

LA CARACTERIZACIÓN LINGÜÍSTICA DEL DISCURSO ESPECIALIZADO¹

M. TERESA CABRÉ CASTELLVÍ
CARME BACII MARTORELL
JOSEP M. CASTELLÀ LIDON
JAUME MARTÍ LLOBET

RESUMEN

*Las teorías terminológicas tienden a evolucionar desde p
descrip
lenguajes de esp
satisfactoriamente la terminología dentro del marco comunicativo esp
comunicación, que se enmarca en la Teoría Comunicativa de la Terminología, p
en el conocimiento de los asp
común o general.
Aunque los mencionados asp
(Kocourek, Hoffmann, Lerat, entre otros), en esta comunicación se ap
lingüística de corp
que p
Creemos que nuestro estudio y sus resultados se encuentran en el camino que nos conducirá, p
p
de las causas de las características lingüísticas halladas, y, p
herramientas de identificación semiautomática de los discursos esp*

PALABRAS CLAVE

Terminología, discurso especializado, características lingüísticas, lingüística de corp

ABSTRACT

*Terminology theories tend to evolve from lexicalist p
lexical items (terms), and their insertion in sp
claim it is p
p
Communicative Theory of Terminology, p
of the linguistics asp
Although aforementioned linguistic asp
among others), this p
studies, and allow new relevant features to ap
In our op
and more rounded account of sp
obtained, and on the other hand, to the p
discourses.*

*codes, to discourse p
investigating the causes of the linguistic features*

KEY WORDS

Terminology, sp

1. INTRODUCCIÓN

Durante décadas el estudio de los llamados *lenguajes de especialidad* ha adoptado un punto de vista básicamente lexicalista, priorizando la descripción de los términos como unidades de códigos especializados. Más recientemente, varias líneas de investigación han evolucionado hacia posiciones discursivas, que consideran que puede explicarse de forma más satisfactoria el funcionamiento de las unidades léxicas y de las redes conceptuales que representan en el marco de la comunicación especializada. Esta perspectiva implica la consideración interrelacionada de los muchos factores que inciden en el discurso especializado²:

- ▣ *Factores de orden sociodiscursivo*: emisor y receptor (los especialistas productores del discurso y los lectores oyentes posibles, especialistas o no según el caso), y los ámbitos científicos, profesionales e institucionales en los que están inmersos.
- ▣ *Factores de orden textual*: los géneros, los tipos de texto y todo el conjunto de convenciones retóricas que regulan, a veces de manera muy estricta, la construcción del discurso especializado.
- ▣ *Factores de orden cognitivo*: las ontologías y los conjuntos de proposiciones adoptados como núcleo conceptual, epistemológico —e incluso ideológico— en cada disciplina.
- ▣ *Factores de orden gramatical*: las construcciones sintácticas y los elementos morfológicos gramaticales y léxicos más habituales y, por lo tanto, más útiles para la transmisión del contenido especializado.
- ▣ *Factores de orden terminológico*: los diversos tipos de unidades terminológicas, descritas según su constitución morfológica, sintagmática, etc. y las clases conceptuales que representan.

La terminología, en este conjunto, continúa siendo un elemento determinante para la definición de los textos especializados, pero deja de ser autosuficiente para explicarse por medio de una adaptación ecológica al entorno comunicativo (de hecho, como cualquier elemento lingüístico cuando es considerado desde una perspectiva discursiva). De acuerdo con Cabré (1999: 131), “en nuestra propuesta teórica, que denominamos *Teoría Comunicativa de la Terminología* (TCT) (...), partimos del supuesto de que los términos no son unidades aisladas que constituyen un sistema propio, sino unidades que se incorporan en el léxico de un hablante en cuanto adquiere el rol de especialista por el aprendizaje de conocimientos especializados”, y este aprendizaje y la transmisión de conocimientos se efectúan a través de los discursos especializados. Desde este punto de vista no es anecdótico el cambio de la denominación *lenguaje de especialidad* a *discurso especializado*, ya que el primero representa una concepción muy estática (código, conjunto de piezas de un sistema), mientras que el segundo se inscribe en una perspectiva más procesual y holística cognitiva y social³.

Como es lógico, el análisis de los textos especializados desde un enfoque discursivo convierte en especialmente relevantes cuestiones como la variación y la densidad. La variación afecta al discurso especializado desde dos puntos de vista: el interno, relativo a la variación denominativa, la sinonimia y la polisemia dentro de cada texto, y el externo, concerniente a los géneros, registros y tipos de texto asociados con las secuencias y con el texto en su conjunto, ahora *situado* en un ámbito social de conocimiento. De este modo, el estudio de la variación vertical se podrá enriquecer a partir de su interrelación con otros factores del registro (formalidad, relación interpersonal, tenor funcional, oralidad y escritura, etc.) y con las convenciones retóricas propias de cada género especializado. La densidad del discurso, por su parte, remite a aspectos lingüísticos (densidad terminológica estricta) y a

otros de carácter más cognitivo y comunicativo (densidad conceptual, densidad informativa y, en último término, dificultad de lectura).

Los variados aspectos del discurso especializado que acabamos de enumerar dibujan un programa amplio de investigación que deberá abordarse en los próximos años y a cuyo desarrollo deseáramos poder contribuir. El trabajo que presentamos en esta comunicación, en la línea de la comparación entre el discurso general y el especializado, pretende ser un inicio en esta dirección. Creemos que en la distinción entre ambos tipos de discurso se dan diferencias de orden comunicativo y cognitivo que son las esenciales, pero también de orden lingüístico, que no se han descrito hasta ahora y que pueden resultar relevantes para una caracterización más precisa de la noción de especialización, aspecto crucial en la ruta que hemos emprendido⁴.

La comparación de los elementos gramaticales predominantes en el discurso general y especializado ha sido abordada numerosas veces (Kocourek, Hoffmann, Lerat, entre otros), pero, a nuestro juicio, sin una comprobación empírica suficientemente generalizable basada en corpus reales representativos de la diversidad de las producciones especializadas. El objetivo concreto de la comunicación que presentamos en esta ocasión es caracterizar gramaticalmente el discurso especializado a partir de un conjunto amplio de textos, para verificar o confirmar o refutar los datos aportados por los estudios anteriores.

El valor de este estudio para la lingüística y para el análisis del discurso radica en la caracterización misma del discurso especializado, y para la lingüística aplicada se puede concretar en utilidades tan diversas como el reconocimiento automático de los textos especializados en Internet, la enseñanza de la lengua para fines específicos —es decir, de los discursos especializados— e, incluso, la valoración y la adaptación de la dificultad de los textos en los ámbitos académico, profesional y de los medios de masas (p. ej., divulgación periodística de la ciencia).

2. METODOLOGÍA DEL ESTUDIO

Para establecer el contraste gramatical entre los dos tipos de discurso, general y especializado, utilizamos el *Corpus Técnico del IULA* de la UPF (Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra de Barcelona). Este corpus está constituido por decenas de miles de documentos en catalán, español e inglés, la mayoría de contenido especializado, que suman un total de más de 22 millones de palabras. Todos los documentos de este corpus han sido digitalizados y etiquetados morfológicamente de acuerdo con los criterios del proyecto europeo *EAGLES* (vid. Morel y otros, 1998). De este conjunto, hemos seleccionado únicamente documentos en español, distribuidos en dos grandes grupos:

- A. *Textos generales*: 5.002.121 palabras en 155 documentos del ámbito periodístico. Se trata de 74 ejemplares de *El País* (entre el 20.10.1999 y el 01.05.2005) y 81 ejemplares de *La Vanguardia* (entre el 27.10.2003 y el 09.04.2005), y consisten en la versión digital casi completa de cada periódico, con la mayoría de secciones y géneros, excepto las imágenes y los apartados que son meros listados de datos.
- B. *Textos especializados*: 5.018.193 palabras en 251 documentos de los ámbitos temáticos *Derecho*, *Economía*, *Informática*, *Medio ambiente* y *Medicina*, con aproximadamente un millón de palabras en cada ámbito. Los géneros más habituales son: en *Derecho*, leyes, decretos, convenios, artículos y sentencias; en *Economía*, manuales universitarios; en *Informática*, artículos de revista; en *Medio ambiente*, artículos de revista y manuales; y en *Medicina*, también artículos de revista y manuales.

Por lo que respecta a la representatividad de la muestra, partimos de las consideraciones del famoso trabajo de Biber (1988), que explota un corpus de un millón de palabras repartidas en 23 géneros y lo considera extrapolable al conjunto de la lengua inglesa. Por lo tanto, nuestra cifra total de 10 millones de palabras debería estar fuera de toda duda. Para asegurar aún más, si cabe, la validez de la muestra, hemos tenido en cuenta los géneros discursivos, uno de los aspectos metodológicos más discutibles del estudio de Biber, que presenta un número de palabras desproporcionado entre géneros y en algunos en concreto, claramente escaso (Castellà, 2002 y 2004)⁵.

Después de la constitución del corpus, hemos efectuado búsquedas automáticas de carácter avanzado con el instrumento *hwanaNet* del *Corpus Técnico del IULA*, con el posterior cálculo de proporciones y porcentajes. Tenemos en proyecto una aplicación del análisis multivariante tipo *cluster*, para detectar correlaciones significativas entre los datos obtenidos, y un análisis cualitativo más detenido sobre documentos concretos, con búsquedas manuales en los casos que sea preciso.

3. RESULTADOS

A continuación presentamos una muestra de resultados cuantitativos agrupados en tres apartados: nombres, verbos y adjetivos. Expresamos entre paréntesis el valor porcentual relativo de las cifras obtenidas para el discurso general y especializado en cada elemento gramatical y marcamos en **negrita** los elementos gramaticales más relevantes para el contraste entre los dos tipos de discurso —consideramos que tienen un grado suficiente de relevancia los que superan una diferencia del 20 % (es decir, del 40 % al 60 % o viceversa).

TABLA 1	DISCURSO GENERAL	DISCURSO ESPECIALIZADO
Nombres	1.218.815 (48,4 %)	1.302.211 (51,6 %)
Nombres comunes	935.355 (44,6 %)	1.164.296 (55,4 %)
Nombres propios	283.460 (67,3 %)	137.915 (32,7 %)
Nombres en <i>-ción</i>	64.120 (31,5 %)	139.224 (68,5 %)

TABLA 2	DISCURSO GENERAL	DISCURSO ESPECIALIZADO
Verbos	684.530 (52,3 %)	624.766 (47,7 %)
Verbos en presente	287.983 (47,9 %)	312.423 (52,1 %)
Verbos en pasado	148.318 (78,7 %)	40.079 (21,3 %)
Verbos en 1ª pers. sing.	23.270 (65,1 %)	12.472 (34,9 %)
Verbos en 2ª pers. sing.	4.214 (56,8 %)	3.210 (43,2 %)
Verbos en 1ª pers. plural	12.973 (48,6 %)	13.718 (51,4 %)
Verbos en 2ª pers. plural	311 (74,6 %)	106 (25,4 %)

TABLA 3	DISCURSO GENERAL	DISCURSO ESPECIALIZADO
Adjetivos calificativos	381.813 (47,0 %)	430.576 (53,0 %)
Nombre + adjetivo calif.	150.386 (39,9 %)	225.856 (60,1 %)
Nombre + sintagma prep.	244.635 (44,6 %)	303.469 (55,4 %)

De entre los rasgos gramaticales consignados en las tablas precedentes, los más destacados en el contraste entre discurso general y especializado son los siguientes:

- ❑ *Nombres propios*: 67,3 % (general) vs. 32,7 % (especializado). El número mayor de nombres propios en la muestra de discurso general puede estar motivado por su pertenencia al ámbito de la prensa. De todos modos, en muchos otros ámbitos no especializados, desde la conversación cotidiana a la narración literaria, la presencia de nombres también suele ser considerable, por lo que este resultado puede ser considerado pertinente.
- ❑ *Nombres en -ción*: 31,5 % (general) vs. 68,5 % (especializado). Se trata de un resultado esperable: una concurrencia superior de las nominalizaciones deverbales en el discurso especializado, en correspondencia con su carácter más conceptual y abstracto.
- ❑ *Verbos en pasado*: 78,7 % (general) vs. 21,3 % (especializado). Predominio de los pretéritos en el corpus de prensa, que tiene la mayoría de géneros (noticia, reportaje, crónica) referidos al pasado y, principalmente, al día anterior. Esta característica idiosincrática de la prensa no merma validez al resultado, ya que corrobora el hecho de que el discurso especializado se muestra reticente a la narración y cuando la incorpora, suele rebajar automáticamente el grado de especialidad.
- ❑ *Verbos en 1ª persona del singular*: 65,1 % (general) vs. 34,9 % (especializado). Como era de esperar, los verbos en primera persona del singular se encuentran más en el discurso general que en el especializado, pero cabe destacar que el peso que tienen en el especializado es considerable, especialmente si se compara con la 1ª persona del plural en el mismo tipo de discurso (13.718 y 51,4 %), que teóricamente tendría que haber sido mucho más habitual.
- ❑ *Verbos en 2ª persona del plural*: 74,6 % (general) vs. 25,4 % (especializado). Este dato resulta poco válido a causa del bajo número absoluto de ocurrencias (331 y 106, respectivamente, sobre un corpus de 10 millones) y, por el mismo motivo, poco aplicable a instrumentos automáticos de reconocimiento.
- ❑ *Nombre + adjetivo calificativo*: 39,9 % (general) vs. 60,1 % (especializado). El discurso especializado se distingue por la expansión adjetiva de los nombres, en su afán por conseguir la precisión, a menudo mediante unidades terminológicas satisfactoriamente formadas por N + adjetivo.

Por otra parte, respecto a estudios anteriores en el campo de la terminología y los lenguajes de especialidad, hemos obtenido resultados innovadores. De entre los presentados en las tablas, vamos a destacar dos: a/ se sabía que los nombres propios abundaban en el discurso especializado, pero no que fueran la mitad que en el general, y b/ se había dicho y repetido que la 1ª persona del plural era la típica de los textos especializados y ha resultado que la 1ª del singular lo es casi en la misma medida.

4. CONCLUSIONES

A la vista de los datos analizados, existen contrastes cuantitativos importantes en la frecuencia de aparición de determinados elementos gramaticales en el discurso general y especializado. Esas diferencias habrán de ser justificadas en un futuro próximo a partir de análisis más detallados sobre los textos.

Por otra parte, los resultados muestran diferencias con las caracterizaciones del discurso especializado realizadas con anterioridad. La presente investigación, desarrollada en base a corpus reales, ha permitido confirmar o bien refutar parte de las asunciones vigentes hasta el momento. A nuestro parecer, será necesario ahondar en esta línea de trabajo, explicando las causas de los fenómenos registrados y, posteriormente, continuar con las siguientes tareas de descripción del discurso especializado esbozadas en el programa de investigación que hemos expuesto al inicio de esta comunicación.

NOTAS

1. La redacción de esta comunicación, así como la investigación en la que se sustenta, se han realizado como actividades del proyecto *TEXTTERM2 - Fundamentos, estrategias y herramientas para el procesamiento y extracción automática de la información especializada*, BFF2003-02111, Dirección General de Enseñanza Superior, Dirección General de Ciencia y Tecnología, 01.01.04-31.12.06 y en el marco del grupo consolidado IULATERM del IULA (Institut Universitari de Lingüística Aplicada) de la Universitat Pompeu Fabra de Barcelona.

2. Esta lista de factores no se aleja de las más habituales en lingüística del texto y análisis del discurso y es próxima a la que presenta Ciapuscio (2003: 103), aunque tenga algunas particularidades. El trabajo de Ciapuscio es una muestra del enfoque discursivo que comentamos, que esta autora explica con carácter programático de la forma siguiente:

"(...) el propósito general de intentar mostrar que una confluencia entre dos disciplinas como la lingüística del texto y la terminología de base lingüística, apoyada en un enfoque léxico comprensivo como la lexicología vertical (Wichter, 1994), puede proveer un marco interesante para la solución de problemas teórico-metodológicos que incumben a ambas disciplinas. La aspiración es superar cierta escisión entre las dos tradiciones de investigación, aspiración que evidentemente es compartida por otros estudiosos dedicados a problemas de las unidades léxicas especializadas y de los textos especializados." (Ciapuscio, 2003: 115)

3. La crítica a la denominación *lenguaje de especialidad* se basa también en su carácter exclusivista:

"La denominació *llenguatges esp* suposa, en la nostra opinió, un ús abusiu del terme llenguatge. En efecte, si en lloc de partir de la noció de llenguatges d'especialitat com un supòsit assumit, analitzem les produccions lingüístiques que es donen en els contextos de comunicació especialitzada per inferir si s'han fet servir per a la seva producció sistemes d'elements lingüístics específics (és a dir, diferenciats dels del llenguatge comú), fàcilment observarem que, des del punt de vista dels recursos gramaticals que els constitueixen, els textos especialitzats són textos codificats a partir de la mateixa gramàtica que s'utilitza en totes les situacions comunicatives." (Cabré, 1999: 152)

4. Existe, además, la interesante discusión teórica sobre el carácter dicotómico o gradual de la frontera entre generalidad y especialización, que no podemos tratar aquí y que nos gustaría abordar a fondo en otra ocasión.

5. Somos conscientes de que el presente trabajo deberá completarse con corpus especializados de clara función didáctica y divulgadora, ya que los documentos del corpus del IULA son básicamente ensayísticos. El corpus general, por su parte, también deberá completarse con textos de procedencia distinta a la periodística. A pesar de ello, el corpus constituido nos sirve para explorar nuestras ideas de partida.

BIBLIOGRAFÍA

- Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Cabré, M. T. 1999. *La terminologia: representació y comunicació. Elements para una teoria de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Cabré, M. T. 2005. «Quelques indices linguistiques pour l'identification des textes de spécialité». Conferencia por invitación en el Laboratoire CLILLAC (CIEL) de la Université Paris 7 y el Département de langues de l'ENS Cachan.
- Castellà, J. M. 2004. *Oralitat i escriptura: dues cares de la complexitat en el llenguatge*. Barcelona: Curial-Publicacions de l'Abadia de Montserrat.
- Ciapuscio, G. E. 2003. *Textos especializados y terminología*. Barcelona: Institut Universitari de Lingüística Aplicada (Universitat Pompeu Fabra).
- Hoffmann, J. 1998. *Llenguatges d'especialitat. Selecció de textos*. Edición a cargo de J. Brumme. Barcelona: Institut Universitari de Lingüística Aplicada (Universitat Pompeu Fabra).
- Kocourek, R. 1991 (1982). *La langue française de la technique et de la science*. 2ª ed. Wiesbaden: Brandstetter.
- Lerat, P. 1995. *Les langues spécialisées*. París: Presses Universitaires de France.