

Hacia la búsqueda de
rasgos discursivos específicos
de dominios especializados

Iria da Cunha
Gerardo Sierra

1. Introducción

El análisis del discurso especializado es un tema que suscita interés desde hace años. Se han estudiado los rasgos específicos de los diferentes géneros (por ejemplo, Swales, 1981, 1990; van Dijk, 1980, 1989; Kaplan *et al.*, 1994, Ciapuscio, 1998), o las diferencias entre textos especializados y textos de ámbito general (por ejemplo, L'Homme, 1993; Teufel y Moens, 2002; Cabré, 2007; Cabré *et al.*, 2010), teniendo en cuenta que podemos considerar que existen tres niveles de especialización (Cabré, 1998): alto (emisor experto y receptor experto), medio (emisor experto y receptor aprendiz de la materia) y bajo (emisor experto y receptor lego en la materia). Para diferenciar textos especializados de diferentes dominios, normalmente se emplean rasgos léxicos, principalmente, la terminología empleada (Cabré y Estopá, 2003; Ciapuscio, 2003); para caracterizar un tipo de género, en ocasiones se emplean rasgos gramaticales (véase por ejemplo la caracterización del género “resumen de artículo científico” de Graetz, 1985); para detectar los rasgos específicos de un tipo de género asociado a un dominio especializado, puede emplearse asimismo la estructura textual (véase, por ejemplo, el trabajo sobre la caracterización de los resúmenes médicos de Salager-Meyer, 1991).

El objetivo de este trabajo es realizar el análisis discursivo de un corpus de textos de un mismo nivel de especialidad (con emisor y receptor expertos del ámbito) y de un mismo género (resumen de artículo científico), pero de diferente dominio especializado (medicina y terminología). Ambos tipos de textos, aunque especializados, se refieren a dominios muy diferentes (científico-técnico vs. lingüístico), por lo que presuponemos que pueden existir ciertos rasgos discursivos característicos de cada uno. En este artículo pretendemos mostrar la existencia de estos rasgos.

Para ello, empleamos la *Rhetorical Structure Theory* (RST) de Mann y Thompson (1988). La RST es una teoría organizativa del texto que describe su estructura a partir de las relaciones que se establecen entre sus diferentes elementos discursivos. Estas relaciones pueden ser “núcleo-satélite” (si uno de los elementos aporta cierta información sobre el otro, como por ejemplo Elaboración, Concesión, Propósito, Condición, etc.) o “Multinucleares” (si no existe un elemento principal, como por ejemplo Contraste, Unión, Lista, etc.).

En el segundo apartado exponemos la metodología de nuestro trabajo. En el tercer apartado mostramos el análisis y los resultados obtenidos. Finalmente, en el cuarto apartado extraemos algunas conclusiones y señalamos el trabajo futuro que llevaremos a cabo.

2. Metodología

La metodología del trabajo es la siguiente: en primer lugar, conformamos un corpus de resúmenes de artículos científicos, que incluye un subcorpus de 20 resúmenes de artículos médicos¹ y un subcorpus de 13 resúmenes de artículos sobre terminología², cada uno de 4000 palabras. La diferencia en el número de textos se debe a que los resúmenes de artículos médicos suelen ser más breves que los resúmenes de artículos terminológicos.

En segundo lugar, realizamos el análisis discursivo de los textos del corpus mediante la RST. La Figura 1 muestra un árbol discursivo de un fragmento de uno de los textos del corpus médico, y la Figura 2, de uno de los textos del corpus terminológico.

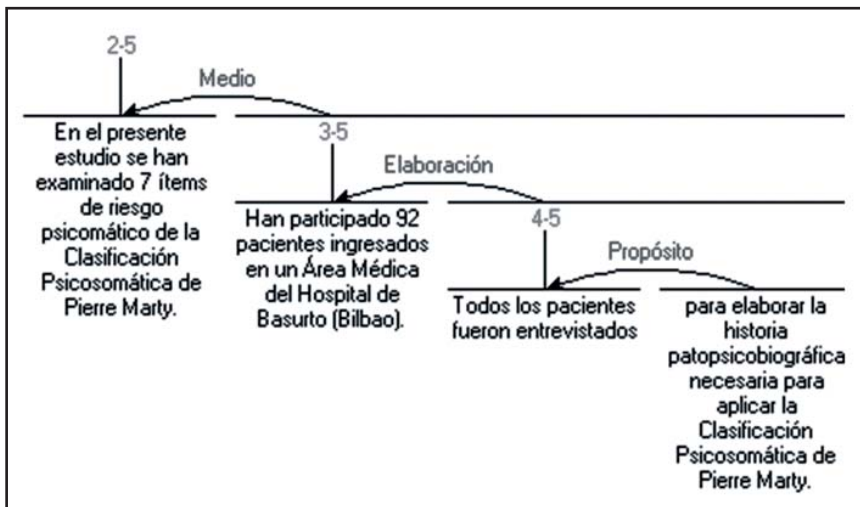


Figura 1. Árbol discursivo de la RST de un fragmento del corpus médico.

1- Los textos del corpus médico se extrajeron de la Gaceta Médica de Bilbao. Disponible en: <http://www.gacetamedicabilbao.org/web/es/>.

2- Los textos del corpus sobre terminología se extrajeron de las actas del Congreso Internacional de Terminología celebrado en Donostia y Gasteiz en 1897. Disponible en: <http://www.uzei.com/antcatalogo.asp?nombre=1687&hoja=0&sesion=14>.

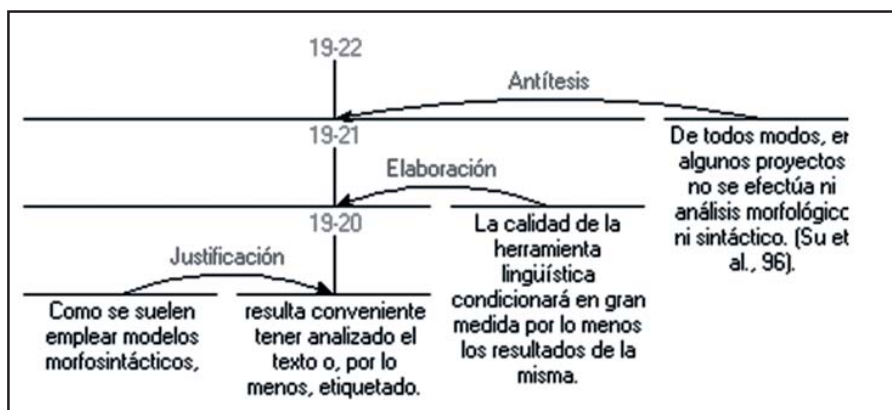


Figura 2. Árbol discursivo de la RST de un fragmento del corpus terminológico.

En tercer lugar, contabilizamos el tipo de relaciones detectadas en cada dominio y los marcadores discursivos que las evidencian. Tomamos el término “marcador discursivo” en un sentido amplio, es decir, considerando cualquier marca léxica que pueda ser indicadora de una relación discursiva, sin restringirnos a clasificaciones existentes, como serían las de Portolés (2001) o Montolío (2001). Ejemplos de marcadores serían: “aunque”, “por lo que”, “de todas maneras”, “no obstante”, “pero”, etc.

En cuarto lugar, establecemos algunas conclusiones sobre las diferencias discursivas encontradas en ambos subcorpus.

3. Análisis y resultados

Una vez realizado el análisis, observamos que existen diferencias relevantes entre los textos de ambos dominios. La principal diferencia se refiere al número de relaciones discursivas detectadas: 224 relaciones en el corpus médico y 174 en el corpus terminológico. Probablemente esta diferencia, bastante significativa si tenemos en cuenta que el número de palabras de cada subcorpus es el mismo, se deba a que en el ámbito médico se expresen más ideas en menor cantidad de texto y a que en el ámbito de la terminología las ideas se expresen con un mayor número de unidades léxicas.

Otra de las diferencias más relevantes entre ambos subcorpus tiene que ver con las relaciones discursivas empleadas. En primer lugar, en el subcorpus médico se emplean más relaciones de Resultado que en el corpus terminológico (17 y 7, respectivamente), lo cual quiere decir que

en el corpus médico se emplea un 7,6% de relaciones de Resultado, mientras que en el corpus terminológico se emplea un 4,02%. Este hecho puede deberse a que, en los textos que reflejan investigaciones médicas, se llega a resultados más tangibles que en las investigaciones sobre terminología. Un ejemplo de relación de Resultado en el subcorpus médico es el siguiente:

[Hasta la aparición de los fármacos tuberculostáticos, la toracoplastia y el neumotórax eran dos técnicas utilizadas para el tratamiento de la tuberculosis pulmonar.]
NÚCLEO_RESULTADO [El neumólogo afronta hoy las complicaciones secundarias, consistentes en paquipleuritis, deformidad de la caja torácica y escoliosis.]
SATÉLITE_RESULTADO

En segundo lugar, observamos un mayor número de relaciones de Interpretación en el subcorpus médico (13) que en el subcorpus terminológico (2); esto supone el empleo de un 5,8% de relaciones de Interpretación en el corpus médico y un 1,15% en el corpus terminológico. Esta diferencia podría deberse a que los médicos son más propensos a realizar interpretaciones sobre los resultados de sus investigaciones, mientras que los terminólogos son más cautos al hacerlo. A continuación mostramos un ejemplo de la relación de Interpretación en el subcorpus médico:

[En los 7 ítems se han encontrado diferencias estadísticamente significativas entre el grupo de pacientes oncológicos con los pacientes afectos de otro tipo de patologías ($p < 0.05$).]
NÚCLEO_INTERPRETACIÓN [Estos ítems diferencian a los pacientes con neoplasias de otro tipo de pacientes y permiten una valoración global de los mismos, ofreciendo una idea de las expectativas del proceso.]
SATÉLITE_INTEPRETACIÓN

En tercer lugar, detectamos un mayor número de relaciones de Medio en el subcorpus médico (19) que en el subcorpus terminológico (6), lo cual supone el uso de un 8,48% de relaciones de este tipo en el corpus médico y un 3,44% en el corpus terminológico. Suponemos que esto se debe a que en los resúmenes médicos, se reflejan las herramientas y métodos empleados en los experimentos, en mayor medida que en los resúmenes sobre terminología, como puede apreciarse en el siguiente ejemplo, extraído de uno de los textos del subcorpus médico:

[Desde enero de 1996 hasta junio de 1996 hemos revisado a un total de 30 pacientes con colapsoterapia, con 70.817 años (60-63 años) de edad media.]
NÚCLEO_MEDIO [Los procedimientos quirúrgicos consistían en: toracoplastia en 13, neumotórax en 15 casos, toracoplastia bilateral en 1 y la combinación de toracoplastia y neumotórax en otro caso.]
SATÉLITE_MEDIO

Hemos observado otras diferencias entre ambos corpus, pero estas son las más significativas.

Detectamos, asimismo, que una de las relaciones más empleadas en ambos subcorpus, como era de esperar ya que se trata de la relación más general, es la de Elaboración (58 ocurrencias en el corpus terminológico y 56 en el médico, lo que supone un 33,3% y un 25% de las relaciones empleadas en cada subcorpus, respectivamente). A continuación mostramos un ejemplo extraído del subcorpus médico y otro ejemplo extraído del subcorpus terminológico, respectivamente:

[Estudios recientes utilizando la técnica de “microarrays” han identificado un subgrupo de cánceres de mama con pésimo pronóstico.]_{NÚCLEO_}
ELABORACIÓN [Se trata de tumores que no expresan receptores hormonales ni el oncogén c-erb-B2, y sí a cambio citoqueratinas propias de las células del estrato basal epitelial.]_{SATÉLITE_ELABORACIÓN}

[El proyecto VOCALL (Vocational Language Learning for Less Widely Used and Taught Languages) es un proyecto fundado por la Comisión Europea en el marco del Programa Leonardo da Vinci.]_{NÚCLEO_ELABORACIÓN}
[El proyecto está coordinado por la Dublin City University (Dublín, Irlanda).]_{SATÉLITE_ELABORACIÓN}

En la Tabla 1 se muestran los datos cuantitativos del análisis discursivo realizado en ambos corpus: nombre de la relación, tipo (N-N: relación Multinuclear; N-S: relación Núcleo-Satélite) y número de relaciones detectadas en ambos subcorpus.

Relación	Tipo	Corpus terminológico	Corpus médico
Elaboración	N-S	58	56
Lista	N-N	29	59
Preparación	N-S	17	21
Fondo	N-S	13	10
Resultado	N-S	7	17
Justificación	N-S	6	0
Concesión	N-S	6	1
Medio	N-S	6	19
Propósito	N-S	5	2
Secuencia	N-N	4	2
Antítesis	N-S	4	2
Causa	N-S	4	4
Condición	N-S	3	2
Motivación	N-S	3	1
Interpretación	N-S	2	13
Circunstancia	N-S	2	2
Evaluación	N-S	2	0

Reformulación	N-S	1	0
Resumen	N-S	1	0
Solución	N-S	1	0
Disyunción	N-N	0	0
Capacitación	N-S	0	0
Conjunción	N-S	0	0
Evidencia	N-S	0	1
Alternativa	N-S	0	0
Unión	N-N	0	0
Contraste	N-N	0	12
Condición inversa	N-S	0	0
<i>Unless</i>	N-S	0	0
Total relaciones		174	224
Total N-N		33	73
Total N-S		141	151

Tabla 1. Análisis cuantitativo de las relaciones discursivas detectadas en ambos corpus.

En cuanto al número de marcadores empleados, también detectamos diferencias significativas: en el subcorpus médico se emplean 48 marcadores, mientras que, en el subcorpus terminológico, se utilizan 59. Esta diferencia es relevante, ya que, teniendo en cuenta que en el subcorpus médico se emplea un 22,33% más de relaciones discursivas que en el subcorpus terminológico; esto significa que el número de relaciones evidenciadas por marcadores discursivos en estos subcorpus es 21,42% y 33,9%, respectivamente. Creemos que esta diferencia se debe a que los lingüistas (es decir, los autores de los textos sobre terminología), por su formación, tienen más capacidad que los médicos para aportar cohesión a sus escritos, lo cual se evidencia en un mayor uso de marcadores discursivos. Ejemplos de marcadores que evidencian relaciones discursivas serían, en la Figura 1, “para”, que refleja una relación de Propósito, y, en la Figura 2, “de todos modos”, que marca una relación de Antítesis.

En la Tabla 2 se muestran los datos cuantitativos del análisis de los marcadores de ambos corpus: nombre de la relación, tipo de relación, marcadores detectados (con el número de ocurrencias de cada uno entre paréntesis) y el total de marcadores en ambos corpus (terminológico y médico).

Relación	Tipo	Marcadores del corpus terminológico	Total	Marcadores del corpus médico	Total
Elaboración	N-S	de acuerdo a (1) que (1) y (2) asimismo (1) lo que (1) efectivamente (1) como ejemplo de lo anterior (1) además (1)	9	como (2) en consecuencia (1) de ellos (1)	4
Lista	N-N	por un lado (1) y, por otro, (1) y (8) también (1) y por último (1) por otra parte (1)	13	y (16) así mismo (1) así como (3) aparte de (1) igualmente (2)	23
Preparación	N-S	—	0	—	0
Fondo	N-S	—	0	—	0
Resultado	N-S	por lo tanto (2)	2	—	0
Justificación	N-S	como quiera que (1) en la seguridad de que (1) como (1)	3	—	0
Concesión	N-S	aunque (3) si bien (2) de cualquier forma (1)	6	aunque (1)	1
Medio	N-S	—	0	para ello (1) mediante (1)	2
Propósito	N-S	con este fin (1) el objetivo es (1) a fin de (1) con tal fin (1) el objetivo de nuestro proyecto es (1)	5	para (1) con el fin de (1)	2
Secuencia	N-N	seguidamente (2) a continuación (1)	3	—	0
Antítesis	N-S	de todas formas (1) pero (2) de todos modos (1)	4	pero (1)	1
Causa	N-S	ya que (2) habida cuenta de que (1) por consiguiente (1)	4	debido a (1) por lo que (2)	3
Condición	N-S	si (2) siempre que (1)	3	—	0

Motivación	N-S	de manera que (1) con esto en mente (1)	2	por lo que (1)	1
Interpretación	N-S	por lo que (1)	1	lo que (1)	1
Circunstancia	N-S	desde que (1) mientras (1)	2	a la hora de (1)	1
Evaluación	N-S	—	0	—	0
Reformulación	N-S	es decir (1)	1	—	0
Resumen	N-S	tal y como hemos indicado más arriba (1)	1	—	0
Solución	N-S	—	0	—	0
Disyunción	N-N	—	0	—	0
Capacitación	N-S	—	0	—	0
Conjunción	N-S	—	0	—	0
Evidencia	N-S	—	0	como (1)	1
Alternativa	N-S	—	0	—	0
Unión	N-S	—	0	—	0
Contraste	N-N	—	0	mientras que (2) y (2) o (1) por lo demás (1) frente a (1) por el contrario (1)	8
Condición inversa	N-S	—	0	—	0
<i>Unless</i>	N-S	—	0	—	0
Total marcadores			59		48
Total N-N			16		32
Total N-S			43		16

Tabla 2. Análisis cuantitativo de los marcadores detectados en ambos corpus.

4. Conclusiones

Este trabajo supone una primera aproximación para la detección de rasgos discursivos específicos de textos de dominios especializados diferentes pero que comparten un mismo nivel de especialización y un mismo género textual. Hemos obtenido algunos resultados que parecen indicar que existen algunos rasgos discursivos específicos de cada dominio, pero para constatarlos

es necesario realizar el análisis de un corpus más amplio (que incluya un mayor número de textos, tanto de los dominios aquí tratados como de otros dominios especializados) del que puedan obtenerse resultados estadísticamente significativos.

En este trabajo, las diferencias más relevantes entre los dos dominios (médico y terminológico) tienen que ver con las relaciones discursivas empleadas. En el corpus médico se emplean más relaciones de Resultado, Medio e Interpretación que en el corpus terminológico (en concreto, de Resultado: 7,6% y 4,02%; de Interpretación: 5'8% y 1,15%; de Medio: 8,48% y 3,44%, respectivamente). En cuanto al número de marcadores, en el corpus médico se emplean 48, mientras que en el corpus terminológico se utilizan 59. Esto quiere decir que el número de relaciones evidenciadas por marcadores discursivos en estos corpus supone un 21,42% y un 33,9%, respectivamente.

Como trabajo futuro, seguiremos la metodología de este trabajo para realizar el análisis de otros corpus más amplios de diversos dominios especializados. En concreto, conformaremos un corpus que incluirá, además de los dos ámbitos tratados en este estudio, textos de los siguientes ámbitos: sexualidad, matemáticas, biología y economía. Los textos de sexualidad se extraerán del corpus sobre sexología realizado en el marco de un proyecto del Grupo de Ingeniería Lingüística (GIL) del Instituto de Ingeniería (II) de la Universidad Nacional Autónoma de México (UNAM). Los textos de matemáticas y de biología se extraerán del corpus sobre textos de manuales escolares realizado en el marco del proyecto COCIEM, del Colegio de México, en colaboración también con el GIL. Los textos de economía se extraerán del Corpus Técnico del Institut Universitari de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra (UPF) de Barcelona.

Asimismo, se prevé analizar textos de ámbito general, en concreto noticias de la prensa escrita, para poder realizar comparaciones entre la estructura discursiva de textos especializados y la de textos generales. Estos textos también se extraerán del Corpus Técnico del IULA.

Iria da Cunha

Instituto de Ingeniería (UNAM)

Institut Universitari de Lingüística Aplicada (UPF)

Laboratoire Informatique d'Avignon (UAPV)

Gerardo Sierra

Instituto de Ingeniería (UNAM)

Agradecimientos

Agradecemos el apoyo de los siguientes proyectos: CONACyT (número 82050) y PAPIIT-DGAPA (número IN403108).

Referencias

- CABRÉ, M. T. (1998). “Variació pel tema. El discurs especialitzat o la variació funcional determinada per la temàtica: noves perspectives”. *Caplletra, Revista Internacional de Filologia, Tardor*, pp. 137-194.
- CABRÉ, M. T.; ESTOPÀ, R. (2003). “On the units of specialised meaning uses in professional communication”. *Terminology Science and Research*, N.º 1.
- CABRÉ, M. T. (2007). “Constituir un corpus de textos de especialidad: condiciones y posibilidades”. En Ballard, M.; Pineira-Tresmontant, C. (eds.). *Les corpus en linguistique et en traductologie*, pp. 89-106. Arras: Artois Presses Université.
- CABRÉ, M. T.; BACH, C.; DA CUNHA, I.; MORALES, A.; VIVALDI, J. (2010). “Comparación de algunas características lingüísticas del discurso especializado frente al discurso general: el caso del discurso económico”. En *Actas del XXVII Congreso Internacional de AESLA: Modos y formas de la comunicación humana*. Ciudad Real: Universidad de Castilla-La Mancha.
- CIAPUSCIO, G. (1998). “Los resúmenes de la revista *Medicina*: Un enfoque diacrónico-contrastivo”. *Revista Signo y Seña*, N.º 10, pp. 217-243.
- (2003). *Textos especializados y terminología*. Barcelona: IULA.
- GRAETZ, N. (1985). “Teaching EFL students to extract structural information from abstracts”. En Ulijn, J. M.; Pugh, A. K. (eds.). *Reading for professional purposes: Methods and materials in teaching language*. Leuven: Acco.
- KAPLAN, R. B.; CANTOR, S.; HAGSTROM, C.; KAMHI-STEIN, L. D.; SHIOTANI, Y.; ZIMMERMAN, C. B. (1994). “On abstract writing”. *Text* N.º 14 (3), pp. 401-426.
- L'HOMME, M. C. (1993). *Contribution à l'analyse grammaticale de la langue de spécialité : le mode, le temps et la personne du verbe dans quelques textes scientifiques écrits à vocation pédagogique*. Québec: Université Laval.
- MANN, W. C.; THOMPSON, S. A. (1988). “Rhetorical structure theory: Toward a functional theory of text organization”. *Text* N.º 8 (3), pp. 243-28.

-
- MONTOLÍO, E. (2001). *Conectores de la lengua escrita. Contraargumentativos, consecutivos, aditivos y organizadores de la información*. Barcelona: Ariel.
- PORTOLÉS, J. (2001). *Marcadores del discurso*. Barcelona: Ariel.
- SALAGER-MEYER, F. (1991). "Medical English abstracts: How well structured are they?". *Journal of the American Society for Information Science* N.º 42, pp. 528-532.
- SWALES, J. (1981). *Aspects of Article Introductions*. Birmingham: The University of Aston.
- (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- TEUFEL, S.; MOENS, M. (2002). "Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status". *Computational Linguistics* N.º 28, pp. 409-445.
- VAN DIJK, T. A. (1980). *Macro-Structures. An interdisciplinary study of global structures in discourse, cognitions an interaction*. Hillsdale NJ: Erlbaum.
- (1989). *La ciencia del texto*. Barcelona: Paidós.