

7. Esquemas de representación para el análisis intensivo de la cohesión

A la hora de plasmar gráficamente la cohesión, hemos necesitado más de una forma de representación dado que la extensión de los textos sobre cáncer y la exhaustividad de nuestro análisis dejan bastante limitada la capacidad de las *matrices de repetición*, presentadas en 4.5.

El primer tipo de representación utilizado en esta tesis es una adaptación de dichas matrices que permite un análisis detallado. Se aplicará esta representación al análisis de textos breves, ya que la segmentación en oraciones permite detectar empírica y sistemáticamente todos los recursos cohesivos desplegados en el texto. En efecto, los textos breves nos brindan la oportunidad de concebir la cohesión como un fenómeno *interoracional* que se amolda a algunas exigencias *intraoracionales*, como, por ejemplo, la de minimizar la repetición del mismo lexema⁹⁷ en la misma oración o en el cotexto cercano⁹⁸. Esta doble faceta interoracional e intraoracional de la cohesión permite tomar la oración como punto de referencia para analizar exhaustivamente textos breves. Este tipo de análisis *interoracional* susceptible de ser representado mediante matrices de repetición es el que hemos llamado *análisis intensivo*.

El otro modelo de representación del que nos vamos a servir se basa en el principio de *lexical chain* (Morris y Hirst 1991) y la aplicación del mismo al reconocimiento automático de la cohesión por parte de Ellman (véase 4.8.2.). Estimamos que este puede ser idóneo para el análisis de textos más extensos y de un corpus textual, como parece indicar el hecho de que este modelo haya sido aplicado al procesamiento automático de textos.

7.1. Modificaciones necesarias a las matrices de repetición de Hoey (1991)

La representación gráfica para el análisis intensivo de la cohesión es un desarrollo de la propuesta por Hoey y permite un reflejo detallado de las relaciones semánticas entre términos que crean cohesión. Para esto han sido necesarias la modificación de las categorías propuestas por Hoey, la inclusión de la repetición sintagmática y la descripción de las relaciones semánticas entre términos.

⁹⁷ Un lexema es el significado que se actualiza mediante diversas formas léxicas asociadas a la misma raíz.

A pesar de haber modificado las categorías de repetición de Hoey, hemos intentado mantener la idea de que la repetición es un elemento fundamental en la construcción de la cohesión, de ahí que hayamos utilizado también la palabra *repetición*. En la siguiente tabla, exponemos la posible correspondencia entre nuestra propuesta y la de Hoey. A pesar de definir las, Hoey (1991) excluye del análisis la repetición superordinada e hiponímica y la paráfrasis compleja porque considera que no contribuyen demasiado a la cohesión del texto.

TABLA 16: Modificación de las categorías de cohesión propuestas por Hoey (1991)

HOEY (1991)	NUESTRA PROPUESTA
Repetición léxica simple	Repetición exacta: <i>cancer – cancer</i> Repetición simple: <i>treat – treated</i>
Repetición léxica compleja	Repetición simple: <i>release (verb) – release (noun)</i> Repetición variada paradigmática sintáctica
Paráfrasis simple	Repetición paradigmática semántica (sinonímica)
Paráfrasis compleja por antonimia ⁹⁹	Repetición paradigmática opositiva (antonímica)
Paráfrasis compleja ¹⁰⁰	Repetición paradigmática asociativa
Correferencia	-----
Repetición superordinada e hiponímica	Repetición paradigmática semántica (hiponímica)
-----	Repetición paradigmática semántica (meronímica)
-----	Repetición paradigmática opositiva (conversión y complementariedad)
-----	Repetición paradigmática asociativa
-----	Repetición sintagmática (exacta, simple y variada)
-----	Repetición mixta de unidades fraseológicas
-----	Repetición mixta de unidad fraseológica y lexema

En esta tabla, las casillas superiores contienen los recursos que, según Hoey, aportan al texto un mayor grado de cohesión. Estos son más explícitos y más reconocibles de forma automática, por ejemplo, mediante listas de frecuencia lematizadas que indiquen los lexemas con mayor peso específico en los textos. Sin embargo, la presencia consistente de los demás recursos en redes semánticas puede resultar tan cohesiva como la repetición exacta de lexemas.

⁹⁸ Con esto queda patente que el nivel intraoracional sí tiene importancia en la cohesión.

⁹⁹ Esta categoría desaparece en Hoey (1995).

Por otra parte, creemos que cualquier análisis sistemático de la cohesión debe incluir la repetición sintagmática y la que hemos denominado repetición mixta. Pensamos que las matrices de repetición deben reflejar la cohesión entre oraciones vinculadas no sólo por lexemas sino también por unidades fraseológicas. De hecho, nuestro análisis pone de manifiesto que en el lenguaje científico, la cohesión se crea principalmente por la reiteración de unidades terminológicas construidas a partir de más de una palabra, es decir, por repetición sintagmática y mixta.

De ahí que, a diferencia de Hoey, propongamos incluir en las casillas de las matrices no sólo lexemas, sino también unidades fraseológicas, cuyos componentes están separados por guiones bajos (Cáncer_de_pulmón).

En cuanto a la representación de la cohesión mediante funciones léxicas, se identificarán las FL que más contribuyen a la cohesión. En algunos casos se verá cómo al determinar la función léxica que vincula dos unidades léxicas¹⁰¹, hay que distinguir entre si se está tomando un punto de vista lexicográfico-terminológico o textual. Desde la primera perspectiva, se considera el papel de la FL que relaciona dos lexemas en el sistema de la lengua, que normalmente va a ser el mismo que se actualiza en los textos concretos (perspectiva textual).

Por la complejidad que conlleva la identificación de relaciones sintagmáticas y de FL, ha sido conveniente elaborar para cada texto dos tablas con el formato de una matriz de repetición. La primera tabla representa las relaciones cohesivas más obvias en el texto y se denominará *matriz de repetición de las relaciones cohesivas primarias*. En la segunda, llamada *matriz de repetición de las relaciones cohesivas secundarias*, se identifican una serie de relaciones que, aunque más difusas, crean cohesión. Desdoblaremos la primera tabla en dos, una tabla A y una tabla B, con el objetivo de visualizar en papel el análisis detallado que se llevará a cabo en el capítulo 10. La tabla de relaciones más difusas sirve meramente para que el lector se haga una idea de cómo recursos cohesivos que no son tan efectivos ni obvios contribuyen también significativamente a crear cohesión, si se activan frecuentemente.

¹⁰⁰ El carácter más difuso de esta categoría hace que Hoey (1995) prescinda de ella.

7.1.1. Matriz de repetición de relaciones cohesivas primarias (Tabla A)

En este apartado se perfila el formato de las casillas de una tabla de tipo A, extraída de la matriz de repetición del texto analizado en el capítulo 10. En concreto, presentamos las casillas (2, 1), (3, 1) y (3, 2), donde la casilla (2, 1) objetiva la cohesión de la segunda oración con respecto a la primera¹⁰². Dentro de cada casilla, hemos otorgado un número a cada instancia de cohesión para facilitar la referencia. Al lado de cada número aparece, o bien una unidad léxica aislada, en el caso de que ambas oraciones contengan exactamente el mismo lexema (riesgo), o bien un par de unidades léxicas separadas por una barra oblicua (*Cáncer_pulmonar/ carcinoma_pulmonar*). En este último caso, la cursiva señala aquellos elementos que, por no ser iguales, crean repetición léxica variada.

TABLA 17: Fragmento de TABLA tipo A

	(1)		
2	1. Alteraciones_del_cromosoma_9 / Alteraciones_en_el_cromosoma_9 2. Cáncer_de_pulmón/ cáncer_ <i>pulmonar</i> 3. Cáncer_de_pulmón/ <i>Tumor</i>	(2)	
3	4. Alteraciones_del_cr9/ Anomalías_del_cr9 5. Cáncer_de_pulmón/ <i>Carcinoma_pulmonar</i>	12. Alteraciones_en_el_cr9/ <i>anomalías_del_cr9</i> 13. Riesgo 14. Investigadores/ investigaciones 15. Cáncer_pulmonar / <i>carcinoma_pulmonar</i> 16. Marcadores / <i>predictor</i> 17. Tumor / <i>carcinoma_pulmonar</i>	(3)

Así pues, la casilla (2, 1) contiene tres instancias de cohesión léxica, es decir, indica que la segunda oración tiene tres vínculos cohesivos con la primera oración. El primer y el segundo caso son ejemplos de repetición variada sintáctica mixta en la que se produce también la repetición exacta, por una parte, de los lexemas *alteraciones* y *cromosoma*, y por otra parte, del lexema *cáncer*:

¹⁰¹ Nos estamos refiriendo a lo que en inglés se conoce como *word-type*, es decir, a “las variantes morfológicas de una misma entrada del léxico” (Márquez 1998: 85).

(61) *alteraciones del cromosoma 9 / alteraciones en el cromosoma 9*: REPETICIÓN SINTÁCTICA MIXTA (SINONÍMICA)

(62) *cáncer de pulmón / cáncer pulmonar*: REPETICIÓN SINTÁCTICA MIXTA (SINONÍMICA)

La repetición mixta aúna una relación cohesiva paradigmática (sinonimia) con una sintagmática entre una alteración y su localización o entre el cáncer y su localización. La relación sintagmática se consigue mediante las funciones léxicas **Degrad_s**, **Loc_{POSS}**, **Loc_{IN}**, y **Loc_A**.

Los términos relacionados en un par están separados por una barra oblicua, de forma que el primero de ellos pertenece a la oración que aparece antes en el texto. Así, por ejemplo, en el par *alteraciones del cromosoma 9 / alteraciones en el cromosoma 9* de la casilla (2, 1), el primer término corresponde a la primera oración y el segundo, a la segunda.

Nótese también cómo la cursiva indica repetición variada, en contraposición a la grafía normal, con la que se marcará la repetición exacta. Asimismo, con la representación tipográfica de estos lexemas en negro indicaremos que comparten una relación de sinonimia, entendiendo esta en el sentido laxo que defendimos en el apartado 6.6.1.2. El color azul, que aparece en el tercer caso de cohesión de la tabla, indica una relación de hiponimia entre los términos.

En definitiva, la tabla A presenta todas las instancias de cohesión del texto y, mediante la tipografía, aporta información básica, que se va a completar en la tabla B.

7.1.2. Matriz de repetición de relaciones cohesivas primarias (Tabla B)

La tabla B objetiva el tipo de repetición y las funciones léxicas que comparten las unidades léxicas y terminológicas identificadas en la tabla A. En esta, la casilla (2, 1) sería algo así:

¹⁰² Recordamos que Hoey consideraba la cohesión como una relación con oraciones precedentes. Por eso, la referencia a las casillas mediante coordenadas sigue este formato: (oración posterior, oración precedente). Esta presentación nos permitirá aludir a las casillas de matrices de repetición que no encajan en formato A4.

TABLA 18: Fragmento de TABLA tipo B

(1)		(2)
2	1. Mixta de UF (v sint) / Syn 2. Mixta de UF (v sint) / Syn 3. Mixta de UF y L (v léx) / Gener <i>Syn (cáncer de pulmón) = este tipo de tumor</i>	

Esta casilla indica que los lexemas del primer par, se relacionan mediante repetición mixta y que son dos unidades fraseológicas. La abreviatura *v sint* nos dice que la repetición es variada y se fundamenta en diferencias sintácticas. Si miramos la información que la tabla A aporta sobre el ejemplo 1, “Alteraciones_*del*_cromosoma_9 / Alteraciones *_en _ el _* cromosoma_9”, la cursiva indica que la variación de los términos lo es con respecto a la preposición utilizada.

Después de la mención al tipo de repetición, la tabla de tipo B especifica la función léxica que vincula esas dos UF desde el punto de vista lexicográfico y textual, que en el ejemplo 1 es la función **Syn**. No obstante, a veces vamos a asociar a un mismo ejemplo dos funciones léxicas. Cuando esto ocurre, la primera función es la que comparten los lexemas desde el punto de vista lexicográfico o en una ontología, es decir, sin tener en cuenta el contexto. En cuanto a la segunda función, es la que se actualiza en el texto en función del cotexto que acompaña al término en cuestión. Este cotexto hace que el valor de la FL desde el punto de vista lexicográfico no coincida con el valor que adquiere en el texto. Una buena ilustración de esto lo constituye el tercer caso de cohesión que se presenta en la casilla (2,1)

(63)

(2, 1)

3. *Cáncer de pulmón / tumor*

Mixta de UF y L (v léx) / Gener

Syn \supset (*cáncer de pulmón*) = este tipo de tumor

En este ejemplo, el término genérico asociado a *cáncer de pulmón* (primera oración) es *tumor* (segunda oración), por lo que podemos decir que desde el punto de vista lexicográfico estos dos términos se relacionan mediante **Gener**. Sin embargo, el cotexto precedente de la palabra *tumor* (*este tipo de*) hace que en este texto *cáncer de pulmón* (palabra clave) y *este tipo de tumor* (valor) queden asociados mediante sinonimia (**Syn**). Es decir, en el texto, la función **Gener** se comporta como la función **Syn** e indica que estos dos términos tienen la misma referencia. Cabe destacar que la expresión que hace de palabra clave de la FL es la que aparece en la oración precedente.

Intuimos que tanto la lengua inglesa como la española se sirven de un repertorio de expresiones que, al acompañar a los términos, contribuyen a tales cambios de comportamiento de la FL¹⁰³. En consecuencia, siempre que ocurra esto haremos explícito el cotexto del valor de la función en la tabla B.

Una vez que se han identificado los tipos de repetición que se dan en un texto, las FL que se ponen en juego y las cadenas cohesivas que se establecen, es posible completar aún más el análisis representando los posibles *marcos conceptuales* que sustentan la producción y la comprensión de textos cohesivos. El hecho de que los marcos correspondientes a cada texto se perfilarán a partir de los datos lingüísticos extraídos de las matrices de repetición o las cadenas léxicas les confiere una gran validez y utilidad para la traducción. Esto es así porque presupone que el marco conceptual para el texto original debe ser muy similar al correspondiente al texto traducido.

Como conclusión, nuestro intento de estudiar la cohesión de un modo exhaustivo que contemple su complejidad semántica y textual, nos ha llevado a proponer nuevas categorías de repetición. Entre estas, destacamos la repetición sintagmática y mixta, que dan cuenta de la importancia de la fraseología en los textos científicos. Además de incluir estas nuevas categorías, hemos modificado las matrices de repetición de Hoey en el siguiente sentido. En primer lugar, nuestras matrices se desdoblaron en dos, una tabla A y una tabla B, para que así puedan especificar las funciones léxicas del texto. En segundo

lugar, hemos hecho que la tipografía de los lexemas de las casillas resulte informativa. Las matrices de repetición en un análisis intensivo son útiles para identificar las oraciones con más vínculos, que, según Hoey (1991), son las que contienen el contenido fundamental de un texto.

7.2. Esquemas de representación complementarios

La información que aportan las matrices de representación se ha extraído de las listas de frecuencia elaboradas por el programa informático de análisis léxico *Wordsmith Tools*. Este programa además de proporcionar una tabla con datos estadísticos para cada texto como, por ejemplo, la media de palabras por oración, nos va a ayudar a detectar qué lexemas son más frecuentes, qué unidades fraseológicas constituyen y qué distribución tienen determinadas palabras a lo largo del texto.

Las listas de frecuencia para cada texto analizado de un modo intensivo van a ser las siguientes: una lista con lexemas individuales y listas con las combinaciones de dos, tres y cuatro palabras del texto. A este último tipo de listas se les denomina en lingüística de corpus *n-grams*.

Presentamos tanto los datos estadísticos como las tablas de frecuencia de un texto sobre posibles mutaciones genéticas asociadas al cáncer de pulmón. Se trata del *abstract* de un artículo experimental publicado en la revista *Cancer* de John Wiley & Sons (véase capítulo 10). Los *abstracts* son el género de mayor circulación entre la comunidad de discurso médica y constituyen una unidad coherente. En consecuencia, creemos que es posible analizar los *abstracts* independientemente del resto del artículo experimental. En el capítulo 10, se analizará la adaptación al español en forma de noticia publicada por la revista *Jano Medicina y Humanidades*.

(64)

1 Lung Carcinoma Patients with a Family History of Cancer and Lymphocyte Primary Chromosome 9 Aberrations

2 #ABSTRACT# BACKGROUND. Deletion of chromosome 9p has been reported in numerous tumor types.

¹⁰³ Se presentará el repertorio correspondiente a la lengua inglesa en 13.1. Expresiones puente entre el lenguaje del especialista y el del lector lego.

3 The authors demonstrated in an earlier study that spontaneous chromosome aberrations on chromosome 9 in peripheral blood lymphocytes (PBLs) were a significant risk predictor for lung carcinoma.

4 METHODS. The current study evaluated the relationship between self-reported family history of cancer and spontaneous chromosome aberrations in the PBLs of 97 previously untreated lung carcinoma patients.

5 The authors' hypothesis was that individuals exhibiting specific chromosome aberrations might have inherited genetic instability and thus be more likely to report a family history of cancer.

6 For each individual, a personal interview was conducted to construct a detailed family history, and 100 metaphases from PBLs were analyzed for spontaneous aberrations by G-banding.

7 RESULTS. The patients reported having 829 first-degree relatives, including 74 (8.9%) with cancer.

8 A significantly elevated odds ratio (OR) of 2.7 was noted in smokers for having chromosome 9 aberrations and a first-degree relative with cancer.

9 When the family history of cancer was dichotomized into lung carcinoma or other cancers, the OR associated with chromosomal aberrations was 8.5 for lung carcinoma but only 2.3 for other cancers.

10 In addition to chromosome 9 aberrations, other spontaneous chromosome aberrations and family history of cancer were also evaluated, but no associations were found.

11 There were no associations between age, gender, ethnicity, or smoking status and the chromosome 9 aberration profile.

12 CONCLUSIONS. The findings of this study suggest that chromosome 9 aberrations may be a marker of cancer susceptibility and may be associated with familial aggregation of cancer.

(Cfr. Apéndice II, texto 7)

7.2.1. Datos estadísticos sobre el texto/ corpus

A la hora de estudiar la cohesión, puede ser de utilidad conocer la proporción de formas distintas (*type*) por cada 100 ocurrencias léxicas (*token*), es decir, la *type-token ratio*. Si se comparan corpórea amplios de igual extensión, una *type-token* elevada puede indicar un mayor número de formas distintas y una mayor complejidad conceptual. A efectos de cohesión, puede ser el indicio de una mayor repetición variada, lo que supone un esfuerzo cognitivo por parte del lector y un despliegue mayor de recursos cohesivos. De todas formas, este índice no es demasiado informativo para corpórea pequeños, en los que adquiere un valor muy elevado.

Otros datos de la tabla son el número de oraciones (*sentences*) y el número de palabras por oración (*sent. length*), que siempre es mayor para la lengua española en relación con la lengua inglesa. Un número elevado de letras en las palabra puede indicar la

presencia de términos médicos de origen grecolatino, ya que estos suelen ser más extensos que los vocablos anglosajones.

TABLA 19: Datos estadísticos sobre el *abstract* en inglés

Text File	CHROMOS.TXT	3-letter words	37
Bytes	1.919	4-letter words	34
Tokens	285	5-letter words	14
Types	122	6-letter words	22
Type/Token Ratio	42.81	7-letter words	21
Standardised Type/Token		8-letter words	18
Ave. Word Length	5.53	9-letter words	17
Sentences	15	10-letter words	22
Sent.length	19.00	11-letter words	21
sd. Sent. Length	10.62	12-letter words	7
Paragraphs	1	13-letter words	2
Para. length	285.00	14-letter words	1
1-letter words	31		
2-letter words	38		

7.2.2. Listas de frecuencia

A continuación presentamos listas de frecuencia con los lexemas individuales y las combinaciones de dos, tres y cuatro palabras que aparecen más de dos veces. En la primera columna (*N*), se nos indica el rango de cada lexema con respecto a su frecuencia.

Bajo *word* aparece la forma léxica más frecuente del lexema en cuestión. El resto de formas léxicas asociadas a esa misma raíz están en la casilla *lemmas*. De los dos sentidos del término *lema*¹⁰⁴, se optará por el más amplio, es decir, lema es el conjunto de formas derivadas de la misma raíz independientemente de su categoría gramatical. Y así, las formas *findings* y *found* estarían agrupadas bajo el mismo lema, que etiquetaremos con la forma más frecuente del texto/corpus en mayúscula: FINDINGS.

Las tablas también indican la frecuencia absoluta (*Fr.*) y relativa (%) de un determinado lema. Nótese cómo después de la vigésimo cuarta unidad léxica, hemos sumado las frecuencias relativas para calcular la representación de los lexemas que

¹⁰⁴ El sentido más restringido de *lema* es el conjunto de formas flexivas asociadas a un mismo lexema. Desde esta perspectiva, *water** sería el lema que agrupa las formas {*water*, *waters*} del lexema *water*. {*Water*, *watered*, *watering*, *waters*} están asociadas al lexema verbal *water** (Cfr. *Corpora List*, 3/11/1999).

participan en relaciones de repetición exacta, simple y variada sintáctica. Obviamente, son aquellas palabras plenas lematizadas cuya frecuencia relativa es superior a 1. La suma, 34,7%, es mucho menor que el porcentaje de repetición léxica simple y compleja (variada sintáctica en nuestra terminología) obtenido por Benbrahim (1996), que era del 92-97%, o Ellman (1998), entre un 60 y un 80%. La única explicación para esta divergencia es que estos autores han computado como repetición léxica, la repetición entre palabras forma (*function words*). Además, ellos hablan de número de conexiones y nosotros, de número de lexemas que participan en esas conexiones, un número a todas luces diferente. Creemos que nuestro porcentaje es bastante más objetivo porque no caben errores en el cómputo de conexiones y porque no hemos fijado la oración como medida de la cohesión.

TABLA 20: Frecuencia de lexemas aislados

N	Word	Freq.	%	Lemmas
1	Chromosome	12	4,21	chromosomal(1)
2	Aberrations	11	3,86	aberration(1)
3	Cancer	11	3,86	cancers(2)
4	Family	7	2,46	familial(1)
5	History	6	2,11	
6	Carcinoma	5	1,75	
7	Lung	5	1,75	
8	Associated	4	1,40	associations(2)
9	Spontaneous	4	1,40	
10	Patients	3	1,05	
11	PBLs	3	1,05	
12	Reported	3	1,05	report(1)
13	Study	3	1,05	
14	Authors	2	0,70	
15	Evaluated	2	0,70	
16	Findings	2	0,70	found(1)
17	First-degree	2	0,70	
18	Having	2	0,70	
19	Individual	2	0,70	individuals(1)
20	Lymphocyte	2	0,70	lymphocytes(1)
21	May	2	0,70	
22	Relative	2	0,70	relatives(1)
23	Significant	2	0,70	significantly(1)
24	Smokers	2	0,70	smoking(1)
Repet. exacta, simple y variada sintáctica		99	34,7	
25	Abstract	1	0,35	
26	Addition	1	0,35	
27	Age	1	0,35	
28	Aggregation	1	0,35	
29	Analyzed	1	0,35	
30	Background	1	0,35	
31	Blood	1	0,35	

32	Conclusions	1	0,35	
33	Conducted	1	0,35	
34	Construct	1	0,35	
35	Current	1	0,35	
36	Deletion	1	0,35	
37	Demonstrated	1	0,35	
38	Detailed	1	0,35	
39	Dichotomized	1	0,35	
40	Earlier	1	0,35	
41	Elevated	1	0,35	
42	Ethnicity	1	0,35	
43	Exhibiting	1	0,35	
44	G-banding	1	0,35	
45	Gender	1	0,35	
46	Genetic	1	0,35	
47	Hypothesis	1	0,35	
48	Including	1	0,35	
49	Inherited	1	0,35	
50	Instability	1	0,35	
51	Interview	1	0,35	
52	Likely	1	0,35	
53	Marker	1	0,35	
54	Metaphases	1	0,35	
55	Methods	1	0,35	
56	Might	1	0,35	
57	Noted	1	0,35	
58	Numerous	1	0,35	
59	Odds	1	0,35	
60	Peripheral	1	0,35	
61	Personal	1	0,35	
62	Predictor	1	0,35	
63	Previously	1	0,35	
64	Primary	1	0,35	
65	Profile	1	0,35	
66	Ratio	1	0,35	

67	Relationship	1	0,35	
68	Results	1	0,35	
69	Risk	1	0,35	
70	Self-reported	1	0,35	
71	Specific	1	0,35	
72	Status	1	0,35	

73	Suggest	1	0,35	
74	Susceptibility	1	0,35	
75	Tumor	1	0,35	
76	Types	1	0,35	
77	Untreated	1	0,35	

TABLA 21: Combinaciones de dos palabras

N	Word	Freq.	%
1	Of cancer	7	2,45
2	Chromosome #	6	2,10
3	Family history	6	2,10
4	History of	5	1,75
5	Lung carcinoma	5	1,75
6	Chromosome aberrations	4	1,40
7	Spontaneous chromosome	3	1,05
8	A family	2	0,70
9	Aberrations and	2	0,70
10	Associated with	2	0,70

11	Cancer and	2	0,70
12	Carcinoma patients	2	0,70
13	For lung	2	0,70
14	May be	2	0,70
15	No associations	2	0,70
16	Of #	2	0,70
17	Other cancers	2	0,70
18	PBLs were	2	0,70
19	The authors	2	0,70
20	With cancer	2	0,70

TABLA 22: Combinaciones de tres palabras

N	Word	Freq.	%
1	Family history of	5	1,75
2	History of cancer	5	1,75
3	Chromosome # aberrations	4	1,40
4	Spontaneous chromosome aberrations	3	1,05
5	A family history	2	0,70
6	For lung carcinoma	2	0,70
7	Lung carcinoma patients	2	0,70
8	Of cancer and	2	0,70

TABLA 23:
Combinaciones de cuatro palabras

N	Word	Freq.	%
1	Family history of cancer	5	1,75
2	A family history of	2	0,70
3	History of cancer and	2	0,70

7.2.3. Matriz de repetición

Aunque el texto sólo cuenta con 12 oraciones¹⁰⁵, resulta bastante difícil representar la matriz de repetición en soporte papel, como se constata al representar sólo la primera columna de la matriz de repetición (Tabla 24). En esta hay un boceto de las conexiones entre la primera oración y las demás.

Según esta matriz de repetición, el título (primera oración) es muy informativo, ya que está vinculado con todas las oraciones. La relación se establece con respecto a las grandes áreas conceptuales del eje horizontal (MÉTODO CIENTÍFICO, ENFERMEDAD,

¹⁰⁵ En la tabla estadística aparecían 15 oraciones pero hemos considerado los encabezados (*Abstract*, *Background*) como parte de la oración que preceden.

PACIENTE, FACTORES DE RIESGO, CÉLULA, INSTRUMENTO, CARACTERÍSTICAS, DIAGNÓSTICO, TRATAMIENTO SÍNTOMAS, GENÉTICA). Todas estas áreas configuran un marco que podríamos denominar BIOLOGÍA MOLECULAR DEL CÁNCER DE PULMÓN (CITOGENÉTICA).

[TABLA 24](#): Primera columna de la matriz de repetición del texto 6 (desarrollo de Hoey 1991)

En el capítulo 10, se verá cómo es posible exportar estos datos a diagramas de sectores. Con estos se consigue una forma de representación empírica de *marcos conceptuales*.

De todas formas, por la extensión del texto, es preciso representar la matriz de repetición en forma de pares de coordenadas. En la siguiente tabla, se presenta cómo la primera oración está vinculada con el resto de las oraciones mediante 35 vínculos cohesivos, y la segunda, mediante 19. En el apéndice VI se incluyen las matrices para el resto de oraciones, en las que se llega a identificar un total de 153 vínculos. Cada instancia de cohesión está representada por un par de unidades léxicas pertenecientes a dos oraciones distintas (*par cohesivo*) o por un solo par en el caso de la repetición exacta.

TABLA 25: Cohesión con respecto a la primera oración (tabla A)

PAR COHESIVO	UNIDAD LÉXICA EN ORAC. PRECEDENTE / EN ORACIÓN POSTERIOR		
(2, 1)	1. Lung carcinoma / tumor types 2. Chromosome 9 aberrations / Deletion of chromosome 9p	(7, 1)	17. Patients 18. Family / Relatives 19. Cancer
(3, 1)	3. Chromosome 9 aberrations / Spontaneous chromosome aberrations 4. Lymphocyte / Peripheral blood lymphocytes (PLBs) 5. Lung carcinoma	(8, 1)	20. Cancer/ smokers 21. Chromosome 9 aberrations 22. Family / relative 23. Cancer
(4, 1)	6. Family history of cancer 7. Chromosome 9 aberrations / Spontaneous chromosome aberrations 8. Lymphocyte / PLBs 9. Lung carcinoma patients	(9, 1)	24. Lung carcinoma (2) 25. Family history of cancer 26. Cancer (2) 27. Chromosome 9 aberrations / chromosomal aberrations
(5, 1)	10. Patients / individuals 11. Chromosome 9 aberrations / chromosome aberrations 12. Family history of cancer	(10, 1)	28. Chromosome 9 aberrations 29. Chromosome 9 aberrations / spontaneous chromosome aberrations 30. Family history of cancer
(6, 1)	13. Patients / individual 14. Family history of cancer / family history 15. Lymphocyte / PLBs 16. Chromosome 9 aberrations / spontaneous aberrations	(11, 1)	31. Cancer / smoking 32. Chromosome 9 aberrations / chromosome 9 aberration
		(12, 1)	33. Chromosome 9 aberrations 34. Cancer / smoking 35. Family history of cancer / familial aggregation of cancer

TABLA 26: Cohesión con respecto a la segunda oración (tabla A)

PAR COHESIVO	UNIDAD LÉXICA EN ORAC. PRECEDENTE / EN ORACIÓN POSTERIOR		
(1, 2)	Véase (2, 1)	(7, 2)	45. Tumor types / cancer
(3, 2)	36. Deletion of chromosome 9p / spontaneous chromosome aberrations 37. Tumor types /lung carcinoma	(8, 2)	46. Deletion of chromosome 9p / chromosome 9 aberrations 47. Tumor types / cancer
(4, 2)	38. Deletion of chromosome 9p / spontaneous chromosome aberrations 39. reported / self-reported 40. Tumor types/lung carcinoma	(9, 2)	48. Tumor types / cancer 49. Deletion of chromosome 9p / chromosomal aberrations
(5, 2)	41. Deletion of chromosome 9p / chromosome aberrations 42. report / reported 43. Tumor types/ lung carcinoma	(10, 2)	50. Deletion of chromosome 9p / chromosome 9 aberrations 51. tumor types / cancer
(6, 2)	44. Deletion of chromosome 9p / spontaneous aberrations	(11, 2)	52. Deletion of chromosome 9p / chromosome 9 aberration (profile)
		(12, 2)	53. Deletion of chromosome 9p / chromosome 9 aberrations 54. tumor types / cancer

Por último, para poder computar todos estos enlaces, proponemos dos tablas en las que aparecen las conexiones en las que participa cada oración y el número de oraciones con las que cada oración está vinculada.

TABLA 27: Cómputo de las conexiones en las que participa cada oración

Oración precedente con la que está vinculada (Nº de conexiones)											Nº Orac	Oración posterior con la que está vinculada (Nº de conexiones)
1	2	3	4	5	6	7	8	9	10	11	1	
2											2	
3	2										3	
4	3	3									4	
3	3	3	4								5	
4	1	1	3	3							6	
3	1	1	2	3	2						7	
4	2	2	2	2	2	1					8	
4	2	2	2	2	2	1	3				9	
3	2	2	3	2	2	1	2	3			10	
2	1	2	2	2	1	1	2	3	3		11	

3	2	2	3	2	2	1	2	4	3	3	12
---	---	---	---	---	---	---	---	---	---	---	----

TABLA 28: Tabla de las conexiones en las que participa cada oración

Oración	Cantidad de oraciones con las que se vincula (precedentes, posteriores)
1	(0, 11)
2	(1, 10)
3	(2, 9)
4	(3, 8)
5	(4, 7)
6	(5, 6)
7	(6, 5)
8	(7, 4)
9	(8, 3)
10	(9, 2)
11	(10, 1)
12	(11, 0)

En vista de la tabla, se infiere que el texto está muy cohesionado ya que cada una de las oraciones está vinculada con once oraciones. Por su carácter sintético, su concisión y su brevedad, los *abstracts* son géneros con un alto grado de cohesión. Las oraciones que incluye son aquellas que contienen la información principal y las palabras clave del texto. Como ya se dijo, estas son las que tienen mayores vínculos cohesivos.