejado de ser un segundón. El taxol (paclitaxel en su nombre genérico), el avance quimioterápico frente al cáncer más importante de la última década, ha

adquirido hoy la categoría de producto oncológico de primera línea. Con la publicación en el número de este jueves del New England Journal of Medicine de un trabajo en el que se demuestra la superioridad de una combinación de taxol y cisplatino sobre la quimioterapia estándar del cáncer de ovario, los oncólogos tendrán que pensar en él desde el principio cuando se enfrenten a un tumor maligno de la gónada.

En el trabajo, realizado con el patrocinio de los Institutos Nacionales de la Salud de EEUU, se comprobó el resultado de tratar inicialmente a 386 mujeres con cáncer del ovario con dos tipos de quimioterapia diferentes. La mitad de ellas - elegidas de forma aleatoria- recibió cisplatino y ciclofosfamida y la otra mitad cisplatino y taxol. Los resultados al cabo de cuatro años de seguimiento clínico fueron significativamente mejores -tanto en supervivencia como en número de recidivas- para el grupo que recibió inicialmente el taxol que para las féminas tratadas de forma estándar.

Al empezar a leer el texto, el lector se concentra en el título. Con el fin de procesarlo, su memoria de trabajo conecta con la memoria semántica, que, como dijimos en el apartado 2.1., es un compartimento permanente de la memoria a largo plazo. La memoria semántica dispone de un repositorio con las palabras más frecuentes, otro con las estructuras sintácticas más frecuentes y, según Hoey, de un archivo donde están almacenadas las concordancias más frecuentes.

La memoria de trabajo busca en la memoria semántica las palabras y unidades poliléxicas del título con el fin de llegar a una interpretación preliminar del mismo. Si la búsqueda es infructuosa, entran en acción los mecanismos de interpretación descritos por Bell: el mecanismo de búsqueda léxica y el analizador sintáctico. Estos mecanismos son también activados por la memoria de trabajo, que es la que controla la interacción entre la memoria semántica y la memoria episódica.

Una vez que la memoria de trabajo ha llegado a una primera interpretación del título, las palabras plenas quedan almacenadas en la memoria a corto plazo en forma de líneas de concordancia en cuyo nodo central puede aparecer el término *taxol* o *cáncer de ovario*.

(19)

Oncología. El triunfo del **taxol**. El fármaco pasa a primera línea en el tratamiento del cáncer ovárico. tratamiento del **cáncer ovárico**.

Según Hoey, estas líneas de concordancia se activarán cada vez que aparezca en el texto el lexema que hace de nodo. Puede haber más de una línea de concordancia por oración, según las palabras plenas que se consideren como nodo.

Conforme el lector avanza en la lectura del texto, se van almacenando en la memoria a corto plazo más líneas de concordancia. Hoey (1995: 79) lo explica del siguiente modo:

[...] every time we encounter a sentence we sort the words in it into a set of on-going concordances for those words, thereby creating a multiple record of the sentence which is then capable of being accessed as information through those records.

Aún se desconoce cuántas palabras se almacenan en cada línea de concordancia. Los nodos de estas concordancias son las palabras que la memoria episódica a corto plazo ha reconocido como instancias de otras palabras del texto. Por ejemplo, cuando el lector ha acabado el primer párrafo, es probable que haya almacenado las siguientes concordancias, extraídas con el programa *Microconcord* de Oxford University Press.

(20)

```

1          ONCOLOGIA El triunfo del taxol El fármaco pasa a primera línea
2 ico Ha dejado de ser un segundón. El taxol (paclitaxel en su nombre genérico),
3 a superioridad de una combinación de taxol y cisplatino sobre la quimioterapia

1 primera línea en el tratamiento del cáncer ovárico Ha dejado de ser un segundó
2 el avance quimioterápico frente al cáncer más importante de la última década,
3 sobre la quimioterapia estándar del cáncer de ovario, los oncólogos tendrán
4 primera línea en el tratamiento del cáncer ovárico Ha dejado de ser un segundó

1 línea en el tratamiento del cáncer ovárico Ha dejado de ser un segundón. El
2 quimioterapia estándar del cáncer de ovario, los oncólogos tendrán que pensar

1 n de taxol y cisplatino sobre la quimioterapia estándar del cáncer de ovario,
```

Posteriormente, cuando el lector se encuentra de nuevo uno de los lexemas que hacen de nodo, por ejemplo, el lexema *taxol* del segundo párrafo, su memoria de trabajo lo busca en el archivo de concordancias, activando así las concordancias de otras

palabras de su radio colocacional: *cáncer*, *ovario* y *quimioterapia*. Se han señalado con un subrayado. Como resultado de esta operación, el lector reconoce las conexiones entre lexemas similares a *taxol* y entre los lexemas que aparecen en su cotexto, por ejemplo, *oncología* y *quimioterapia*. Si las conexiones entre los elementos de dos oraciones son más frecuentes de lo habitual, el lector establecerá un vínculo entre estas dos oraciones.

(21)

ONCOLOGIA El triunfo del **taxol** El fármaco pasa a primera línea en el
ico Ha dejado de ser un segundón. El **taxol** (paclitaxel en su nombre genérico), e
a superioridad de una combinación de **taxol** y cisplatino sobre la quimioterapia

primera línea en el tratamiento del **cáncer** de ovárico Ha dejado de ser un segund
sobre la quimioterapia estándar del **cáncer** de ovario, los oncólogos tendrán que
el avance quimioterápico frente al **cáncer** más importante de la última década,

línea en el tratamiento del cáncer **ovárico** Ha dejado de ser un segundón. El
quimioterapia estándar del cáncer de **ovario**, los oncólogos tendrán que pensar en
n de taxol y cisplatino sobre la **quimioterapia** estándar del cáncer de ovario, lo

El proceso que se acaba de describir no es el único que tiene lugar en la creación de conexiones entre oraciones. También son necesarias una serie de operaciones léxicas, sintácticas y discursivas. De estas operaciones, podemos mencionar la identificación de relaciones de sinonimia, o de correferencia, la habilidad para hallar el referente de un pronombre o de expandir una expresión léxica utilizando el conocimiento enciclopédico.

De todas formas, Hoey hace hincapié en los procesos que pueden explicar que, en los textos, las conexiones se establezcan entre las oraciones y no simplemente entre las palabras. Por este motivo, Hoey sostiene que el lenguaje se almacena casi sin procesar en forma de concordancias y que continuamente nuevas ocurrencias se añaden a estas conforme leemos y hablamos. Cuando el sistema que guarda en nuestra memoria las concordancias indica que una concordancia tiene más conexiones de lo normal, nuestra mente la interpreta en función de las líneas de concordancia con las que está relacionada. En consecuencia, este sistema permite que se emparejen nuevos enunciados con antiguos enunciados y que se creen conjuntos de concordancias interconectadas gracias a los procesos léxicos, sintácticos y discursivos arriba

mencionados. Estos procesos hacen que exista un mayor número de paralelismos entre las oraciones y permite que se emparejen unas con otras.

Esta hipótesis sobre el fenómeno de vinculación tiene como ventaja que aporta una explicación a otros fenómenos como la adquisición del lenguaje, la colocación, las citas, las bromas y la intertextualidad. También parece estar conforme con la idea de que aprendemos las palabras en contexto y que somos capaces de guardar en la memoria a largo plazo frases y oraciones que se han activado con frecuencia en la memoria a corto plazo. De esta forma, podemos pronunciarlas de una forma automatizada que reduce el esfuerzo a realizar por la memoria de trabajo.

Sin embargo, son perceptibles algunas inconsistencias en su explicación. En primer lugar, nos ha sorprendido la primacía que Hoey otorga a las líneas de concordancia, a pesar de que no haya incluido la colocación como recurso cohesivo.

Por otra parte, si Hoey pretende dar una explicación cognitiva a los procesos de vinculación, debería haber comentado algo sobre la teoría de marcos de Fillmore (1977) y haber citado algunos modelos del lexicón mental como el de Aitchison (1987, 1994) o el de Martín Mingorance (1998). Estos modelos explican cómo las palabras almacenadas en nuestro lexicón están relacionadas entre sí formando una red y cómo al utilizar una palabra se activan sus conexiones con otras palabras que aparecen con frecuencia en el mismo contexto o que tienen un significado o un significante parecido. Estas conexiones podrían ser la base de las líneas de concordancia a las que hace alusión Hoey. Asimismo opinamos que la investigación de Hoey podría haberse enriquecido con una mención a los sistemas de la memoria humana. De esta forma, podría haber encontrado una base neurológica al fenómeno de la vinculación.

4.6.2.3. Intertextualidad

Hoey sostiene que, si los procesos descritos en el apartado anterior se producen en la realidad, constantemente estamos almacenando líneas de concordancia independientemente de si estas proceden del mismo texto o no. En consecuencia, los vínculos que establecen los hablantes no son sólo entre oraciones del mismo texto, sino también entre textos producidos en fechas similares o extraídos de las mismas fuentes. Incluso, Hoey sostiene que ha encontrado vínculos entre textos que no tienen una fuente

común inmediata. Todas estas hipótesis están respaldadas por una argumentación coherente, por un estudio experimental previo (Abraham 1993) y por tres textos sobre astronomía.

Para apoyar la idea de intertextualidad entre textos extraídos de una misma fuente, Hoey argumenta que, cuando escuchamos o leemos un texto y posteriormente decidimos hablar o escribir sobre el tema del que trata, utilizamos léxico proveniente de dicho texto porque tenemos almacenadas en nuestra mente líneas de concordancia provenientes del mismo. De esta forma, establecemos vínculos no sólo con el texto que hemos procesado con anterioridad, sino con las palabras que lo codifican. De ahí que se puedan encontrar vínculos entre el texto de un escritor y los textos que consultó antes de escribirlo.

En cuanto a la vinculación entre textos que no tienen una fuente común inmediata o cuyas fuentes se basan en un texto muy remoto, Hoey ha constatado su existencia en diversos textos, aunque, claro está, la vinculación es menor. Su opinión al respecto queda clara en la siguiente cita:

If we indeed bond the sentences/utterances we write or speak with those we have received and are storing our mental concordances, then, we would expect the authors of the secondary sources to write sentences that bond with those in the primary sources and in turn would expect the authors of the newspaper articles to write sentences that bond with their secondary sources.

(Hoey 1995: 90)

Ante la posible objeción de que no es extraordinario que se utilicen las mismas palabras para hablar de un mismo tema, Hoey arguye que sí es llamativo que las oraciones que están vinculadas sean coherentes y pertinentes entre ellas.

Haciendo balance, la metodología de Hoey puede ser un instrumento valioso para encontrar una relación entre el fenómeno de la intertextualidad y los procesos mentales por los que almacenamos los enunciados con los que interactuamos al comunicarnos. El hecho de que podemos vincular unos textos con otros es una evidencia indirecta de la vinculación constante que hacemos entre nuestros enunciados y los que hemos escuchado con anterioridad.

Una consecuencia de este almacenamiento de colocaciones y expresiones previas en la memoria a largo plazo es que, al interactuar con los textos de una comunidad de discurso, inconscientemente vamos adoptando las convenciones y el sistema de valores de la misma y, en consecuencia, construimos y hacemos que se consolide esa comunidad de discurso (Bazerman y Paradis 1991). Hoey comparte esta idea cuando afirma (ibid: 90):

If one bonds with what one has heard and read, and if everyone else in the field is doing the same, then certain expressions are going to become typical of that field and to distinguish the well-read from the novice.

Hasta aquí se han repasado algunas aplicaciones interesantes de la metodología de Hoey. Se ha hecho hincapié sobre todo en la última, puesto que fundamenta la intertextualidad y la existencia de diversos géneros en las relaciones cohesivas.

4.7. Morris (1988), Morris y Hirst (1991)

Morris y Hirst sostienen que el contenido de un texto se crea a partir de *cadena léxica* que tratan sobre el mismo tema. A diferencia de los autores que hemos citado, su perspectiva se encuadra dentro de la lingüística computacional. En consecuencia, dan menos importancia a la cohesión como fenómeno lingüístico interoracional y conceden la primacía al contenido semántico del texto.

En Morris y Hirst (1991: 22-23) se definen las cadenas léxicas como “sequences of related words (...) spanning a topical unit of the text”. Estas cadenas de palabras que tratan sobre el mismo tema pues, están formadas por lexemas semánticamente relacionados en un texto. Por ejemplo, si un texto contiene las palabras *cáncer* y *melanoma* estas pertenecerían a una misma cadena léxica porque *melanoma* es un tipo de *cáncer*.

Estos autores crean un algoritmo para la identificación de estas cadenas, aunque realizan el análisis de forma manual, buscando las palabras plenas susceptibles de formar parte de una cadena y la relación entre ellas en un tesoro, en concreto, el *Roget's Thesaurus*. Se asigna a cada lexema un número que lo asocia a una categoría

del tesauo. El tesauo permite que se computen no sólo las relaciones entre palabras idénticas, sino también entre palabras que pertenecen a la misma categoría semántica o entre palabras pertenecientes a una categoría distinta pero que están directamente relacionadas. Por ejemplo, estos autores incluyen dentro de la misma cadena cohesiva las formas *cow*, *cows* y *sheep*, pero no computan la relación indirecta que *cow* y *wool* establecen mediante *sheep*³³.

Asimismo, también especulan sobre si se debe incluir dentro de la misma cadena cohesiva lexemas que están distantes en el texto y deciden establecer una distancia máxima de tres oraciones. Si la distancia es mayor, se computa lo que Morris y Hirst denominan *chain return*.

Así pues las cadenas léxicas son conjuntos de palabras de un texto que están relacionadas por vínculos léxicos y por proximidad. El *Roget's Thesaurus* (en adelante, *Roget*) sirve de base de datos para establecer vínculos léxicos o relaciones que, según Morris y Hirst, pueden ser de varios tipos. Las relaciones de *identidad* (*identity*) son las que se establecen entre lemas o inflexiones de la misma palabra (*aberrations-aberration*). Las relaciones *categoriales* (*category*) vinculan miembros que pertenecen a una misma categoría semántica del tesauo, por ejemplo, la relación entre *aberration* y *abnormalities*, que se establece mediante la categoría ABNORMALITY, que sirve de hiperónimo de ambas. Las relaciones de *grupo* (*group*) se establecen entre palabras pertenecientes a categorías vecinas dentro de un mismo *grupo de categorías*, teniendo en cuenta que las entradas del tesauo aparecen asociadas en grupos de categorías jerárquicamente organizados. Esta relación es la que existiría entre dos lexemas pertenecientes al grupo **DEFICIENCY: prognosis**, asociado a la categoría PATHOLOGY, y *disease*, asociado a la categoría DISEASE.

Por otra parte, las palabras contenidas en una categoría del *Roget* a veces hacen referencia a otras categorías. Por eso, se establece una relación entre dos palabras siempre que la entrada de una de ellas haga alusión a la entrada que contiene la otra palabra. A esta relación, Morris y Hirst la denominan *one-point relation*. Por último, Morris y Hirst proponen otra categoría en la que las palabras en cuestión contienen

³³ Estos autores denominan la distancia entre los miembros de distintas categorías transitividad y fijan la transitividad en un vínculo. Por eso excluyen la relación de dos vínculos entre *cow* y *wool*.

entradas que apuntan hacia una tercera categoría. Este último tipo de vínculo genera relaciones muy difusas que pueden distorsionar las cadenas léxicas ya identificadas.

Finalmente, esta aproximación postula que cada cadena cohesiva representa una porción de la estructura cohesiva de un documento. Por este motivo, se afirma que la organización de las cadenas léxicas de un documento refleja en cierto modo la estructura discursiva de ese documento.

Si bien es verdad que los creadores de la noción de cadena léxica no la aplicaron al análisis automatizado de textos, sus investigaciones han servido de base a una serie de estudios cuyo objetivo es la identificación automática de la cohesión con la ayuda de un tesaurus. La utilización de un tesaurus tiene la ventaja de permitir un análisis más completo puesto que, además de la cohesión entre palabras idénticas (repetición léxica), el ordenador también es capaz de reconocer la que existe entre palabras semánticamente relacionadas.

4.8. Aplicaciones informáticas para el reconocimiento automático de la cohesión

En la década de los noventa, una serie de autores han estudiado el modo de reconocer la cohesión de forma automática. Las diversas aproximaciones siguen principalmente dos tendencias, la de Hoey y la de Morris y Hirst. La primera se deriva de los estudios de Hoey y, ante la dificultad inherente al estudio de la cohesión, limita la cohesión a la repetición de formas léxicas. La segunda tendencia además trata de reconocer la cohesión derivada de relaciones conceptuales. Seleccionamos a continuación aquellas que han servido de base a programas informáticos que, en la mayoría de los casos, están aún en periodo de prueba.

4.8.1. Aproximaciones basadas en la repetición léxica

Entre las aplicaciones de la metodología de Hoey (1991), Renouf y Collier (1995) computan las conexiones y vínculos dentro de textos periodísticos expositivos con la intención de generar resúmenes a partir de los mismos. En este estudio sólo se reconoce la repetición léxica simple (*parliament – parliament*), limitación presente en el

programa informático derivado, *Automatic Abridgement System*³⁴, cuya versión comercial está preparándose en la Universidad de Liverpool.

En esta misma universidad, Berber Sardinha (1997a) ha estudiado cómo contribuye la cohesión a la organización del texto, en concreto, cómo la cohesión puede ayudar a identificar las secciones en las que se divide un texto. Para ello, diseña un programa de *segmentación textual*³⁵, con el que analiza un corpus de 300 textos pertenecientes a tres géneros: el artículo experimental, el informe comercial (*business report*) y la entrada de enciclopedia. Los resultados obtenidos confirman que las oraciones que delimitan las secciones tipográficas de un texto son las más relevantes desde el punto de vista de la cohesión léxica. Berber Sardinha sostiene que estas oraciones son las que están vinculadas con más oraciones, y no las que tienen un mayor número de conexiones con una determinada oración. Para justificar el hecho de que el programa sólo reconoce instancias de repetición léxica simple (*important – important*) y compleja (*important – importance*), Berber Sardinha remite a Benbrahim (1996), donde se estima que, de todos los recursos cohesivos identificables en un texto, entre un 92 y un 97% se fundamentan en estos dos tipos de repetición.

La investigación de Mohamed Benbrahim y Khurshid Ahmad se encuentra en la línea divisoria que separa los programas citados y los que reconocen algo más que la repetición del mismo lexema en un texto. Benbrahim y Ahmad (1994) aplica la metodología de Hoey (1991) con miras a la producción automática de resúmenes y bancos terminológicos. Este estudio cristaliza en el programa de análisis textual *Telepattan*, que identifica las conexiones y vínculos de un texto y las representa gráficamente³⁶. En Benbrahim (1996) se ofrece un análisis detallado del número y tipo de conexiones que aparecen en cinco textos académicos en inglés y se extraen las conclusiones expuestas en el párrafo anterior. El cómputo de las instancias de paráfrasis simple y compleja se realiza en función de las categorías del tesoro del proyecto *Wordnet* (Miller 1991).

³⁴ Comunicación personal con Antoinette Renouf (ajrenouf@liv.ac.uk).

³⁵ El programa se denomina *Words* aunque, por falta de financiación, su autor aún no ha procedido a su distribución.

³⁶ Mohamed Benbrahim (mohamed@ceanet.po.my) está programando en C++ una versión que se distribuirá de forma gratuita entre investigadores.

Telepattan ha servido de prototipo de dos programas de gestión de terminología desarrollados también por la Universidad de Surrey: *Sebriba* y *System Quirk*. *Sebriba*³⁷ genera automáticamente un corpus buscando en la red aquellos textos cuyo contenido se ajuste mejor a la palabra clave de búsqueda e identifica los términos del corpus resultante. *System Quirk* permite elaborar córpora textuales, aunque a partir de las páginas de Internet determinadas por el usuario. Asimismo es de utilidad en la gestión de córpora textuales, ya que dispone de una tipología textual que clasifica los textos en función de siete atributos: *texto*, *autor*, *publicación*, *lengua*, *dominio*, *derechos de autor* y *textos relacionados*. Por último, *System Quirk* también extrae términos. Por este motivo, la investigación que sustenta este proyecto se centra en la actualidad en el estudio de unidades fraseológicas y la elaboración de ontologías específicas para un determinado campo del saber³⁸.

4.8.2. Aproximaciones que incluyen relaciones conceptuales

Dado que la cohesión también se construye a partir de relaciones semánticas, el concepto de *cadena léxica* ideado por Morris y Hirst (1988, 1991) está presente en una serie de estudios que pretenden la identificación de la cohesión mediante un ordenador dotado de tesoro (Green 1997; Ellman 1998, 1999; Ellman, Klincke y Tait 1998).

Green (1997)³⁹ utiliza las nociones de *cadena léxica* y *similitud semántica* en la creación automática de vínculos de hipertexto entre las palabras de un artículo periodístico o entre distintos textos periodísticos. La base de datos léxica de la que se sirve es *Wordnet*.

Ellman (1997, 1998, 1999) crea un generador de cadenas léxicas para optimizar la recuperación de información en Internet mediante búsquedas basadas no sólo en la forma de las palabras, sino también en su significado, de ahí que incorpore un tesoro. Este sistema, que su autor denomina *Hesperus*, analiza las cadenas léxicas de cualquier

³⁷ Los autores de *Sebriba* son Steve Collingham (S.Collingham@surrey.ac.uk) y K. Ahmad (k.ahmad.mcs@surrey.ac.uk).

³⁸ En el sitio de Internet de *System Quirk* [<http://www.mcs.surrey.ac.uk/SystemQ>] también hay una aplicación que elabora en poco tiempo un resumen de cualquier texto que se introduzca en el recuadro destinado a tal efecto.

³⁹ Documento de Internet disponible en <http://sjgreen.mpce.mq.edu.au/~sjgreen/Documents/sjgthesis.ps.gz>.

texto y otorga un perfil genérico del documento (*Generic Document Profile*) que permite comparar la similitud semántica entre ese texto y otros. A partir de textos prototípicos (*exemplars*) de un determinado tema, el sistema elabora el perfil correspondiente y activa un motor de búsqueda que encontrará textos cuyo perfil indique una similitud conceptual.

Al igual que Morris y Hirst, Ellman utiliza el *Roget's Thesaurus* en lugar de *Wordnet* (Miller 1991). Justifica esta elección aduciendo que el *Roget* computa un mayor número de relaciones y relaciones más simples. Además, su estructura es más sencilla y equilibrada, 1024 categorías, en contraposición a las 91595 de *WordNet 1.5*, lo cual permite una puesta en marcha más eficaz.

Hesperus extrae el conocimiento léxico de una versión electrónica que Ellman confecciona a partir de dos versiones del *Roget*⁴⁰. Este tesoro contiene bastante terminología médica, en parte porque su autor, Peter Mark Roget, era un médico que pretendía aumentar su capacidad expresiva⁴¹.

Aunque se constatará esto en el análisis, veamos ahora qué encontramos en el índice bajo la palabra *cancer*. Esta palabra nos remite a *carcinosis* 651 n, es decir, a la categoría ILL - HEALTH - DISEASE, que comprende una serie de sustantivos (n). Bajo *carcinosis* aparecen los siguientes lexemas:

(22)

Epithelioma, cancer; neoplasm, growth; tumour, indolent t., benign t.; malignant tumour, cancerous growth; melanoma, black cancer.

[*Roget's Thesaurus of English Words and Phrases* (1982), Harmondsworth: Penguin books: 251]

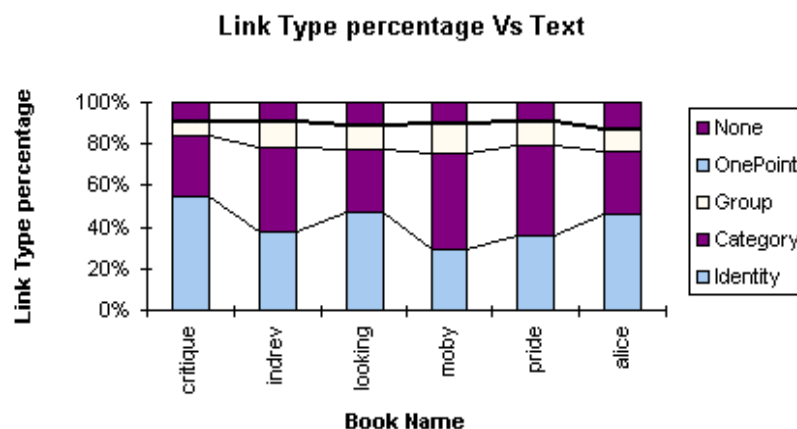
Teniendo en cuenta que se mencionan otras enfermedades en este apartado y que los síntomas y el tratamiento constituyen apartados por sí mismos, parece que el *Roget* puede ser útil en la detección automática de relaciones conceptuales en textos sobre oncología, a pesar de ser un tesoro de la lengua general. De hecho, con esta base de conocimiento léxico, *Hesperus* halla la relación entre *carcinoma* y *melanoma*.

⁴⁰ La primera, publicada en 1911, consta de 56000 palabras y ha sido utilizada por el proyecto Gutenberg. La segunda cuenta con 250000 palabras.

⁴¹ La primera edición, publicada en Londres en 1852 se llamaba *Thesaurus of English Words and Phrases classified and arranged so as to facilitate the expression of ideas and assist in literary composition*.

Por otra parte, la etiqueta que *Hesperus* concede a los vínculos es la de ID (*identity*), CAT (*category*), GRP (*group*) y ONE (*one-point*). De estos tipos de vínculos, en contra de lo esperado, la relación ID es la más frecuente. En Ellman (1998) se afirma que el 60-80% de las cadenas léxicas se deben a la repetición de términos. A diferencia de Morris y Hirst (1991), el programa de Ellman sólo computa como relación ID la existente entre formas idénticas o entre sustantivos que cambian sólo respecto al número. Quedan por tanto fuera de esta categoría las inflexiones de un mismo verbo (*includes-include*), que cuenta como CAT. Siguen en importancia a la relación ID, las denominadas CAT y GROUP. La categoría ONE es muy inusual. Al reconocer la preponderancia de relaciones basadas en repetición léxica, Ellman coincide con lo defendido en Benbrahim (1996) y Berber Sardinha (1997a). Veamos en la siguiente tabla el porcentaje que corresponde a cada tipo de vínculo en seis libros analizados por Ellman, que son los que aparecen en el eje horizontal:

DIAGRAMA 8: Porcentaje de los principales tipos de conexión en seis libros (Ellman, 1999)⁴²



Critique:	<i>The Critique Of Pure Reason</i> de Immanuel Kant
Indrev:	<i>Lectures on The Industrial Revolution in England</i> de Arnold Toynbee
Looking:	<i>Through The Looking Glass</i> de Lewis Carroll
Moby:	<i>Moby Dick</i> de Herman Melville
Pride:	<i>Pride And Prejudice</i> de Jane Austen
Alice:	<i>Alice's Adventures In Wonderland</i> de Lewis Carroll

⁴² Disponible en <http://osiris.sunderland.ac.uk/~cs0jel/web1.gif>.

Aunque describiremos los datos que proporciona *Hesperus* sobre los textos en el apartado de metodología, por el momento, anticipamos que vincula palabras separadas por un máximo de 500 palabras. También enumeramos los pasos que sigue *Hesperus* para elaborar el *perfil genérico del documento*, que, en adelante, denominaremos *perfil*.

1. Lematización y anotación del texto.
2. Supresión de las palabras forma (*stopwords*).
3. Búsqueda de cada lexema en la versión electrónica del tesoro.
4. Creación de cadenas léxicas.
5. Elaboración del *perfil*.

De todos los programas que hemos revisado, creemos que *Hesperus* es el que mejor se ajusta al estudio de la cohesión en textos extensos, de ahí que este sea el programa utilizado en nuestro análisis. *Hesperus* tiene la ventaja de disponer de un tesoro, lo cual permite reconocer relaciones semánticas entre palabras. Además, proporciona datos relevantes sobre los textos, tales como el número de palabras plenas, el tipo de relación cohesiva y el porcentaje de activación de los conceptos en el texto.

Aunque *Hesperus* aún no está disponible, nuestra investigación ha sido posible gracias a la desinteresada ayuda de Jeremy Ellman, quien sometió parte de nuestro corpus de textos sobre oncología al programa *Hesperus* y nos envió los resultados como fichero comprimido.

4.9. La cohesión desde la Teoría Sentido Texto

4.9.1. Consideraciones generales

La Teoría Sentido Texto (*Meaning Text Theory*), propuesta en Mel'cuk y Zholkovskij (1970, 1988), es un modelo lingüístico que pretende relacionar los textos con sus representaciones semánticas y presentar estas relaciones de una forma rigurosa, ya que, al fin y al cabo, en la lengua las relaciones léxicas están institucionalizadas.

La TST ha sido aplicada principalmente en lexicografía, en concreto, en el *Explanatory Combinatorial Dictionary* (Mel'cuk *et al.* 1984), que es un diccionario formal cuyo objetivo es especificar el universo sintáctico, semántico y léxico de las unidades léxicas. Asimismo, según Saint-Dizier y Viegas (1995: 27), se ha tenido muy en cuenta este modelo en la lingüística computacional. Desde la traducción, Faber (1999a) y Tercedor (1999: 149-159) también destacan la validez de este modelo para la descripción léxica y terminográfica en dos lenguas, estableciendo como *interlingua* las funciones léxicas propuestas dentro de la TST.

Las 53 funciones léxicas propuestas inicialmente⁴³ son un intento de sistematizar las relaciones semánticas y sintácticas que se dan en los textos. Las funciones léxicas, de ahora en adelante FL, son relaciones léxicas institucionalizadas según la definición de Wanner (1996: 1):

A lexical relation is institutionalized if it holds between two lexical units L1 and L2 and L2 has the following characteristics: if L1 is chosen to express a particular meaning M, its choice is predetermined by the relation of M to L2 to such an extent that in case M and L2 are given, the choice of L1 is a language-specific automatism.

Como ilustración, Wanner cita la relación existente entre *aircraft – crew*, *sheep – flock*, *influence – exert* y *attention – pay*. Esta relación se manifiesta de un modo sistemático en la comunicación entre los miembros de una comunidad de discurso.

La representación formal de las FL es similar a la de las funciones matemáticas. Mel'cuk (1996: 39-40) define *función léxica* (f) como la función o correspondencia que asocia una determinada expresión léxica (X), que se denomina *argumento* o *palabra clave*⁴⁴, con un conjunto ($\{Y_i\}$) de expresiones léxicas—*valor* de f— que expresan, al depender de X, un significado específico asociado con f:

⁴³ A estas se irán sumando algunas más. Mel'cuk (1996) propone 64 funciones léxicas simples y sugiere una serie de funciones léxicas complejas derivadas de estas.

⁴⁴ Otros términos para denominar al *argumento* (*palabra clave*) y al *valor* son los de *base* y *colocador* (Hausmann 1985: 119; Irsula 1991: 160; Tercedor 1999: 90). Opinamos que estos dos últimos hacen alusión a relaciones sintagmáticas (fraseología). Debido a que nuestro estudio va a estudiar no sólo estas sino también las relaciones paradigmáticas, creemos oportuno utilizar la nomenclatura de *palabra clave* y *valor*.

$$(23) f(X) = \{Y_i\}$$

$$f(\text{palabra clave}) = \{\text{valor}\}$$

$$\text{Anti}(\text{malignant}) = \text{benign}$$

En este ejemplo, la función **Anti** (*Antónimo*) relaciona el lexema *malignant* (palabra clave) con un término de significado opuesto, el lexema *malignant* (valor).

Según Mel'cuk (1996: 41), las expresiones léxicas que constituyen el valor de una determinada función léxica pueden ser de cinco tipos: lexema, frase hecha, ejemplo libre⁴⁵ (*free phrase*), base de palabra compuesta y morfema derivativo.

TABLA 11: Tipos de Funciones léxicas (Mel'cuk 1996: 41)

Type	LF example
a lexeme	Magn (smoker) = heavy
a phraseme	Magn (to smoke) = like a chimney Magn (to rain) = cats and dogs
a free phrase	Magn (to smoke) = without stopping lighting up a cigarette from the other, ...
a compounding stem	Magn (smoker) = chain- [chain-smoker]
a derivational means	S ₁ (to smoke) = ~ -er [smoker]

Como este modelo ya ha sido aplicado a la representación del conocimiento terminológico sobre oncología por Tercedor (1999: 194), en lugar de incluir un listado de FL referentes a la lengua general, se incluirá el repertorio de funciones terminológicas que esta autora elabora a partir de las FL de Mel'cuk.

A ₀	adjective derived from keyword A ₀ (tumor/tumour) = <i>tumoral</i>
A ₁ , A ₂	typical adjective for numbered participant A ₁ (tumor/tumour) = <i>malignant</i> A ₂ (tumor/tumour) = <i>benign</i>

⁴⁵ Proponemos esta traducción para *free phrase* siguiendo a Rodríguez Reina (2000: 326), que define el *ejemplo libre* como un "tipo de frases o combinaciones de palabras que no se someten a las leyes de frecuencia de aparición ni a reglas de colocación".

Able ₁ , Able ₂	ability of numbered participant Able ₁ (<i>tumour</i>) = <i>tumorous</i>
Adv ₀	adverb from keyword Adv ₀ (<i>cancer</i>) = <i>oncogenically (growth)</i>
Adv ₁ , Adv ₂	adverb from numbered participant Adv ₁ (<i>treatment</i>) = <i>innovative</i>
Anti	antonym (exact or near) Anti (<i>malignant</i>) = <i>benign</i>
Bon	standard praise for keyword Bon (<i>growth</i>) = <i>normal</i>
Caus	cause Caus ₁ (<i>cancer</i>) = <i>smoking</i>
Centr	center of Centr (<i>cell</i>) = <i>nucleus</i>
Cont	continue Cont (<i>grow</i>) = <i>keep, keep on</i>
Contr	non-antonymic contrast Contr (<i>apoptosis</i>) = <i>necrosis</i>
Conv _{ijk}	conversive (opposite where participants switch roles) Conv (<i>attack</i>) = <i>suffer an attack</i>
Culm	culmination of Culm ₁ (<i>cancer</i>) = <i>invasion</i> Culm ₂ (<i>cancer</i>) = <i>metastasis</i>
Degrad	degradation of (dimension can be specified: color, movement, etc) Degrad (<i>cancer: evolution</i>) = <i>stages</i>
Epit	standard epithet (representing a part of keyword) Epit (<i>cancer</i>) = <i>advanced</i>
Excess	excessive functioning of (dimension can be specified as in Degrad) Excess (<i>cancer</i>) = <i>metastasis</i>

Fact _{0,1,2}	verb meaning “the realization of” with the keyword as grammatical subject and the participants as objects Fact ₁ (<i>diagnosis</i>) = <i>confirm</i>
Figur	metaphor Figur (<i>cancer</i>) = <i>war</i>
Fin	stopping of Fin (<i>cancer</i>) = <i>cure</i> Fin (<i>tumor</i>) = <i>surgery</i>
Func _{0,1,2}	verb which takes the keyword as subject of first participant, second participant, etc. Func ₁ (<i>diagnosis</i>) = <i>includes CT scans</i>
Gener	generic word Gener (<i>MR</i>) = <i>diagnostic test</i>
Germ	the core of Germ (<i>problem</i>) = <i>crux</i>
Imper	the command associated with Imper (<i>care</i>) = <i>Watch out!</i>
Involv	verb meaning non-participant involvement Involv (<i>scent</i>) = <i>fill</i>
Incep	the beginning of Incep (<i>cure</i>) = <i>treat</i>
Labor _{ij}	verb which takes the participants <i>ij</i> as subject and object and the keyword as secondary object Labor ₁₂ (<i>esteem</i>) = <i>hold</i> (i.e., holds <i>y</i> in esteem)
LabReal _{ij}	verb meaning “the realization of” with the first two participants as subject and object and the keyword as a secondary object (a combination of Labor and Real) LabReal ₁₂ (<i>mind</i>) = <i>bring to</i> (x brings <i>y</i> to mind)
Liqu	the elimination of Liq (<i>cancer</i>) = <i>eradicate</i>
Loc _{in}	preposition for “in”
Loc _{ab}	preposition for “from”
Loc _{ad}	preposition for “to”

	$\text{Loc}_{\text{in}}(\text{cancer}) = \text{in}$
Magn	intensity (can specify dimension of quantity or duration) $\text{Magn}_1(\text{treatment: time}) = \text{duration}$ $\text{Magn}_1(\text{treatment: amount}) = \text{dose}$
Manif	is manifest in, with the keyword as subject $\text{Manif}(\text{small-cell lung cancer}) = \text{lung}$
Minus	less of $\text{Minus}(\text{process}) = \text{revert}$
Mult	regular aggregate of $\text{Mult}(\text{cell}) = \text{tissue}$
Nocer	to harm, injure, or impair $\text{Nocer}(\text{tissue}) = \text{affect}$
Obstr	to function with difficulty $\text{Obstr}(\text{organs}) = \text{malfunction}$
Oper _{1,2,3}	verb which takes numbered participant as subject and keyword as object $\text{Oper}_1(\text{cancer}) = \text{develop}$ $\text{Oper}_2(\text{cancer}) = \text{suffer from}$
Perm	permit or allow $\text{Perm}(\text{treatment}) = \text{tolerate}$
Plus	more of $\text{Plus}(\text{cancer}) = \text{metastasis}$
Pos _{1,2,3}	positive attributes of numbered participants $\text{Pos}_1(\text{cancer}) = \text{curable}$
Pred	copula for nouns and adjectives $\text{Pred}(\text{cancer}) = \text{to develop}$
Propt	preposition for “because of” $\text{Propt}(\text{cancer}) = \text{as a result of}$
Prox	to be on the verge of $\text{Prox}(\text{tumor}) = \text{precancerous}$

Qual _{1,2,3}	Highly probable qualities of numbered participants Qual ₁ (<i>cancer</i>) = <i>mortal or fatal</i> Qual ₂ (<i>cancer</i>) = <i>curable</i> Qual ₃ (<i>cancer</i>) = <i>preventive</i> Qual ₄ (<i>cancer</i>) = <i>detectable</i>
Real _{1,2,3}	verb meaning to realize with keyword as object and numbered participants as subject Real ₁ (<i>ambition</i>) = <i>realize</i>
S ₀	noun for keyword S ₀ (<i>immune</i>) = <i>immunity</i>
S _{1,2,3}	typical noun for numbered participant S ₁ (<i>cancer</i>) = <i>victim</i> S ₂ (<i>cancer</i>) = <i>patient</i>

Según Tercedor, las FL denominadas con la letra **S** y un subíndice son útiles para detallar los instrumentos y pruebas utilizados en el diagnóstico y el tratamiento del cáncer, la localización del tumor, el medio y el modo, junto con el resultado. Estas funciones pueden combinarse entre sí para una descripción exhaustiva de los conceptos.

S _{instr}	typical instrument S _{instr} (<i>cancer of the stomach</i>) = <i>endoscope</i>
S _{loc}	typical location S _{loc} (<i>endocrinal cancer</i>) = <i>pancreas</i>
S _{med}	typical means S _{means} (<i>drug</i>) = <i>oral</i> ⁴⁶
S _{mode}	typical mode S _{model} (<i>cancer</i>) Lique ₁ = <i>chemotherapy</i>
S _{res}	typical result S _{res} (<i>apoptosis</i>) = <i>cell death</i>
Sing	one instance of Sing (<i>treatment</i>) = <i>course</i>

⁴⁶ Tercedor utiliza esta función para designar factores de riesgo, que opinamos que están relacionados con S_{cause}.

	Sing (<i>tratamiento</i>) = <i>pauta</i>
Son	to emit typical sound Son (<i>difficult breathing</i>) = <i>whistling</i>
Sympt	to be a physical symptom of Sympt (<i>small-cell lung cancer</i>) = <i>coughing up blood</i>
Syn	synonym and near synonym Syn (<i>cancer</i>) = <i>malignant neoplasm</i>
V ₀	verb for keyword V ₀ (<i>metastasis</i>) = <i>metastasize</i>
Ver	true, correct, or proper Ver (<i>treatment</i>) = <i>effective</i>

En 6.6. *Funciones léxicas en la generación de textos cohesivos*, clasificaremos estas funciones de acuerdo con diversas aproximaciones dentro de la TST y con la forma en que se crea cohesión.

4.9.2. Aproximaciones a la cohesión

En la TST hemos encontrado una gran precisión en la descripción semántica de los lexemas y algunas alusiones a la cohesión (Alonso *et al.* 1995: 355-365; Lee 1996: 299-306; Mel'cuk 1996: 95-96) pertinentes para nuestra investigación, a pesar de que no logren explicar el fenómeno cohesivo en textos completos. Sus ejemplos sólo implican una relación semántica unidireccional o bidireccional entre dos oraciones pertenecientes al mismo párrafo, cuando, como se ha visto en este capítulo, las relaciones semánticas normalmente son multidimensionales y trascienden más allá del párrafo.

Mel'cuk alude a cómo las FL pueden ser útiles en la selección de expresiones capaces de establecer relaciones anafóricas tales que garanticen la cohesión del texto y la secuencia oracional sin incurrir en repeticiones tediosas. La base de todas estas relaciones se fundamenta en el conocimiento léxico. El ejemplo de cohesión que presenta es el siguiente:

(24) Sales *increased* slightly in Quebec and Ontario. Modest *gains* were also reported in British Columbia.

En este fragmento, en lugar de repetir elementos léxicos de la primera oración y decir *Sales also increased in British Columbia*, se utiliza el valor de la función léxica S_2 ⁴⁷ asociada al lexema *increase*. Formalmente, S_2 (increase) = gain_N [the amount by which X increased]. La aplicación de esta FL genera un texto más variado y elegante desde el punto de vista retórico.

Alonso Ramos *et al.* (1995) también estudian la referencia anafórica como fenómeno cohesivo y cómo las FL paradigmáticas pueden formalizar fenómenos léxico-semánticos implicados en la generación de textos. Al hablar de cohesión, la sitúan en el plano paradigmático y en un contexto textual amplio que abarca las oraciones anteriores y posteriores que contienen elementos léxicos en relación semántica con el lexema en cuestión.

En su análisis de la anáfora como recurso cohesivo léxico, estos autores establecen tres criterios que ha de cumplir toda expresión que se considere anáfora de otra:

1. Tener la misma referencia que su antecedente
2. Ser una palabra plena, con lo que se restringe el campo de estudio a la cohesión léxica
3. Compartir un vínculo léxico-semántico con el antecedente. De ahí que no se considere como anáfora vínculos de naturaleza extralingüística como el existente entre *the Prime Minister of France* y el nombre de la persona que ocupa ese cargo.

Para poder delimitar el vínculo léxico-semántico arriba mencionado, esquematizan en forma de marco la representación conceptual subyacente a los lexemas analizados. En este marco se sistematizan los roles semánticos de los argumentos que aparecen en cada oración. Para sistematizar aún más estas relaciones léxico-semánticas

⁴⁷ Typical noun for numbered participant.

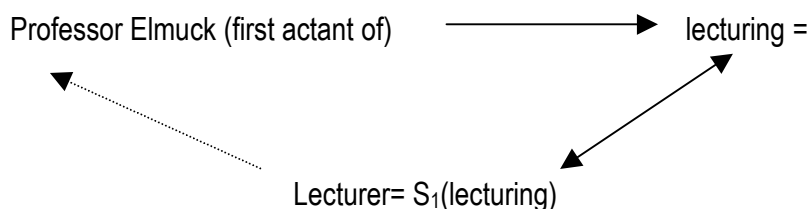
y darle un alcance más allá de la oración, estos autores aplican las funciones paradigmáticas de la TST al estudio de relaciones anafóricas. En concreto, utilizan las funciones léxicas **Gener**, **Syn**, **Conv**, **S₀**, **S_{instr}**, **S_{med}**, **S_{res}**, **S_{loc}** y **S_{mod}**, **Contr** y **Anti**, tanto en su forma simple (**Gener**) como en su forma compuesta (**Gener S₀**).

Por otra parte, según estos autores, la referencia entre el antecedente y la anáfora se puede establecer directamente (*direct lexical coreference*) o de un modo indirecto (*indirect lexical coreference*). En el último caso, la referencia se establece a través de otra expresión que está relacionada paradigmáticamente con el antecedente. Como ejemplo podemos mencionar la relación referencial indirecta entre *Professor Elmuck* y *lecturer* en el siguiente texto:

(25) *Professor Elmuck* was *lecturing* on Lexical Functions to third year students. The *lecturer* was interesting.

Esta relación indirecta entre *Professor Elmuck* y *lecturer* se establece mediante el lexema *lecturing*, cuyo primer actante o argumento es *lecturer* [**S₁** (lecture) = lecturer]. Formalmente, hemos presentado esta relación correferencial indirecta mediante líneas discontinuas, mientras que las relaciones semánticas las hemos marcado con líneas continuas.

DIAGRAMA 9: Referencia anafórica directa e indirecta



En cierto modo, esta forma de referencia es similar a la que Hoey (1991: 64-65) denomina paráfrasis compleja mediante triángulo de conexiones (*link triangle*).

En conclusión, si bien los criterios de definición de la referencia anafórica, especialmente el que la restringe a la que se establece entre palabras plenas, y el

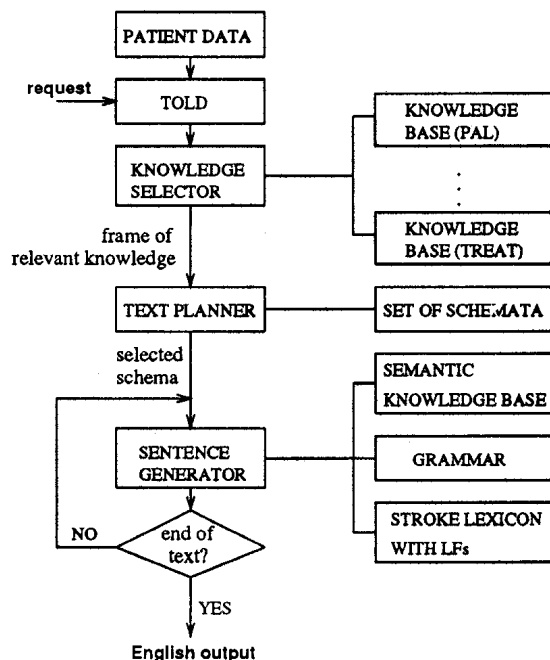
concepto de marco propuesto por Alonso Ramos *et al.* nos pueden ser de utilidad para nuestro estudio, creemos conveniente aumentar la taxonomía de FL implicadas en la cohesión. También consideramos necesario incluir las relaciones sintagmáticas, dado que la repetición de las mismas tiene un valor cohesivo y aplicar todos estos planteamientos a textos completos y a un corpus extenso de textos.

Otros estudiosos de la cohesión dentro de la TST como Lee y Evens (1996) aplican el concepto de cohesión léxica a la creación de un sistema experto capaz de generar párrafos explicativos sobre el diagnóstico y tratamiento de pacientes afectados de apoplejía. Su análisis de la cohesión es el más completo dentro de la TST y su alusión a los trabajos de McKeown (1985), Morris y Hirst (1991) y Myers (1991) demuestra un mayor conocimiento de dicho fenómeno.

Al igual que nosotros, Lee y Evens reconocen dos tipos de cohesión léxica, aunque los denominan con la nomenclatura de Halliday y Hasan (1976): reiteración y colocación, en lugar de acuñar un término en línea con la TST, algo así como relaciones cohesivas paradigmáticas y sintagmáticas.

Este generador de textos cuenta con: a) un selector de conocimiento que consulta una base de datos y elige el marco más apropiado; b) un planificador textual que selecciona entre un conjunto de superestructuras (*schemata*); c) un generador de oraciones conectado a tres repositorios, uno con información semántica, otro con información gramatical y el tercero, un lexicón en el que se especifican las FL en las que pueden participar los lexemas del mismo. Los componentes de este generador no difieren demasiado con respecto a los propuestos en el modelo cognitivo de la traducción de Bell (1991).

DIAGRAMA 10: La organización del generador (Lee y Evens 1996: 301).



Estos tres elementos se corresponden con los tres pasos que sigue el generador a la hora de producir textos. El primer paso consiste en la selección de uno de los marcos contenidos en la base de datos conceptual. Como ilustración del tipo de marcos que contiene esta base de datos, veamos el correspondiente a la aspirina (ibid: 301):

(26) Drug frame (

(generic	: Aspirin)
(brand_name	: Bufferin, Excedrin, Easprin)
(class	: Anti-Platelet)
(avg_oral_dose	: 325_mg)
(min_oral_dose	: 325_mg)
(indication	: cerebral_embolism_of_cardiac_origin)
(contraindication	: peptic_ulcer_disease)
(risk	: NIL)
(efficacy	: NIL)
(chemistry	: NIL))

En un segundo paso, se procede a la planificación de los textos. Un planificador textual selecciona el esquema textual o *superestructura* más apropiada para la situación

comunicativa que se presenta. En el caso de la receta de un medicamento, la superestructura que guía el mensaje del facultativo es básicamente la siguiente: el nombre del medicamento, la posología, la vía (oral por defecto), y las indicaciones y contraindicaciones.

Por último, el generador presta atención al nivel oracional y a la elección de lexemas para transmitir el significado pretendido y para mantener la cohesión del texto pero sin incurrir en una excesiva repetición. En esta etapa son importantes dos facetas: la generación de expresiones referenciales y la generación de colocaciones apropiadas.

En la primera faceta, las FL relacionan los lexemas y determinan cómo se articulan y cómo se pueden intercambiar estos según la información que se presente sea nueva o no, lo cual determina el artículo que debe acompañar a cada lexema. Así por ejemplo, el hecho de que *aspirin* esté en relación con *anticoagulant* mediante la función léxica **Gener**, hace que la expresión *this anticoagulant*, donde *anticoagulant* es un término genérico de *aspirin*, pueda hacer referencia a *aspirin*. A este tipo de referencia Lee y Evans lo denominan *sustitución superordinada*:

(27) *Aspirin* was prescribed for Mrs. Washington. The dosage was 325 mg., the average oral dose. *This anticoagulant* is indicated in cerebral embolism of cardiac origin.

En la búsqueda de colocaciones apropiadas para un término, el generador busca aquellas que hacen de *valor* de la función léxica que se pretenda expresar, una información que tiene almacenada. En uno de los ejemplos que presentan Lee y Evans, el generador tiene que seleccionar la expresión adecuada para expresar la idea de descenso de la temperatura corporal. Dado que, entre la información almacenada por el generador figura que el lexema *temperature* se relaciona mediante la función **CausPredMinus** con la expresión *[to]reduce*, es posible producir la oración:

(28)

Dilation of the blood vessels in the skin reduces the temperature of the body

CausPred (*temperature*) = *[to] reduce*

En definitiva, esta última aproximación pone de manifiesto las aplicaciones de la TST a la lingüística computacional. En concreto, las funciones léxicas pueden servir de fundamento a la especificación de sublenguajes, es decir, de "subsistemas de la lengua restringidos para hacer referencia a un determinado campo del saber" (Kittredge y Lehrberger 1982). Como afirma Frawley (1988: 344), los análisis textuales de sublenguajes científicos que han llegado a buen término tienen una base léxica.

4.9.3. Implicaciones

En vista de todas estas aproximaciones al fenómeno cohesivo, parece plausible sistematizar las relaciones semánticas entre los términos que crean cohesión mediante funciones léxicas abstractas. No obstante, creemos que la TST deja un poco de lado la faceta textual de la cohesión, el hecho de que la cohesión se crea entre palabras que conforman un texto y no entre palabras aisladas. A raíz de esto, sus representantes circunscriben el análisis a fragmentos de texto y no se despegan demasiado de las especulaciones sobre la cohesión de Halliday y Hasan (1976). Los estudios que más se aproximan a nuestra propuesta son los de Alonso Ramos et al. (1995), Lee y Evens (1996) y Escalier y Fournier (1997). En este último estudio, se propone una representación del significado que tenga en cuenta las diferentes realizaciones léxicas y sintácticas de este en una lengua determinada y cómo estas quedan relacionadas mediante la paráfrasis. Intuimos que estas expresiones con un significado muy parecido se van repitiendo a lo largo de todo un texto mediante determinadas funciones léxicas paradigmáticas. Se estudiará esto en el capítulo 10.

Creemos pues posible aplicar la rigurosidad de descripción semántica de la TST al estudio de la cohesión en textos no muy extensos (análisis intensivo de la cohesión). En este tipo de análisis, se combinarán las funciones léxicas con una adaptación del modelo de Hoey, que concede gran relevancia a criterios formales y textuales. En el capítulo 7, *Esquemas de representación para el análisis intensivo de la cohesión*, se aplicará la TST a la clasificación de los recursos cohesivos léxicos de los textos y se interpretarán estos recursos a la luz de las cadenas cohesivas del texto analizado en el capítulo 10.

4.10. Cohesión y tipo textual

Hasta aquí se han repasado distintos estudios que apuntan hacia la doble naturaleza léxica y textual de la cohesión. Es ahora oportuno prestar atención a las líneas de investigación que han explorado la relación que puede existir entre la cohesión, principalmente, la cohesión léxica y los tipos textuales.

4.10.1. Cohesión *referencial* y *género*

Uno de los primeros estudios que pretende encontrar una relación entre *recursos cohesivos* y *género* es el de Frawley y Smith (1989).

Desde el presupuesto de que un texto es un texto porque sus constituyentes presentan coherencia entre ellos, Frawley y Smith (1989) quieren determinar si la coherencia y la forma de establecer relaciones cohesivas basadas en la referencia cambia de unos géneros a otros.

Basándose en la premisa de que la referencia es la forma cohesiva más utilizada⁴⁸, estos autores se concentran en el estudio de la distribución y la función cohesiva de los pronombres, que establecen principalmente conexiones endofóricas y exofóricas en los textos. Su hipótesis es que la forma en que se distribuyen estos pronombres es indicativa del tipo de discurso en el que se establecen esas conexiones. Frawley y Smith utilizan un corpus de 16 000 palabras extraído del *Brown Corpus* y que contiene instancias de los siguientes discursos, que Frawley y Smith denominan erróneamente géneros: periodismo, ficción, religión y ciencia.

Aunque reconocemos el carácter pionero de su estudio y el valor de haber utilizado un corpus informatizado, hemos encontrado algunos fallos metodológicos. En primer lugar, los autores no aclaran qué recursos establecen la coherencia y qué pronombres concretos son los que se están teniendo presentes. Tampoco definen qué entienden por *género*, de forma que llegan a confundir este concepto con el de *discurso*, que implica dominios como la ciencia y la religión.

En un primer momento, estudian la distribución de pronombres nominativos de tercera persona del singular, que, según Frawley y Smith, son indicio de cohesión

⁴⁸ En contraposición a esta postura, Hoey (1991) sostiene que el enlace cohesivo más frecuente es la cohesión léxica.

endofórica o intratextual en un texto concreto. Los resultados correspondientes a los cuatro tipos de discurso mencionados indican claras variaciones. Por ejemplo, se constata una parquedad de pronombres en los textos científicos, en comparación con los otros tipos de discurso. Esta falta de referencias pronominales ha sido señalada por Ellegard (1978) y Gopnik (1972) y pretende evitar la ambigüedad. Otra explicación de este fenómeno se encuentra en que, en los textos científicos, la cohesión endofórica se establece mediante repetición léxica o la mención de palabras léxicamente relacionadas.

En cuanto a la utilización de pronombres de referencia exofórica, que, según Halliday y Hasan, son los de primera y segunda persona, también se perciben diferencias. Y así, los pronombres exofóricos también son mucho menos frecuentes en los textos científicos.

A grandes rasgos, sus conclusiones son las siguientes. El género de ficción es el que más utiliza la endófora y la exófora, al igual que los textos religiosos. Los textos periodísticos recurren con frecuencia a la endófora y con poca frecuencia a la exófora. Los textos científicos hacen un uso escaso de la endófora y la exófora. A partir de estos resultados, se puede inferir que distintos *tipos textuales* llevan asociados distintas configuraciones de la *cohesión referencial* mediante pronombres, que tiene una naturaleza textual.

4.10.2. Cohesión léxica y conocimiento científico en textos especializados y divulgativos

Los patrones cohesivos que se dan en los artículos científicos especializados difieren con respecto a los que se dan en los artículos científicos de divulgación. Myers (1991) encuentra una explicación a esto en el hecho de que los lectores de dichos textos no poseen los mismos conocimientos sobre el mundo científico. Este conocimiento es el que permite al lector unir una oración a la siguiente, relacionar referencias a la misma entidad y distinguir referencias a distintas entidades.

Los lectores de textos científicos necesitan conocer las relaciones léxicas que existen entre los lexemas del texto para ver la cohesión implícita del mismo. Por el contrario, los lectores de artículos de divulgación tienen que reconocer las relaciones cohesivas del texto, que están marcadas de una forma explícita, para inferir relaciones

léxicas que desconocen debido a que son legos en la materia de la que trata el artículo. Mediante estas relaciones cohesivas son capaces de unir los campos semánticos pertenecientes a un dominio especializado al dominio de la vida diaria.

Al leer textos científicos especializados, lectores no expertos como estudiantes, profesores de idiomas, traductores y periodistas tienen problemas a la hora de entender la terminología y a la hora de conectar una oración con la siguiente. Esto es así porque no es suficiente saber el sentido de los términos y la función de los *conectores*. También es necesario conocer las relaciones léxicas existentes dentro del texto, que no son explícitas en los artículos especializados.

En el caso de los textos de divulgación, el lector no necesita aportar este conocimiento léxico al texto porque la cohesión explícita en el mismo le ayuda a inferir los conocimientos sobre la materia que no posee. Estos recursos cohesivos, que son mucho más variados en los textos de divulgación, construyen una especie de puente entre el vocabulario cotidiano de los lectores y la terminología científica que aparece en el texto. Mientras que, en los textos especializados, el recurso más empleado es la cohesión léxica, en los textos divulgativos, además de este recurso, se emplean con frecuencia pronombres, conjunciones y elipsis.

Para poder entender la cohesión de un texto, es necesario tener una serie de conocimientos (Brady y Berwic 1983; Lehnert y Ringle 1982). Con la intención de especular sobre el conocimiento que el lector lego requiere para entender textos especializados, Myers se adentra en dos dominios cuya principal preocupación es la representación del conocimiento: el procesamiento del lenguaje natural y la sociología de la ciencia.

Para los primeros, el conocimiento se entiende de un modo cognitivo, como algo que puede ser representado en una serie de proposiciones (Brady y Berwick 1983; Lehnert y Ringle 1982). El conocimiento se encuentra en los enunciados y es posible representarlo en un gráfico como predicados lógicos y redes semánticas. Para la sociología de la ciencia (Latour 1987; Bazerman 1988), el conocimiento es algo que poseemos los seres humanos para poder encontrar sentido a nuestra interacción diaria con otros seres humanos y con objetos, con nuestro entorno. Desde la sociología de la

ciencia, el conocimiento está enfocado a la toma de decisiones y a la acción; es algo eminentemente práctico.

Myers defiende que ambos tipos de conocimiento son necesarios para poder detectar la cohesión de un texto. Por una parte, el lector ha de comprender la estructura semántica de un texto y, por otra parte, debe entender qué función juega el texto en la comunidad de discurso y en la construcción del conocimiento científico.

Por ejemplo, en un artículo de genética, el lector tiene que tener dos tipos de conocimiento. En primer lugar, tiene que saber qué moléculas componen el ADN. En segundo lugar, tiene que saber qué hacen los expertos con esas moléculas, los procesos, técnicas e instrumentos implicados en la investigación. También tiene que estar familiarizado con el género del artículo experimental y los recursos retóricos y movimientos (Swales 1981, 1990, 1994) que se emplean en este género. Este último conocimiento sobre el género tiene implicaciones en la cohesión porque permite que los lectores establezcan relaciones entre elementos aparentemente inconexos. Un lector consciente de que, en la introducción de un artículo experimental, se indica lo inapropiado de aproximaciones anteriores y lo que pretende aportar ese artículo al campo de la genética, tendrá más posibilidades de interpretar correctamente las relaciones entre palabras.

En su estudio, Myers compara los recursos cohesivos empleados en la introducción a algunos artículos científicos con los que aparecen en textos divulgativos. El corpus incluye distintos tipos textuales: artículos científicos, noticias científicas, artículos de divulgación y secciones de libros de texto.

4.10.2.1. Textos especializados: requisitos epistemológicos y cohesión

El conocimiento especializado que ha de tener el lector lego para identificar las entidades que aparecen en los textos e inferir conocimiento de los mismos comprende básicamente los siguientes aspectos:

(a) Reconocer casos de homonimia y polisemia

El lector lego necesita saber con qué sentido se utilizan los términos especializados y qué sentido adquieren en el contexto de la ciencia palabras de la lengua

general. Así, el lector lego necesita ampliar el marco de algunos verbos o nombres de la lengua general y aprender qué argumentos llevan en ese dominio.

Por ejemplo, en un texto especializado, el verbo *to infect* puede tener el siguiente marco según el primer argumento tenga un rol semántico de *fuerza* o de *agente* (Dik 1989).

(29) <i>one organism infects another organism</i> <i>a researcher can infect one organism with another organism</i>	<i>(organism)</i> <small>Fuerza</small> <i>(researcher)</i> <small>Agente</small>
--	--

Este marco no se encuentra en la lengua general común, en la que el primer argumento del verbo *to infect* tiene el rol semántico de *Fuerza*.

También se encuentran casos de unidades léxicas repetidas con la misma forma aunque se refieren a distintas entidades o a distintos aspectos de la misma entidad. Por ejemplo, en artículos sobre genética, el ADN aparece con distintos sentidos: *unidad*, *molécula con dos filamentos*, *sustancia de la que se toman cantidades* o *calificador de entidades*.

(b) *Asociar palabras que tienen el mismo o un significado parecido: sinonimia, hiponimia y antonimia*

El lector no especializado tiene problemas a la hora reconocer la misma entidad bajo distintas formas. No obstante, esto no es demasiado común en los artículos especializados, en los que se utiliza la repetición léxica en lugar de la sustitución por sinónimos, para evitar la falta de precisión. Cuando hay una sustitución de una palabra por lo que parece un sinónimo, los expertos infieren que tiene que haber una diferencia en su significado porque, de lo contrario, se hubiera recurrido a la misma palabra.

Además de sinónimos (*paráfrasis simple*), la cohesión léxica también depende de la utilización de antónimos (*paráfrasis compleja*) y términos pertenecientes a un conjunto ordenado de palabras plenas (*sistemas de incompatibilidad múltiple*). Para entender estas oposiciones, el lector necesita tener conocimiento experto puesto que las

pistas sintácticas y estructurales en artículos especializados no son tan explícitas como para permitir reconocer esos contrastes.

Las relaciones de hiperonimia e hiponimia son más frecuentes en los textos de divulgación que en los especializados. Si aparecen en los últimos, normalmente indican un cambio de énfasis y no se hacen explícitas mediante la cohesión.

(c) Determinar relaciones de dominancia en los sintagmas

El lector ha de conocer distintas colocaciones y saber cuál es el nodo en esa colocación para reconocer las relaciones de dominancia que se dan en los extensos sintagmas nominales típicos de los textos científicos. Este conocimiento es compartido por los expertos en la materia.

Además de la extensión de los sintagmas, otro problema es que algunos de los lexemas se pueden interpretar como sustantivos o como adjetivos. En contraposición con los textos especializados, los textos de divulgación evitan esta confusión convirtiendo estos sintagmas en oraciones de relativo y en sintagmas preposicionales. Para poder detectar estas relaciones de dominancia es necesario una labor terminográfica seria que establezca todas las relaciones posibles entre los términos y las relaciones que se activan cada vez que se activa un término.

(d) Definir segmentos discursivos en los que existen funciones comunicativas distintas

Este conocimiento va más allá del conocimiento de relaciones léxicas y supone reconocer unidades estructurales superiores a la oración. Supone la capacidad de reconocer la intención del autor en las distintas secciones de un artículo. Por ejemplo, un lector comprenderá mejor la introducción de un artículo si sabe que el contenido de esta se organiza en función de tres funciones comunicativas:

Move 1	Establishing a research territory
Move 2	Establishing a niche (Obligatory)
Move 3	Occupying the niche

(véase *Modelo de Swales* en 3.1.2.1.)

El desconocimiento de esta estructura probablemente impide que el lector lego entienda la función social del artículo y que encaje las oraciones dentro de un *esquema* que codifique los actos de habla que se suceden en el texto.

4.10.2.2. Textos de divulgación: recursos cohesivos y adquisición de conocimiento

La importancia de los textos no especializados para el lector lego radica en que le permiten adquirir conocimiento de experto mediante recursos cohesivos explícitos. Estos recursos ayudan al lector a reconocer las relaciones léxicas entre los términos y establecen un puente entre dos registros, el del especialista y el del aficionado.

En los textos de divulgación, el lector hace inferencias a partir de la cohesión que encuentra en el texto. Como Halliday y Hasan (1976: 299) defienden:

It is the continuity provided by cohesion that enables the reader or listener to supply all the missing pieces, all the components of the picture which are not present in the text but are necessary to its interpretation

Uno de los recursos cohesivos más frecuentes es la utilización de términos especializados definidos en función de términos de la lengua común o de términos semiespecializados. Por ejemplo, Myers (ibid: 17) menciona la siguiente definición:

(30) chromosomes ... are chemical threads that carry the instructions needed to run every *cell*.

Después de estas definiciones, se repite el término especializado para que el lector aumente su conocimiento. También aparecen casos en los que hay un término especializado y, unas líneas más abajo, aparece un término superordinado que ayuda al lector a comprender el significado del término especializado.

Las metáforas son también un recurso frecuente en los textos divulgativos, como se percibe en este ejemplo:

- (31) Instructions for the protein repertoire in both prokaryotes and eukaryotes are **written** in the **four letter alphabet** of DNA.

También los lectores de textos divulgativos pueden percibir relaciones entre oraciones viendo el uso explícito de conjunciones. En los artículos especializados estas relaciones no suelen ser explícitas.

4.10.2.3. Recapitulación

Myers aporta a los estudios sobre cohesión una interesante reflexión sobre la interrelación entre conocimiento y cohesión, y entre registro y cohesión. Los distintos tipos de conocimiento que requieren los lectores de artículos científicos y los lectores de artículos de divulgación determinan que se utilicen distintos recursos cohesivos.

Mientras que los lectores de artículos especializados tienen que conocer las relaciones léxicas entre términos para detectar la cohesión implícita propia de estos textos, los lectores de textos de divulgación tienen que reconocer las relaciones cohesivas, que son explícitas en estos tipos textuales, para inferir las relaciones léxicas entre términos.

Esta diferencia puede explicar los distintos recursos cohesivos utilizados en los dos tipos textuales. En los artículos especializados, la cohesión se crea principalmente mediante repetición léxica. En los textos de divulgación, esta también crea cohesión, pero se utilizan otros recursos como la referencia, la sinonimia y la conjunción.

Teniendo en cuenta que en los artículos científicos no se utilizan apenas pronombres o sinónimos como recursos cohesivos, estos textos resultan de difícil lectura para los lectores no especialistas. Por el contrario, los textos de divulgación contienen más recursos que hacen explícita la cohesión y, en consecuencia, es más fácil que el lector detecte las relaciones entre las oraciones y haga inferencias sobre el significado de términos desconocidos.

De todas formas, según Myers, la lectura de textos no especializados no es suficiente. Es necesaria la lectura de distintos géneros que muestren distintas situaciones sociales como artículos de divulgación a distintos niveles, artículos de revisión, informes sobre el desarrollo de la investigación y libros de texto, entre otros. Es decir, la

lectura de artículos de divulgación no puede servir de sustituto a la lectura de textos especializados porque, como dice Myers (ibid: 23),

Learning the language of molecular genetics, like learning any language, involves learning collocations and idioms. The reader (or translator) needs to know which words and phrases count as technical terms (like *gene expressions* or *translation*), which are the most common collocations in the field (such as coding *for* a protein of DNA *strand*), and which are drawn from general usage (such as *directs* or *instructions*). Scientific language is by its nature formulaic: a relatively limited set of verbs goes with a relatively limited set of nouns. Popularizations explore the relations between this register and other registers. But they do not necessarily show institutionalization of expressions.

4.10.3. Cohesión y dificultad comunicativa

Tanskanen (1995) establece que la cohesión léxica es sensible a un aspecto extralingüístico esencial en la configuración de un determinado género: las dificultades comunicativas que impone el contexto situacional. En su estudio, se identifican distintas categorías de cohesión léxica y se estudian su frecuencia y densidad con el fin de descubrir los recursos cohesivos que se adaptan mejor a situaciones en las que la comunicación es más difícil o más fácil. Para ello, investiga cómo se establece la cohesión léxica en tres géneros que responden a una situación comunicativa diferente: la conversación cara a cara, el discurso preparado y la prosa académica.

Su clasificación de los artificios que crean cohesión léxica aúna las taxonomías de Halliday y Hasan (1976) y Hasan (1984). Así, reconoce dos grandes grupos, la reiteración y la colocación y, dentro del primero, identifica unas subcategorías que son bastantes parecidas a las de Hasan (1984).

Las subcategorías de reiteración que Hasan denominaba repetición, sinonimia, antonimia, hiponimia y meronimia se denominan ahora repetición, equivalencia, contraste, especificación (*public transport–underground*) y coespecificación (*buses – underground*) y generalización, respectivamente. La diferencia lleva consigo cambiar una terminología propia de la semántica léxica por una más enfocada a las relaciones que son específicas al discurso.

En cuanto a la colocación, esta comprende principalmente dos categorías:

- a) Pertenencia a un conjunto ordenado de lexemas como los días de la semana o los números: *Monday* – Thursday.
- b) Implicación, en la que la asociación entre elementos no es sistemática sino que se basa en la tendencia que tienen a aparecer en el mismo contexto: *Winter* – *cold*; *armies* – *war*.

Con respecto a la colocación, Tanskanen sostiene que es una categoría difícil de identificar en los análisis de la cohesión léxica porque no implica una relación tan sistemática como la que se da en la reiteración (Halliday y Hasan 1976: 288; Hoey 1991: 83-84). También sostiene que la colocación es una relación más local que la reiteración. Si estas afirmaciones de Tanskanen fueran ciertas, no hubiera sido posible crear programas informáticos que computan colocaciones como *Microconcord* o *Wordsmith Tools* ni tampoco hubiera sido posible el estudio de Phillips (1985) sobre cómo la colocación crea cohesión entre capítulos de libros.

En su análisis no hemos encontrado detalles sobre el corpus estudiado, lo cual nos hace poner en entredicho los resultados de este estudio y destacar la necesidad de replicar investigaciones de este tipo.

En primer lugar, Tanskanen analiza la frecuencia de casos de reiteración y colocación en los tres géneros y detecta que la reiteración es mucho más frecuente que la colocación. Un mayor uso de la reiteración en comparación con la colocación se da en aquellos géneros en los que la comunicación es más difícil como en los discursos preparados y en la conversación cara a cara entre más de dos participantes. En situaciones en las que las demandas comunicativas son más moderadas, como en la prosa académica, las diferencias entre el recurso más frecuente (la reiteración) y el que le sigue en frecuencia (la colocación) se reducen.

El orden de frecuencia de las subcategorías para los tres géneros es bastante similar para los tres géneros. Y así, la repetición es el recurso cohesivo más frecuente, seguido de la implicación, la equivalencia, la generalización, la especificación y la co-especificación con prácticamente la misma frecuencia. El contraste y el conjunto

ordenado aparecen como los recursos menos frecuentes. De aquí se deduce que el orden de frecuencia de las subcategorías no indica una diferencia entre los géneros.

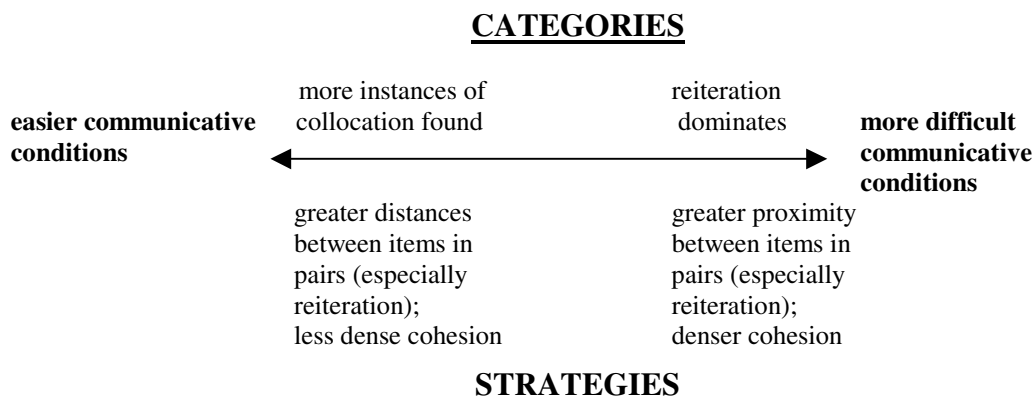
En segundo lugar, Tanskanen observa la variación de la densidad de recursos cohesivos y especula sobre su valor como indicio de situaciones comunicativas difíciles. La densidad de recursos cohesivos depende de las frecuencias normalizadas de cohesión y colocación⁴⁹ y de la distancia, en oraciones, entre los elementos del par.

El discurso preparado es el más denso en cuanto a reiteración, seguido por la prosa académica. La conversación es el género con menor densidad de reiteración. En cuanto a la distancia, los pares entre los que hay una relación reiterativa están separados por más oraciones que los que están relacionados mediante colocación. La distancia entre pares reiterativos es mayor en las conversaciones, sobre todo, cuando hay unos cuantos interlocutores. Es algo menor en la prosa académica y bastante menor en el discurso preparado, en el que los elementos del par cohesivo se encuentran en la misma o en una oración adyacente. También se encuentran relacionados en la misma oración o en una oración adyacente los pares relacionados por colocación, lo cual hace que la densidad de colocación normalmente se mantenga alta.

En esencia, el discurso preparado es el más denso en cuanto a la cohesión léxica, ya que hay más pares cohesivos que en los otros géneros y las distancias entre los elementos de los pares son menores. Tanskanen deduce que, si la densidad de recursos cohesivos es alta, sobre todo cuando son de naturaleza *reiterativa*, nos hallamos ante un género que cobra existencia en condiciones comunicativas difíciles.

⁴⁹ La frecuencia normalizada corresponde al número de *pares* en relación cohesiva por cada mil palabras.

DIAGRAMA 11: Cohesión y dificultad comunicativa (Tanskanen 1995: 537)



Tanskanen presenta sus conclusiones en un diagrama, en el que se indica que la reiteración se utiliza de un modo más frecuente que la colocación, y domina principalmente en situaciones en las que la comunicación es más difícil. La colocación se utiliza en un mayor grado en aquellas situaciones en las que la comunicación es más fácil. Asimismo, las estrategias de utilización de la cohesión léxica también cambian según las condiciones comunicativas, al menos en lo que se refiere a la densidad cohesiva: cuanto más difícil resulta la comunicación, mayor es la proximidad entre los elementos entre los que se establece la cohesión, y por tanto, mayor la densidad cohesiva.

Los recursos empleados en situaciones más difíciles están orientados al receptor, mientras que los que se utilizan en condiciones en las que la comunicación es más fácil están orientados al productor del mensaje.

Esta aproximación parece demostrar que la cohesión léxica se ajusta a las demandas comunicativas planteadas en determinadas situaciones y, en cierto modo, apoya la hipótesis de que puede existir una correlación entre cohesión y tipos textuales.

4.11. Aproximaciones a la cohesión desde la traducción

La redacción de textos cohesivos facilita la transmisión de la función comunicativa de los mismos y su comprensión por parte del lector. En consecuencia, este fenómeno lingüístico es crucial para el traductor, que detecta la conexión entre los elementos del texto origen y reformula esta unidad conceptual en el texto meta mediante recursos cohesivos.

Los estudios de traducción, apenas han explorado la construcción de la cohesión en los textos, con la excepción de los trabajos de Blum-Kulka (1986), Lotfipour-Saedi (1997) y Fernández Polo (1999). En todo caso, la investigación de la cohesión no ha puesto el énfasis en la cohesión léxica y no se ha llevado a cabo ni en textos extensos completos ni mediante métodos informáticos.

Aunque no se centren en la cohesión léxica, estos ponen de manifiesto que la traducción normalmente lleva consigo cambios en la cohesión. El estudio más citado es el de Blum-Kulka (1986: 19), donde se propone la llamada hipótesis de la explicitación, que sostiene que los textos traducidos cuentan con una mayor redundancia que lleva aparejada un mayor grado de explicitación cohesiva.

Blum-Kulka (ibid: 23) define la coherencia como una relación semántica implícita entre las partes del texto que se hace explícita gracias a la interpretación que hace el lector. La cohesión es una relación semántica explícita entre partes de un texto, expresada por marcadores lingüísticos.

Asimismo, añade que la traducción puede llevar consigo cambios en la cohesión en dos sentidos (Blum-Kulka 1986: 18). Por una parte, los textos traducidos son más explícitos o menos explícitos que el original. Por otra, el potencial semántico explícito o implícito del texto origen a veces varía en el texto meta.

Algunos de estos cambios se deben a la gramática, a motivos estilísticos, es decir, al hecho de que diferentes lenguas imponen diferentes normas que rigen la utilización de recursos cohesivos y, por último, al proceso traductor en sí, en el que afloran las normas retóricas correspondientes a la lengua origen y término. Lotfipour-Saedi (1997) comenta que la cohesión léxica contribuye a la textura del texto. Lo aplica a un fragmento literario de sólo 276 palabras en el que Lotfipour-Saedi sólo analiza *una* cadena léxica. Este autor parece desconocer los estudios sobre cohesión de Hoey,

Morris y Hirst, como indica el hecho de que conciba la *colocación* en el sentido que le dan Halliday y Hasan. Este sentido dista del sentido más aceptado en lingüística, que es el que le da Firth, es decir, la colocación es la relación que un lexema establece con otros lexemas que aparecen con una probabilidad significativa en su cotexto inmediato.

De todas formas, propone unas etapas que permiten evaluar si una traducción mantiene el valor discursivo pretendido por el TO:

- a) determinar las principales cadenas léxicas centrales al tema del TO
- b) describir las relaciones semánticas entre los nodos de las cadenas
- c) determinar la distancia en palabras entre los nodos
- d) determinar hasta qué punto los miembros de la cadena están subsumidos en la jerarquía textual
- e) evaluar la naturaleza de la equivalencia traductológica comparando el texto origen y meta según estos rasgos textuales

Por último, presentamos un trabajo que está plenamente relacionado con la hipótesis formulada en esta tesis. Fernández Polo (1995, 1999) detecta cómo al traducir textos se producen cambios en la cohesión y coherencia de los mismos a raíz de la adecuación a las normas retóricas asociadas a una determinada lengua y a un determinado tipo textual. El interés de esta propuesta merece una sección aparte.

4.11.1. Cambios cohesivos en la traducción de textos de divulgación científica

Fernández Polo (1995) parte del presupuesto de Myers (1991) de que en textos especializados se utilizan diferentes elementos cohesivos en comparación con sus versiones de divulgación y sostiene que estas mismas diferencias se encuentran al confrontar textos divulgativos en inglés con sus traducciones al español. Los textos profesionales muestran una preferencia especial por la repetición, mientras que en la divulgación esta es menos habitual.

En los textos divulgativos, la conexión entre los conceptos es mucho más explícita, dado el limitado conocimiento de los lectores, y esto se ve reflejado en una serie de elementos cohesivos: la referencia, la cohesión léxica, la conjunción y el tema,

principalmente. Estos elementos son necesarios para mantener la efectividad comunicativa, y por tanto, hay que prestarle bastante atención a la hora de traducir.

En este trabajo se indagan las causas de la variación con respecto a la cohesión que se produce como resultado de la traducción de textos divulgativos del inglés al español. Para ello, se analiza un reducidísimo corpus formado por dos artículos de la revista *Scientific American* y la versión traducida al castellano que publica *Investigación y Ciencia*. Fernández Polo (1999) amplía el número de artículos de estas publicaciones hasta conseguir un corpus de aproximadamente 65 000 palabras formado por cinco originales en inglés y sus correspondientes traducciones, y cinco artículos originales en castellano.

Las conclusiones de ambos trabajos apuntan hacia distinciones cohesivas que obedecen, entre otros motivos, a la estructuración retórica propia de las lenguas inglesa y española. Asimismo, explican el papel de la repetición en los textos originales ingleses con el fin de detectar posibles pérdidas de coherencia en las versiones traducidas. En esencia, la repetición léxica presenta un determinado referente mediante diversas expresiones que pueden ser sinónimas o estar relacionadas mediante hiponimia. Con estas se establece una red conceptual en torno a la cual se articula el tema. Es decir, se presenta un determinado referente desde distintos ángulos con la esperanza de que al menos una de las formulaciones permita al lector acceder a la idea que se le desea transmitir.

La referencia a entidades previamente mencionadas en los artículos del *Scientific American* se actualiza mediante la repetición léxica, incluso en casos en los que la distancia entre la segunda mención y el antecedente es mínima, como se ve en el ejemplo que aporta Fernández Polo (1995: 258).

- (32) Since then, the sophisticated satellite has been making the first comprehensive astronomical detections of **gamma rays, the most energetic form of electromagnetic radiation**. When viewed **at gamma ray energies**, the universe appears unfamiliarly dynamic and capricious.

No ocurre lo mismo en la versión española, en la que se evita la repetición léxica y en su lugar aparecen formas de referencia menos explícitas cognitivamente: elipsis, pronombres, sinónimos y términos superordinados.

(33) Desde entonces, este satélite de 16 toneladas de peso ha venido realizando rastreos exhaustivos de **los rayos gamma celestes, la radiación del extremo más energético del espectro**. A **esas frecuencias** aparece un universo voluble y dinámico.

En este sentido, la versión española se aproxima más a las convenciones de los textos divulgativos ingleses, que utilizan pronombres y sinónimos en lugar de repetición léxica. En la traducción al español (33), los sinónimos para "rayos gamma celestes", "el espectro" y "esas frecuencias", no dejan muy claro cuál es su referente, ya que, según Polo, este no está explícito en el cotexto anterior.

Para explicar este fenómeno, Fernández Polo remite a la norma retórica que considera la repetición léxica como un signo claro de un estilo pobre (Montaño-Harmon 1991: 420-421; Aznar Más *et al.* 1989: 30). El cumplimiento de esta convención hace que el lector español tenga más problemas a la hora de asignar el referente a un pronombre, establecer la conexión entre dos sinónimos o entre un hiperónimo y su hipónimo, y ver si dos formas similares hacen referencia a la misma entidad. Estas operaciones mentales requieren un mayor conocimiento del tema y un mayor esfuerzo por parte de la audiencia española.

Fernández Polo también presta atención a dos recursos cohesivos que Hoey (1991) no incluye en su análisis: la progresión temática y la conjunción. Mientras que en el texto original la posición temática incluye el rema de la oración precedente, en las traducciones aparecen segmentos de información conocida en posición de rema, con lo cual, se destruye la cohesión. Según este autor, estas diferencias responden a que cada género lleva asociados unos patrones temáticos y estos están determinados en gran medida por la audiencia. Hay géneros cuyos patrones temáticos contribuyen más a la comprensión por parte del lector que otros. Por este motivo, los traductores han de ser conscientes del papel de la progresión temática en el establecimiento de la cohesión.

En cuanto al uso de conjunciones y conectores, estos son más explícitos en los artículos del *Scientific American* que en *Investigación y Ciencia*, en cuyos artículos se suprimen algunos de los conectores presentes en el texto original. Esto parece contradecir la hipótesis de la explicitación de Blum-Kulka (1986: 19), es decir, el hecho de que la traducción supone un mayor grado de explicitación cohesiva. Una respuesta a este fenómeno se puede encontrar en la retórica de contrastes (Connor y Kaplan 1987; Hinds 1987; Mauranen 1993; Tyler 1992), que sostiene que hay diferencias en el grado de explicitación entre lenguas y culturas. El inglés es muy explícito o, en palabras de Hinds (1987) y Hatim (1997) es *writer-responsible*. El español es mucho menos explícito y se sitúa entre el inglés y las denominadas *reader-responsible languages* como el japonés y el coreano. Algunas investigaciones como las de Saint-John, Montañó-Harmon y Reppen y Grabbe (*apud* Fernández Polo 1999: 179) sugieren que los hispanohablantes consideran los textos que siguen la retórica anglosajona como excesivamente simples, aburridos e incluso infantiles; y a la inversa, para un anglosajón, la retórica del español le resulta confusa y difícil de comprender. En consecuencia, la supresión de conectores no es más que una adaptación del texto original a la estructura retórica de la lengua española.

A pesar de lo importante que resulta en traducción respetar las convenciones retóricas de las lenguas de partida y de llegada, una aplicación irreflexiva de las mismas puede conllevar modificaciones en la cohesión textual. Estas modificaciones requieren un conocimiento experto mayor por parte del lector español, que puede pensar que distintas formas léxicas corresponden a distintas realidades cuando, en realidad, corresponden a la misma entidad.

En consecuencia, Fernández Polo concluye que es necesario realizar estudios contrastivos sobre cómo se amolda la cohesión a distintos géneros en las lenguas inglesa y española. También es necesario encontrar el equilibrio entre las convenciones retóricas de una lengua y de un género determinado y los recursos cohesivos asociados a esa lengua y a ese género.