

El grado de especialización textual del discurso médico en español: análisis basado en corpus desde la teoría de la percepción

Sergio Rodríguez-Tapia*

Resumen: Este trabajo analiza el grado de especialización textual del discurso médico. Para ello, a partir de los fundamentos discursivistas de la teoría comunicativa de la terminología y de la teoría sociocognitiva de la percepción, se ha analizado una base de datos en Excel que contiene información sobre seis variables, tanto lingüísticas como extralingüísticas (el tipo de texto, la variedad léxica, la reformulación, la densidad terminológica, el papel de los interlocutores y la clase de texto), empleando un método correlacional con enfoque cuantitativo-cualitativo que ha sido complementado con aprendizaje automático. Entre las conclusiones, cabe destacar el papel que desempeña el receptor como condicionante del grado de especialización o la densidad terminológica como principal factor discriminante del grado de especialización.

Palabras clave: corpus, densidad terminológica, grado de especialización, hablante, percepción, reformulación, texto especializado.

The degree of text specialization of the medical discourse in Spanish: a corpus-based analysis using perception theory

Abstract: This paper analyzes the degree of text specialization in medical discourse. To this end, based on the discursive underpinnings of the Communicative Theory of Terminology and the socio-cognitive theory of perception, an Excel database with data on six variables, both linguistic and extra-linguistic (type of text, lexical variety, rewording, terminological density, role of the interlocutors, and text category), was analyzed using a correlational method with a quantitative-qualitative approach, complemented with machine learning. Among the conclusions, it is worth highlighting the role played by the recipient as a conditioning factor of the degree of text specialization, as well as the terminological density as the main discriminating factor of the degree of text specialization.

Key words: corpus, degree of specialization, perception, rewording, speaker, specialized text, terminological density.

Panace@ 2020; XXI (51): 75-92

Recibido: 31.I.2020. Aceptado: 15.IV.2020.

1. Motivación del estudio

El aumento de la divulgación del conocimiento científico médico (entendida como la publicación de entradas en blogs, la creación de espacios dedicados a la medicina dentro de los periódicos, la edición de monografías y volúmenes de divulgación médica o la emisión de espacios televisivos o series de televisión sobre medicina, entre muchos otros fenómenos) lleva asociado el empleo de terminología actualizada en diferentes tipos de discurso, pues se orientan a todo un abanico de nuevos interlocutores.

Estas diferencias están vinculadas a una recontextualización pragmática, ya que el esquema prototípico de emisor especialista-receptor especialista desaparece: a veces no solo cambian los receptores del mensaje, sino también los emisores, que pueden llegar a ser no especialistas (periodistas, especialistas en divulgación científica, especialistas de ramas afines, etc.). Así pues, la terminología médica recibe un tratamiento diferente, puesto que puede ir vinculada a procedimientos de reformulación diferentes o bien su aumento de frecuencia en los medios de masas puede provocar que sus significados sean percibidos de forma laxa o con imprecisión semántica.

El aumento de la divulgación de este conocimiento guarda relación con la motivación profesional o personal de diferentes tipos de receptores según su competencia cognitivo-comunicativa, lo que hace que las diferentes expectativas informativas de cada receptor puedan ser identificadas en diversos tipos textuales que siguen ciertas convenciones textuales.

La mayor o menor abstracción con la que se aborde el tema del texto hace que pueda hablarse de diferentes grados de especialización, que satisfarán en mayor o menor medida las expectativas informativas del receptor (véanse los trabajos de Domènech, 2006: 49; o Prieto Velasco, 2008: 71; en torno al grado de especialización). He aquí el interés central de nuestro trabajo:

* Universidad de Córdoba, Córdoba (España). Dirección para correspondencia: sergio.rodriguez@uco.es.

describir las características del grado de especialización textual del discurso médico, con especial atención a la vinculación entre la relación emisor-receptor, la variedad léxica, el tipo textual, la reformulación y la densidad terminológica.

2. Objeto, hipótesis y objetivos

El objeto que tratamos en este estudio corresponde al grado de especialización textual, también denominado *eje vertical* (Hoffmann, 1998 [1987]: 69), que constituye el grado de abstracción con el que un emisor se aproxima a un texto. Este grado es actualizable de forma lingüística (Cabré, 2002^a: 89; Cabré *et al.*, 2001: 184), es decir, puede ser reconocido a partir del análisis de su densidad terminológica, variedad léxica, de su reformulación o de la relación entre emisor y receptor desde el punto de vista de su competencia cognitivo-comunicativa.

En este trabajo se esbozan las siguientes hipótesis:

— El receptor es el factor que condiciona las características lingüísticas de los diferentes textos en el marco del grado de especialización.

— La densidad terminológica, en especial la densidad terminológica de términos crípticos, es el principal factor discriminante del grado de especialización textual y lo que permite distinguir entre textos especializados, semiespecializados y no especializados, si bien no es el único elemento que contribuye a su distinción.

— La relación entre emisor y receptor, desde el punto de vista de su competencia cognitivo-comunicativa, está vinculada a los tipos textuales seleccionados por el emisor para transmitir la información al receptor.

— La reformulación terminológica, al contrario que la idea tradicional, está asociada de forma directa al grado de especialización textual, de forma que cuanto mayor es el grado de especialización, mayor es el grado de reformulación al que se somete la terminología.

Nos basamos en la idea de Loureda (2002: 158) de que la predicción de características de ciertos textos sirve de instrumento de elaboración para textos similares del mismo tipo, puesto que «los tipos de texto equivalen a representaciones ideales formadas por el común de una serie de discursos». En definitiva, los objetivos de este trabajo son tres:

— Identificar qué factores se relacionan con el grado de especialización y en qué medida lo condicionan o modifican.

— Considerar la teoría de la percepción como herramienta de análisis que permite identificar convenciones colectivas a partir de análisis subjetivos.

— Esbozar los cimientos teóricos y determinar una propuesta de marco metodológico de partida sobre los cuales puedan realizarse en el futuro investigaciones con grandes cantidades de datos. A partir de la confirmación o refutación de nuestras hipótesis, este estudio pretende servir de fun-

damento teórico para la lingüística aplicada, en concreto a aquellos campos relacionados con la redacción de discursos, la estilística, la traducción, la terminografía o la misma lingüística de corpus para reflexionar sobre los diversos tipos de textos relacionados con el campo médico y su grado de especialización.

3. Fundamentación teórica

Para elaborar este artículo, nos hemos basado en un conjunto de concepciones teóricas que sintetizamos a continuación:

— Adoptamos como marco la lingüística aplicada como dimensión orientada a intentar dar respuesta a los problemas comunicativos de la sociedad actual (Payrató, 1998: 33; Berns y Matsuda, 2006: 394).

— Seguimos una perspectiva funcionalista de la comunicación, que percibimos como un objeto dinámico (Bernárdez, 1995: 137), lo que contribuye a explicar la variación textual (ya sea discursiva o denominativa) por factores extralingüísticos. En concreto, en el presente trabajo, concebimos el texto según señala Bernárdez (1982: 85), acotado por las dimensiones coserianas que sugiere Loureda (2002) de acuerdo con las normas de textualidad de Beaugrande y Dressler (1997: 35) y los niveles de estructuración textual (especialmente la superestructura) de Dijk (1992 [1983]), y caracterizada por los rasgos de «impredictibilidad» y diversa interpretación de Zamorano (2012) sobre la comunicación verbal humana.

— Percibimos el texto especializado como fruto de los procesos inherentes a la comunicación especializada y regido por las normas del sistema lingüístico. Por ello, queda concebido como una variedad de lengua, por una parte, diáfónica (en el sentido de registro funcional) y, por otra, diatrática (en el sentido de actividad verbal humana de ciertos grupos socioprofesionales) (Rodríguez-Tapia, 2019: 89). En tanto que objeto variable, puede ser descrito en diferentes grados de actualización lingüística o *continuum* (Rodríguez-Tapia, 2019: 171), lo que nos permite hablar de un grado de especialización mayor o menor, con características distintivas cada uno de ellos.

— En cuanto a la terminología como campo de análisis, adoptamos la teoría comunicativa de la terminología (ТСТ; Cabré, 1998 y 1999). De ella, tomamos el principio de poliedricidad (Cabré, 2008), el principio de variación (Cabré, 1998: 72) y los fundamentos discursivistas para el análisis del léxico (Adelstein, 2007: 277). No todos los aspectos de la cognición (una de las perspectivas de análisis del término según la ТСТ) nos interesan en este trabajo. Por ello, prestamos atención a la noción de percepción de la teoría de Caravedo (2014), que desarrollamos posteriormente, y precisamos que no abordamos este estudio desde teorías cognitivas de la terminología (como la teoría sociocognitiva de la terminología o la terminología basada en marcos), por tratarse de teorías orientadas, respectivamente, a la terminografía o la cate-

rización de los términos (Temmerman, 2002: 73 y 78) o a la representación del conocimiento especializado (Montero y Faber, 2008: 50), lo cual se aleja de nuestras necesidades teóricas y metodológicas en el actual estudio, si bien se emplearán para explicar otros fenómenos relacionados con el corpus en trabajos posteriores.

El elemento clave de nuestro objeto de estudio, el texto especializado, podría quedar caracterizado desde el punto de vista teórico principalmente según cuatro factores (Cabré *et al.*, 2001: 181; Ciapuscio y Kuguel, 2002: 43; Cabré, 2007: 90; Da Cunha *et al.*, 2011: 266):

— superestructural: factor relacionado con las clases, géneros o tipos textuales —según la terminología que emplee cada autor— que selecciona específicamente dicho texto (Cabré y Estopà, 2005: 36);

— pragmático, especialmente relacionado con la teoría del registro de Halliday (1982 [1976]);

— cognitivo, en tanto que representa y comunica conocimiento especializado; y

— lingüístico (Cabré *et al.*, 2011: 302), entre los que se distinguen elementos característicos en el nivel gramatical (Cabré, 2007: 105 y ss.: predominio de sustantivos, ausencia de segunda persona del singular y plural, presencia de la tercera y del sujeto impersonal, etc.) y, específicamente, en el nivel léxico, ampliamente reconocido por los investigadores por ser característico del texto especializado (Arntz y Picht, 1995: 45; Monterde Rey, 2002: 115; Domènech, 2007: 250; Cabré *et al.*, 2014: 224).

En este estudio, contaremos con estos factores como variables de análisis: el tipo de texto según el factor superestructural, el tipo de interlocutor según el factor pragmático, algunos de los elementos cognitivos serán tratados desde la teoría de la percepción, y los elementos léxicos servirán para determinar la densidad de formas léxicas especializadas con respecto al total de formas léxicas, de modo que pueda identificarse la codificación del conocimiento especializado en las unidades léxicas del texto.

En relación con el factor pragmático, de acuerdo con Díaz Rojo (2000: 70-72), en la comunicación especializada podría hablarse de dos formas de relación entre los interlocutores del acto comunicativo, en principio, opuestos entre sí: *a*) entre especialistas y *b*) entre especialista y sociedad (niveles de investigación y divulgación). Entre ellos, es posible identificar una gradación.

Cabe detenerse a comentar las implicaciones que supone estudiar el papel que desempeñan los interlocutores, especialmente teniendo en cuenta la relevancia de los elementos pragmáticos en la TCT y la concepción de texto como producto social desde la perspectiva de Bernárdez (1995) y Zamorano (2012). Considérese como modelo el siguiente enunciado:

«El virus infecta las vías respiratorias y causa síntomas que van desde un cuadro leve (tos seca, fiebre...) a graves

dificultades para respirar y una neumonía potencialmente mortal» (Güell, 2020).

Si bien la mayoría de estas unidades léxicas es frecuente en el discurso común de los hablantes, podría defenderse que más de un usuario podría considerar terminológicas algunas de ellas, como *virus* o *neumonía* (incluso *fiebre* o *vías respiratorias*). Los criterios que emplean los usuarios de la lengua para clasificar las unidades lingüísticas como terminológicas atienden a su percepción sociocognitiva y no son claros, puesto que dependen individual y colectivamente de forma simultánea. El hecho de contemplar la posibilidad de que una unidad léxica sea considerada terminológica o no implica debatir sobre el valor terminológico de dichas unidades (Cabré, 1999: 132), que queda activo por criterios semántico-pragmáticos, especialmente vinculados a la situación comunicativa y, en concreto, como queda reflejado en Domènech (2006: 72) o Rodríguez-Tapia (2019: 466), quedan condicionados por el receptor. Y es que debemos tener en cuenta que, aunque se trate de la misma denominación, *virus*, *neumonía* o *fiebre* pueden suponer concepciones o dimensiones de los conceptos diferentes para un lego en la materia, para un médico de familia o para un epidemiólogo, debido al diferente nivel de conocimiento experto. A estas diferencias también contribuyen los fenómenos de banalización (aumento de frecuencia de un término en el discurso común) y de terminologización (adquisición de un sentido especializado por parte de una unidad léxica usada prototípicamente en el discurso común), que modifican el grado de contacto que tienen los usuarios con el léxico disponible en una lengua. El principal problema que encontraríamos en este estudio sería, por tanto, contar con la información suficiente para clasificar una unidad como terminológica o como no terminológica.

En este estudio no se plantea analizar la densidad cognitiva o el significado de la terminología de los textos, sino simplemente codificar el fenómeno del valor terminológico de forma cuantitativa para identificar su relevancia en la clasificación de los textos según el grado de especialización. Para ello, emplearemos la información analizada por un conjunto de informantes que detallaremos más adelante.

En relación con los interlocutores, cabe mencionar que la clasificación que aquí se presenta es una propuesta que pretende justificar el *continuum* (Cabré, 2002^b: 34; Ciapuscio y Kuguel, 2002; y Ciapuscio, Kuguel y Otañi, 2005) en el grado de especialización que defendemos, por lo que, partiendo de la dicotomía de Díaz Rojo (2000: 70-72), se distingue una clasificación tripartita de interlocutores según su competencia cognitivo-comunicativa (Rodríguez-Tapia, 2019: 47-49), es decir, según su capacidad de crear, comunicar, comprender y precisar el conocimiento especializado mediante el código lingüístico, especialmente sus innovaciones o reflexiones. Relacionamos *grosso modo* algunas de sus características:

— Especialista: interlocutor con motivación profesional y formación académica o profesional, competente en el desarrollo de reflexiones y argumentos sobre el conocimiento especializado.

— Instruido: interlocutor con motivación personal o profesional que conoce a grandes rasgos el campo especializado, no necesariamente por su formación académica o profesional, sino por su labor de documentación o información complementaria.

— Lego: interlocutor que no es capaz de reflexionar sobre el conocimiento especializado y que no posee formación en la materia, a pesar de que posea (o no) motivación personal al respecto.

Esta clasificación tripartita podría observarse como excesivamente simplista, si bien una clasificación más amplia en seis subcategorías (especialista investigador, especialista no investigador, semiespecialista, aprendiz de especialista, lego formado y lego no formado) fue empleada en otro estudio (Rodríguez-Tapia, 2019) cuyos resultados sugieren que, de entre las seis categorías, tres de ellas siempre suelen ser más frecuentes en la percepción de los usuarios, por lo que suelen quedar sobrerrepresentadas en detrimento de otras categorías. Esto invita a reflexionar que quizá una especificación minuciosa de perfiles de interlocutores según su competencia cognitivo-comunicativa no sea tan útil para el objetivo de taxonomizar o no sea fácilmente distinguible por parte de los usuarios que leen los textos.

4. Metodología

4.1. Método y variables de análisis

Este estudio constituye un análisis correlacional con enfoque mixto cuantitativo-cualitativo: en primer lugar, se analizan los textos del corpus, que pasan a codificarse en datos numéricos recogidos en una base de datos en Excel; a partir de este análisis cuantitativo, se extrae la información y se ponen a prueba las hipótesis mediante un análisis cualitativo. Como herramientas, empleamos la lingüística de corpus, la teoría de la percepción y el aprendizaje automático. En este estudio se analiza un conjunto de siete variables:

TABLA 1. Relación de variables, valores y criterios de análisis

Clase de texto	Valores	— Texto especializado (TE) — Texto semiespecializado (TSE) — Texto no especializado (TNE)
	Observaciones	Variable ordinal que corresponde a la variable dependiente que investigamos, puesto que pretendemos vincular cierto número de características a cada uno de los valores de la clase de texto. Viene determinada por los grupos de informantes.

Tipo de texto	Valores	— Artículo científico (ART) — Artículo de revista o web de divulgación especializada (DIV) — Consultas y respuestas en foro en línea (FOR) — Entrada de blog (BLOG) — Entrevista (ENT) — Noticia (NOT) — Tesis doctoral (TES)
	Observaciones	Variable nominal que se corresponde con la noción de superestructura de Dijk (1992 [1983]). Sigue una clasificación de características prototípicas. Viene determinado <i>a priori</i> por el investigador del estudio. Constituye uno de los criterios de compilación del corpus.
Emisor / Receptor	Valores	— Especialista (Esp) — Instruido (Ins) — Lego (L)
	Observaciones	Variables ordinales según el perfil cognitivo-comunicativo de los interlocutores de los textos del corpus. Vienen determinadas por un grupo de informantes.
Ratio <i>types/tokens</i> estandarizada (TTR)	Valores	Valor numérico mayor que 0.
	Observaciones	Variable cuantitativa que permite analizar la diversidad léxica, pues establece la relación entre tipos de formas léxicas (<i>types</i>) y formas léxicas (<i>tokens</i>), según la fórmula de Templin (1957): cuanto mayor es el porcentaje, mayor es la diversidad léxica de un texto, es decir, es más variado en léxico y no tan repetitivo. Estas operaciones se realizan computando exclusivamente aquellas formas consideradas no gramaticales mediante una lista de exclusión (de manera que no se tienen en cuenta las preposiciones, conjunciones, etc.). Dado el objetivo del estudio, no debatimos acerca de la noción de unidad léxica, y optamos por tratar como forma léxica toda forma lingüística con lexema, ya sea simple o compuesta (por ejemplo, estadísticamente, y no ontológicamente, en este estudio, el segmento <i>neuralgia del trigémino</i> se consideraría constituido por dos formas léxicas que, a su vez, pueden ser consideradas terminológicas o no), si bien desde la fraseología o la morfología podría considerarse una única unidad lingüística. Los valores relacionados con el léxico han sido contabilizados con WordSmith Tools 7.0.

Índice de densidad terminológica (IDT)	<p>Valores</p> <p>Valor numérico entre 0 y 1.</p>
	<p>Observaciones</p> <p>Variable cuantitativa que resulta de dividir el número total de formas léxicas especializadas o con valor terminológico entre el número total de formas léxicas del texto. Permite codificar la cantidad de conocimiento especializado que representan las formas léxicas en un texto. A su vez, el IDT distingue tres subtipos:</p> <ol style="list-style-type: none"> 1. IDT a) o IDT de términos délficos: el usuario conoce la denominación y el concepto, aunque no necesariamente de forma precisa. Por ejemplo, <i>fiebre</i>. En este trabajo se marcarán con el color turquesa. 2. IDT b) o IDT de términos semicrípticos: el usuario conoce la denominación por su frecuencia, pero no el concepto, o conoce este último de forma muy laxa. Por ejemplo, <i>polivitamínico</i>. En este trabajo se marcarán con el color verde. 3. IDT c) o IDT de términos crípticos: el usuario desconoce la denominación y el concepto. Por ejemplo, <i>lipolítico</i>. En este trabajo se marcarán con el color amarillo.
Índice de reformulación (IR)	<p>Valores</p> <p>Valor numérico correspondiente a 0 o mayor que 0. Con el objetivo de facilitar la comparación de datos, estos valores se han normalizado, de forma que aquellos valores mínimos se presentan como próximos a 0, y los valores máximos (que pueden superar la unidad) se aproximan a 1 por debajo. De esta forma, todos los valores de IR se encuentran entre 0 y 1.</p>
	<p>Observaciones</p> <p>Variable cuantitativa que resulta de dividir el número total de procedimientos de reformulación formal de un texto (uso de paréntesis, formas denominativas alternativas, marcadores como <i>es decir</i> u <i>o</i>, etc.) entre el índice de densidad terminológica de dicho texto. Permite codificar el fenómeno de reformulación al que está sometido un texto según la cantidad de terminología relativa a dicho texto. A su vez, el IR distingue tres subtipos (que siguen la propuesta de Garcés, 2008: 73), que ilustramos mediante ejemplos en los que aparece en negrita el término reformulado y subrayado el procedimiento de reformulación:</p> <ol style="list-style-type: none"> 1. IR a) o IR de procedimientos por expansión: «También pueden presentarse sibilancias (<u>un sonido de silbido al respirar</u>) y presión torácica» (Paricio, 2017). 2. IR b) o IR de procedimientos por reducción: «El dolor también puede aumentar repentinamente incluso si está siendo tratado. <u>Esto recibe el nombre de dolor irruptivo</u>» (CancerNet, 2016). 3. IR c) o IR de procedimientos por variación: «Los síntomas diagnósticos son: disnea (<u>ahogo</u>), expectoración anormal» (Paricio, 2017).

4.2. Criterios de compilación del corpus

Siguiendo los criterios de Torruella (2017: 41-42) de clasificación de los corpus, hemos elaborado un corpus que atiende a los siguientes criterios: se trata de un corpus textual monolingüe y sincrónico en español que cubre los años entre 2007 y 2017, de textos escritos y planificados sobre medicina, simple (puesto que no ha sido anotado o etiquetado) y equivalente (puesto que se ha perseguido la homogeneidad cuantitativa en cada texto, principalmente en el número de textos).

Desde el punto de vista de su representatividad cualitativa, el texto está compuesto por los siete tipos textuales relacionados en la sección de variables. Estos textos han sido sometidos a ciertos procesos de selección de la información (por ejemplo, se ha eliminado la información relacionada con autores, referencias bibliográficas, índices de contenido, etc.). En el caso de los textos extensos, como las tesis doctorales, se ha seleccionado aleatoriamente un fragmento textual de cada sección del documento, de forma que se garantice una muestra de cada apartado (Torruella, 2017: 146).

Desde el punto de vista de su representatividad cuantitativa, dado que no es posible seleccionar la totalidad de los textos existentes que cumplan con las características que determinamos como criterio de partida en nuestro corpus (McEnery y Hardie, 2012: 1-23), hemos seleccionado 15 textos por cada tipo textual (en total, 105 textos). Estos 105 textos han sido analizados por 4 grupos de informantes, de forma que los resultados arrojan 420 análisis. Pretendemos garantizar el equilibrio interno del corpus, si bien debe mencionarse que aquellas secciones descompensadas corresponden a la realidad textual: tanto los artículos científicos como las tesis doctorales suelen ser textos más extensos, lo que queda reflejado en el corpus. Debe tenerse en cuenta, no obstante, que el análisis se realiza atendiendo estrictamente al número de patrones (es decir, de textos), que está compensado al máximo. El número de palabras solo sirve para obtener los valores relativos a la densidad terminológica o la reformulación dentro de cada texto, por lo que en ningún momento la descompensación del número de palabras afectaría a otra categoría del corpus (tabla 2).

4.3. La teoría de la percepción como instrumento

Previamente se adelantaba el debate sobre la dificultad que existe para determinar el valor especializado de un término. Entre los criterios empleados para ello se encuentran los sintetizados por L’Homme (2020: 72-75), que distinguen la vinculación a un campo concreto, la naturaleza de los argumentos, las relaciones morfológicas o las relaciones paradigmáticas. Otros criterios están relacionados con el nivel estadístico, como las técnicas aplicadas en el análisis contrastivo de palabras clave (Tercedor y Láinez, 2017: 32) o los grafos de coocurrencia léxica (Nazar y Cabré, 2011: 53), entre otros.

En nuestro trabajo, usaremos como punto de partida cómo los hablantes acudimos a un conjunto de juicios que usamos como elementos discriminantes del valor especializado. La percepción es una capacidad cognitiva que empleamos de forma

TABLA 2. Número de palabras y de formas léxicas por tipos textuales en el corpus

	ART	ENT	BLOG	NOT	FOR	DIV	TES	TOTAL
número de palabras	62 498	23 845	14 284	11 555	13 845	12 056	44 754	182 837
porcentaje total	34	13	8	6	8	7	24	100
número de formas léxicas	31 843	12 353	7237	5832	6949	6105	23 070	93 389
porcentaje total	34	13	8	6	7	7	25	100

generalizada, no exclusivamente sobre elementos lingüísticos, si bien la percepción lingüística es la que más nos interesa en este estudio.

La percepción se origina como un desajuste de una relación biunívoca entre un elemento y su valor prototípico. Por ejemplo, la percepción permite discriminar si consideramos que un enunciado ha sido emitido por un nativo o no, teniendo en cuenta si la entonación, la selección léxica o la sintaxis se ajustan o se alejan del comportamiento prototípico de un hablante nativo (lo que resulta, por tanto, en variación, puesto que se desvía de la norma). Entendido el grado de especialización textual como variación diastrático-diafásica (en registro y tipo de usuario), es posible defender la utilidad de la teoría de la percepción como instrumento de análisis o como base teórica sobre la que ejercer el análisis lingüístico. La percepción lingüística funciona según dos elementos: la ocurrencia y la prominencia (Caravedo, 2014: 115-116). La primera se basa en la recurrencia de ciertos fenómenos en el acto comunicativo (por ejemplo, la ocurrencia de formas léxicas no terminológicas en el discurso no especializado frente a la ocurrencia de formas terminológicas en el discurso especializado). La segunda es el fenómeno contrario: sucede cuando un hablante hiperpercibe de forma subjetiva elementos que sobresalen, normalmente por falta de frecuencia, y que quedan vinculados a otros condicionantes. Por ejemplo, una entonación alterada, una nasalización en el tono o una selección léxica errónea permitirían vincular un enunciado a un hablante no nativo, o, en el caso que nos ocupa, la presencia de numerosas formas léxicas especializadas en un texto no especializado llamaría la atención del hablante, lo que iniciaría los procesos subjetivos hacia una actitud, valoración y posterior configuración de la percepción (ídem), la cual vincula dicho texto con uno no prototípicamente divulgativo. En suma, a partir de los elementos lingüísticos, los hablantes son capaces de tomar decisiones de discriminación o clasificación de fenómenos basándose en su percepción.

Con todo, que el hablante ejecute la percepción desde la subjetividad no implica que sus decisiones no puedan valorarse desde la ciencia, puesto que, en palabras de Caravedo (2014: 173): «identificados determinados rasgos subjetivos, se reconoce que tales rasgos están aceptados colectivamente y que presentan comportamiento regulares y motivados, esto es, que no constituyen producto del azar, de la imaginación o de la preferencia idiolectal del individuo». Así, nuestro interés en emplear como base la teoría de la percepción se halla en acudir a la información ofrecida por un conjunto de 73 informantes,

divididos en cuatro grupos diferentes (de entre 20 y 53 años): usuario medio de la lengua, traductor profesional, estudiante de lingüística y el investigador del objeto de estudio, en el papel de hablante-oyente ideal, desde el punto de vista generativo (Chomsky, 1965: 18, 24). El hecho de elegir cuatro grupos de informantes está relacionado con la intención de contar con percepciones heterogéneas y que los resultados no se vinculen a un único perfil de usuario (es decir, que no se falseen por infrarrepresentación). En este estudio, no discriminaremos entre la información ofrecida por los cuatro grupos de informantes (información que se trata en otro trabajo en preparación), sino que estos serán tratados como un único grupo.

Los grupos de informantes facilitan información tanto cuantitativa como cualitativa. En relación con la primera, determinan, basándose en el criterio de prominencia, el número de formas léxicas délficas, semicrípticas y crípticas que encuentran en el texto. En relación con la información cualitativa, aportan la información relacionada con el emisor, el destinatario y la clase que asignan al texto, como conclusión de su análisis. Debe resaltar que los grupos de informantes desconocen por completo la información que aporta el investigador del estudio, incluido el atributo del tipo de texto, que podría condicionar la decisión del informante. De esta forma, cada informante analiza un conjunto de textos sin formato y sin información adicional de carácter extralingüístico. El objetivo de proponer que los informantes determinen cuál de las tres clases asignan a cada texto, basándose en los elementos lingüísticos hallados en dicho texto, se debe a que nos interesa analizar qué elementos atribuyen a cada clase. De forma contraria, si la clase estuviese determinada *a priori* por el investigador, se estaría condicionando la clase a una serie de características que quizá pertenezcan a otra clase, por lo que la vinculación entre clasificación y elementos característicos sería errónea, además de nada descriptiva (mucho menos explicativa del fenómeno). El investigador, por su parte, proporciona la información relacionada con el tipo de texto, la variedad léxica, la densidad terminológica y sus tipos, el número de procedimientos de reformulación formal y sus tipos, y el índice de reformulación y sus tipos.

En definitiva, el interés de emplear la teoría de la percepción en este estudio reside en contar con las valoraciones de los usuarios sobre cómo vinculan la configuración discursiva de los textos a los diferentes grados de especialización de acuerdo con su experiencia lingüística individual, y, a partir del análisis global, identificar si existen convenciones que puedan caracterizar claramente el grado de especialización del discurso médico.

4.4. Algunos apuntes sobre aprendizaje automático aplicado a la clasificación textual

La última herramienta empleada para tratar de obtener información de los datos que nos permitan falsar nuestras hipótesis es el aprendizaje automático. Mediante al aprendizaje automático, los ordenadores son capaces de identificar patrones que muestran similitudes a partir de un conjunto de datos (Bishop, 2010). En el caso que nos ocupa, el aprendizaje automático nos permite:

- 1) determinar la probabilidad de saber la clase de un texto según su grado de especialización conociendo previamente los valores de las variables que hemos analizado (por ejemplo, ¿qué probabilidad existe de que un texto sea considerado semiespecializado si el texto cuenta con una superestructura de artículo de divulgación?);
- 2) identificar qué combinaciones de estas variables resultan en la clasificación de una determinada clase (por ejemplo, ¿qué clase corresponde a un texto si ha sido emitido por un especialista y cuenta con una densidad terminológica reducida?).

Hemos empleado el software WEKA de aprendizaje automático para analizar nuestra base de datos en Excel. En concreto, para estudiar la probabilidad, hemos empleado el algoritmo de red bayesiana (fundamentado en el teorema de Bayes de relación entre probabilidad condicionada y probabilidad total) y, para estudiar las condiciones de clasificación, hemos diseñado un árbol de decisión (un esquema arbóreo de suma de condiciones) mediante el algoritmo J48, con una configuración de árbol podado (es decir, un esquema simplificado), un intervalo de confianza de 0,045 y una validación cruzada de 20 iteraciones, que arrojan un 79,05 % de éxito en las clasificaciones.

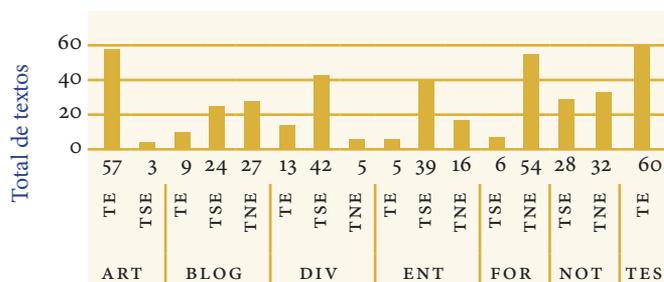
5. Resultados

A continuación, presentamos la información cuantitativa que relaciona las diferentes variables analizadas en esta contribución, vinculando nuestra variable dependiente (la clase de texto, que representa el grado de especialización textual) al resto de las variables.

5.1. Según la clase y los tipos textuales

El gráfico 1 identifica el número total de textos según la clase que han sido vinculados a cada tipo textual. De este gráfico llaman la atención los extremos: por una parte, para el artículo científico y la tesis doctoral eligen la clase de texto especializado casi exclusivamente; por otra, las participaciones en foros y las noticias no se clasifican nunca como textos especializados, sino, más bien, tienden al extremo de texto no especializado. Esto puede deberse a las funciones comunicativas que cumple cada superestructura: más representativa en el primer caso y más comunicativa en el segundo. Por otra parte, los otros tres tipos textuales se relacionan con la categoría de texto semiespecializado, tipo de clase a caballo

GRÁFICO 1. Clases de textos según tipo textual



entre los polos opuestos, si bien, incluso en esta categoría se observan ciertas tendencias: los artículos de divulgación se aproximan más al texto especializado, y los artículos de blogs, cuya moda corresponde al no especializado, tienden al semiespecializado. Determinaremos posteriormente algunas implicaciones adicionales junto con otras variables (gráfico 1).

Debe destacarse que los informantes desconocen el tipo de texto y, en algunos casos, no analizan el texto de forma íntegra, sino fragmentos de ciertas secciones seleccionados de forma aleatoria. Este hecho es relevante, puesto que la típica estructura formal de los artículos científicos médicos (introducción, materiales y métodos, resultados, discusión, conclusiones y referencias) quedaría oculta al informante, por lo que deben existir otros elementos de carácter lingüístico que le sirvan para atribuir la clase o las relaciones discursivas, por ejemplo (y algunos deben ser claros respecto a si el grueso de ese tipo textual es vinculado al texto especializado sin duda).

Si representamos en un eje horizontal las diferentes clases, entendidas como gradación, pero determinando las fronteras discretas entre las tres categorías, los datos obtenidos en la tabla anterior quedarían ilustrados como indica la tabla 3, lo que permite representar gráficamente lo cercanos o lejanos que son ciertos tipos textuales (y, por tanto, su configuración discursiva) respecto a determinadas clases del grado de especialización del discurso médico en español.

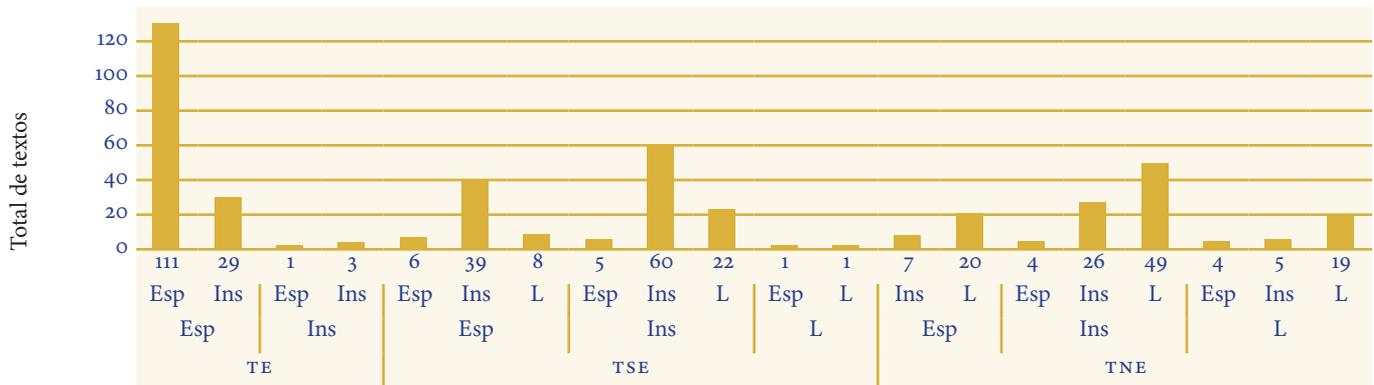
TABLA 3. Distribución de tipos textuales según clases

TE	TSE	TNE
ART		
TES		
	DIV	
	ENT	
		BLOG
		NOT
		FOR

5.2. Según las relaciones discursivas

En el gráfico 2 puede analizarse la relación que existe entre el emisor (variable inferior en el eje horizontal) y el receptor (variable superior en el eje horizontal) según la clase. Así, por ejemplo, destaca el hecho de que el texto especializado no se concibe en ningún momento como un texto dirigido a un lego, sino como fruto de una relación entre especialistas principal-

GRÁFICO 2. Relaciones discursivas según clase de texto



mente (o entre un emisor especialista y un receptor instruido). En este último caso, podría justificarse que el texto especializado se percibe como un texto en el que el emisor presupone en el receptor cierto dominio del conocimiento especializado (lo que justifica que ningún informante vincule esta clase a un receptor lego). Los textos especializados con una relación especialista-instruido podrían vincularse a manuales para especialistas en formación o a artículos dirigidos a estudiantes.

El texto semiespecializado es el más variado en relaciones discursivas, lo que es coherente con su estatus como objeto intermedio de gradación. En él, destaca, por su parte, el receptor instruido, combinado tanto con un emisor especialista como instruido. En la tabla 4 y la tabla 5 se muestran dos ejemplos¹ extraídos de los textos del corpus, donde se marcan algunos de los términos identificados por los informantes. Como se puede comprobar, los elementos lingüísticos no difieren en gran medida ni en calidad ni en cantidad, por lo que, a nuestro juicio, el hecho de que existan pares de interlocutores similares en cantidad en los informes puede deberse a la dificultad que entraña identificar el dominio del conocimiento especializado por parte del emisor.

TABLA 4. Texto semiespecializado con emisor y receptor instruidos

«Los **procedimientos mínimamente invasivos** son una de las respuestas más eficaces en el **diagnóstico y tratamiento de enfermedades**, con beneficios indiscutibles tanto para pacientes, profesionales y la institución: un **tratamiento neuro-cardiovascular** permite una recuperación más rápida para el paciente a la vez que reduce **complicaciones pos-operatorias**; gracias a las nuevas herramientas de **fusión de imagen**, el profesional toma **decisiones clínicas** más seguras que reducen **tiempos de intervención** y aportan seguridad al procedimiento; Con **cirugías mínimamente invasivas** la institución logra reducir estancias hospitalarias, optimizar recursos humanos y tecnológicos lo que se traduce en ahorro de costes» (Salud y Medicina, 2017) [MED DIV03].

TABLA 5. Texto semiespecializado con emisor especialista y receptor instruido

«La **fitoterapia** es la ciencia que estudia el **uso terapéutico** de las **plantas medicinales** y sus derivados para prevenir o tratar las **enfermedades**. Las **afecciones digestivas** se encuentran entre aquellas que pueden tratarse con plantas, y suponen aproximadamente el 20% de las **consultas de atención primaria**. Además, suelen ser **motivo de consulta** frecuente en las **oficinas de farmacia**. Debe tenerse en cuenta también que, aparte de su uso en **fitoterapia**, algunas de estas **plantas** pueden utilizarse también con finalidad alimentaria, como se establece en el **Real Decreto 3176/1983**» (Ediciones Mayo, 2015) [MED DIV02].

Por último, el texto no especializado destaca en la relación entre emisor instruido y receptor lego (en primer lugar; véase la tabla 6) y receptor instruido (en segundo lugar). En este caso, la única diferencia que se aprecia con los ejemplos anteriores se halla en la disminución considerable de términos semicríticos.

TABLA 6. Texto no especializado con emisor instruido y receptor lego

«El **aerosol** ofrecerá, además, autonomía al paciente de cara a la administración de un **ciclo convencional de fármacos**, explicó hoy la investigadora española en un comunicado. “Lo que tratamos es de abolir la dependencia del **paciente** que está dos horas sometido a **tratamiento** en una sala mientras le están suministrando la **quimioterapia**”, afirmó Martín del Valle en el comunicado. Asimismo se pretende disminuir la “cantidad de **fármaco** empleado para que llegue de forma específica”, lo cual reducirá la **toxicidad** y aumentará la **eficacia** del mismo, debido a que prácticamente “el 80% del **fármaco** suministrado no se utiliza, sino que tiene que ser **metabolizado** o expulsado por el **organismo**”» (EFE, 2017) [MED NOT06].

Por último, el caso de relación lego-lego, que también destaca en cantidad, es exclusivamente dominio de las participaciones en foros, como la que se relaciona en la tabla 7, donde se aprecia una densidad terminológica menor.

TABLA 7. *Texto no especializado con emisor y receptor legos*

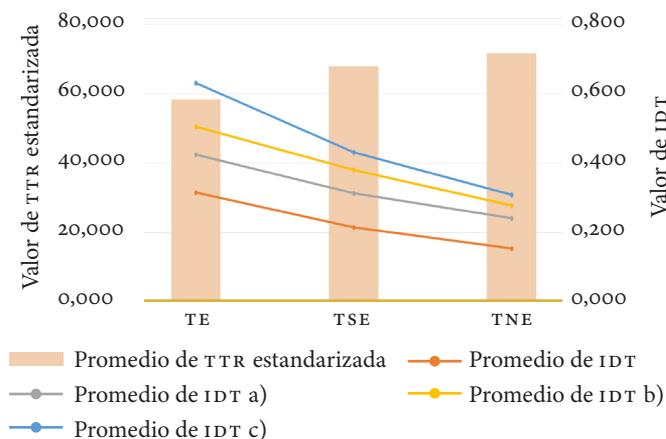
«Hola, en estos momentos, estoy sufriendo los dolores de una **neuralgia intercostal**. La estoy tratando con calor, sesiones de laser y aplicación de ciertas pomadas que dan calor y **anti-inflamatorios**. No se exactamente, a que se debe. Mi **fisio** me dijo que solían ser frecuentes después de una caída, pero no es mi caso. Hace unos diez años me hicieron una **punción lumbar** y la verdad que desde entonces la espalda me ha ido dando la lata, pero como esta vez nunca. El dolor es fuerte y dependiendo de la postura más. Mi trabajo es como secretaria y en muchas ocasiones, tengo que levantarme para cambiar de postura, ya que cuando estoy sentada y camino, es cuando más me molesta. Creo que hay que tener un poco de paciencia y dejar que los nervios se desinflanen para que dejen de rozarse y producir el dolor. El calor es lo más importante y lo que más ayuda» (AA. VV., 2006-2014) [MED FORO1].

Precisamente en estas relaciones se puede apreciar que los informantes pueden considerar las expectativas de conocimiento especializado por parte del receptor (Bernárdez, 1995: 142) como un criterio clave para establecer no solo el tipo de receptor, sino la clase a la que se asocia. Así, se puede señalar el rasgo de la aceptabilidad (Beaugrande y Dressler, 1997: 35-45) como uno de los rasgos empleados para vincular receptor y clase. Debe prestarse también especial atención al papel que ya se infiere de la terminología en los ejemplos anteriores, donde la densidad de colores (y de diferentes colores) va disminuyendo. A pesar de que estos ejemplos corresponden a textos particulares, en el corpus se aprecia un comportamiento similar cuando otros fragmentos conciernen a las relaciones discursivas planteadas anteriormente, lo cual podría servir a la redacción de textos para contar con modelos textuales que garanticen unas características lingüísticas que sirvan para la aceptabilidad de dicho texto. En otras palabras, la caracterización de textos como los anteriores permitiría a los creadores de contenido de blogs o revistas especializadas redactar sus textos atendiendo a las expectativas de conocimiento especializado de los diferentes receptores si ya conocen el tipo de terminología o de estructuras lingüísticas que le son propias a dicha relación discursiva.

5.3. Según la variedad léxica

En el caso de la variedad léxica, el Relación de variedad léxica y densidad terminológica según clase de texto permite identificar una relación clara: la variedad léxica de un texto es inversamente proporcional a su grado de especialización y también a su densidad terminológica. Los textos especializados constituyen la clase menos variada (como puede observarse en el ejem-

GRÁFICO 3. *Relación de variedad léxica y densidad terminológica según clase de texto*



plo que se recoge en la tabla 8), lo que es coherente con la sistematicidad y recurrencia típicas de estos textos. Por el contrario, los textos no especializados son más variados: presentan un promedio del 72 % de formas léxicas diferentes, lo que implica que 7 de cada 10 formas léxicas suelen ser distintas (gráfico 3).

Un análisis rápido de las siguientes tres tablas permite comprobar de forma visual cómo el grado de especialización está vinculado a una disminución de la variedad léxica de un texto (puesto que aparecen más palabras marcadas en negrita, que son las formas que se repiten). Una densidad menor en formas marcadas tipográficamente en negrita implica que existen formas léxicas muy diversas, por lo que el texto se considera poco repetitivo.

TABLA 8. *Texto especializado con una variedad léxica del 48 %*

«Con respecto a las características del **paciente** que se podrían asociar al grado de **ansiedad**, se observó un **mayor grado de ansiedad** en el sexo femenino de manera **estadísticamente significativa** (media 40,4 vs. 32, p = 0,01) y una correlación negativa con la **edad** (r: -0,170, p = 0,02). Se **observó un mayor grado de ansiedad** en los **pacientes** que nunca se habían **realizado** una **colonoscopia** pero sin **diferencias estadísticamente significativas** (media 38,7 vs. 32,3, p = 0,09). Se realizó un subanálisis en los **pacientes** que se habían **realizado** previamente una **colonoscopia** (n = 125). Se **observó** que existía un **mayor grado de ansiedad** en aquellos **pacientes** cuya tolerancia a la **colonoscopia previa** había sido valorada por el **paciente** como “mala” con **diferencias estadísticamente significativas** (media 50,67 vs. 28,44, p = 0,01) o si se correlacionaban con los valores de la EVA referidos por el **paciente** en la **colonoscopia previa** (r = 0,26, p = 0,005)» (Grilo *et al.*, 2016) [MED ART11].

TABLA 9. Texto semiespecializado con una variedad léxica del 67 %

«Siguiendo las recomendaciones de la Asociación Española de Pediatría, la **vacuna** de la **varicela** se ha incorporado recientemente a los calendarios vacunales de las diferentes comunidades autónomas, con una primera dosis a los 15 meses y otra entre los 3-4 años de edad. La **vacuna** de la **varicela** (Varivax®, Varilrix®) contiene virus vivo atenuado, por lo que cabe la posibilidad de que un niño correctamente **vacunado** pueda más adelante desarrollar un **herpes zóster**. Sin embargo, parece que la incidencia de **herpes zóster** en niños inmunocompetentes **vacunados** de la **varicela** es menor que en no **vacunados**» (Taberner, 2017) [MED BLOG13].

TABLA 10. Texto no especializado con una variedad léxica del 89 %

«En enero me percaté de unos **bultos** situados en la zona lumbar derecha (zona de la cadera, un poquitín más arriba de donde nos colocamos habitualmente el cinturón del pantalón). Son como 3 o 4 **bultos**, con **forma** ovoide, algo profundos, de consistencia algo dura o firme y que se mueven, no muy separados entre sí (alguno da la impresión de estar encima de otro), de un tamaño entre 3 a 5 centímetros. Uno de ellos es algo más redondo y tiene como si fuera tejido encima (como una línea que hace ‘clack’ cuando lo palpo). Estoy a la espera de hora con el cirujano, pero mientras tanto ando bastante preocupado por si puede ser algo malo. Simplemente quería preguntar si en esa **zona** hay o no ganglios, ya que en el fondo es lo que me preocupa. El doctor de cabecera considera que pueden ser lipomas (decir que desde octubre he bajado como 4 kilos de peso haciendo dieta, de **forma** intencionada)» (AA. VV., 2014) [MED FOR10].

Si entramos a valorar los tipos textuales menos variados, destacan el artículo científico y la tesis doctoral (coinciden con los textos especializados identificados en el gráfico 1), seguidos del artículo divulgativo, la entrevista y el blog (coinciden con el orden que aparece en la Tabla 3. Distribución de tipos textuales según clases). En concreto, el blog, la participación en foro y la noticia se diferencian en apenas un 3 % de variedad léxica, lo que sugiere que las diferencias no son nada destacables y que su configuración formal suele ser similar (es decir, plantean el uso del léxico de forma parecida). Algunos ejemplos que contrastan entre tipos los encontramos en los casos anteriores: la tabla 8 representa un artículo científico, y la tabla 10, una participación en foro.

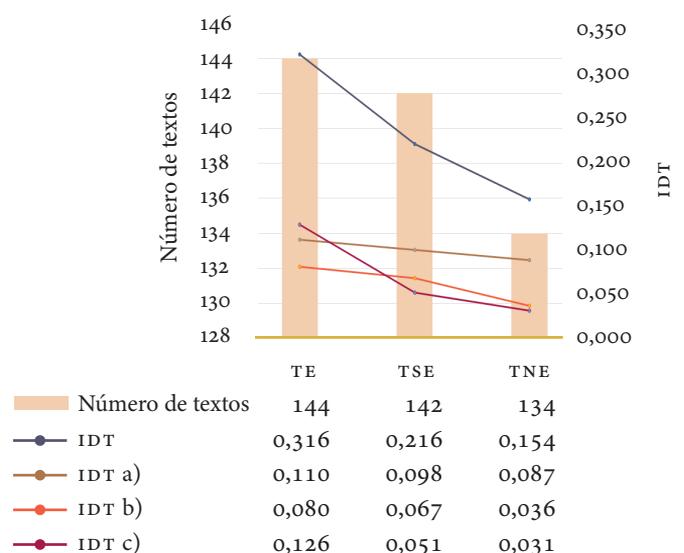
Conviene atender a la variedad léxica como instrumento en la redacción y traducción de textos especializados, especialmente si la relacionamos con la variación denominativa (con el uso de sinónimos, por ejemplo), puesto que los resultados reflejan, si se analizan de forma descriptiva, que los textos es-

pecializados son notablemente repetitivos, si bien con estos resultados desconocemos si esto se debe a una convención del español o a un factor universal. De hecho, Domènech (2006: 216) ya defendía en las conclusiones de su trabajo los beneficios de considerar en los estudios aplicados la variedad léxica como un elemento discriminante en el grado de especialización, idea con la que estamos de acuerdo, al ser fácilmente automatizable. En efecto, si se atiende previamente a factores pragmáticos como el posible receptor o el tipo textual, la relación *types-tokens* estandarizada podría servir para filtrar la especialización de los textos en un primer nivel.

5.4. Según la terminología

La terminología es reconocida como uno de los elementos característicos del texto especializado. En relación con el grado de especialización, el gráfico 4 muestra una relación directa entre el IDT de un texto y su clasificación dentro del grado de especialización: un texto se considera más especializado cuanto mayor sea la densidad terminológica que presenta (no cuanta más terminología, puesto que son valores diferentes). En esta tabla se observa que, con una descompensación de solo diez textos entre TE y TN (144 frente a 134), la diferencia en el IDT es de más del doble (0,316 frente a 0,154).

En relación con los tipos de densidad terminológica, se observa que solo el tipo (c), es decir, la densidad de términos crípticos (aquellos términos desconocidos por los informantes en su forma o contenido), es la única que muestra una tendencia lo suficientemente notable como para poder discriminar el grado de especialización, es decir, para distinguir entre TE, TSE y TNE. A continuación, se presentan dos ejemplos de grados diferentes con la terminología, marcada según se indica en la sección de metodología. En estos ejemplos se puede percibir, también de forma gráfica, cómo las diferencias de color ilustran la distinta densidad terminológica, y no solo eso, sino también cómo se percibe un diferente tipo de terminología: en

GRÁFICO 4. Relación de IDT según clases

la tabla 11, los informantes no distinguen fácilmente el significado de las formas léxicas (es decir, existe gran cantidad de términos crípticos, marcados en amarillo), mientras que en la tabla 12 identifican términos que reconocen como frecuentes en el discurso común (existe gran cantidad de términos délficos, marcados en azul).

TABLA 11. Artículo científico con alta densidad terminológica

«—OCMp tipo I: **Cálculos** pequeños que inician su desarrollo en la zona cóncava en íntimo contacto con el **tejido papilar**, donde se encuentra un **depósito subepitelial de hidroxapatita**, que permite el crecimiento de un **cálculo** que cuando finalmente es liberado, no incorpora la **placa de hidroxapatita** (figura 6).

»—OCMp tipo II: En este caso sí que podemos identificar un **core de hidroxapatita** que se encuentra cerca o en la zona cóncava del **cálculo**, donde observamos el **punto de anclaje**. Al igual que los **cálculos tipo I** se forman a partir de **depósitos de hidroxapatita** en la **papila renal** pero que en este caso, crecen lo suficiente como para erosionar el **epitelio** y contactar directamente con la orina, facilitando el **crecimiento cristalino** (figura 6).

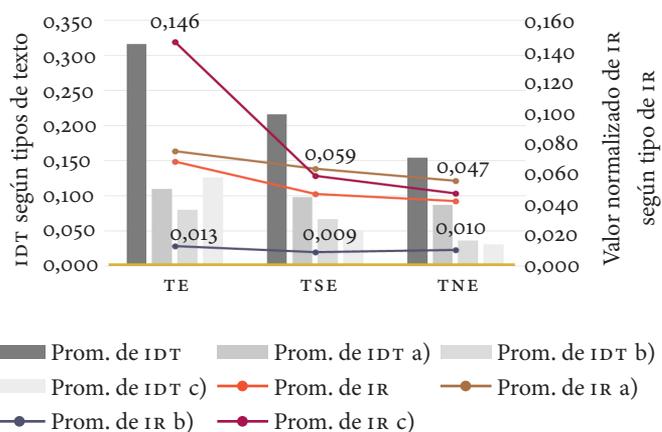
»—OCMp tipo III: Característicamente, estos **cálculos** se desarrollan en la punta de la **papila calcificada**, pudiéndose identificar en la zona cóncava en contacto con la **papila** y el **punto de anclaje, tejido calcificado con hidroxapatita** en el que incluso pueden observarse **túbulos calcificados** (figura 7)» (Ruiz, 2017) [MED TESIS10].

TABLA 12. Blog con baja densidad terminológica

«En distintos **metaanálisis** que extraen conclusiones tras analizar en conjunto las experiencias (algún trabajo valora en **tratamiento** de casi 1.000 **pacientes**) y estudios de diferentes grupos y países, se concluye que parecen dar mejores resultados los **tratamientos quirúrgicos** que los **conservadores** o no **quirúrgicos**. Es llamativo que las cifras se hayan dado la vuelta a favor de la **reparación quirúrgica** a mediados de los años noventa, que es cuando el **tratamiento quirúrgico** se empezaba a manejar con más recursos, precisión y mayor experiencia de los **cirujanos** que se habían aventurado unos decenios antes a iniciarse en la práctica de la **osteosíntesis**» (Villamor, 2015) [MED BLOG09].

Por otro lado, cualitativamente, la distinción de estos tres tipos de léxico es difícil de llevar a cabo de forma estadística, a no ser que se empleen las técnicas anteriormente mencionadas de análisis contrastivo de palabras clave (Tercedor y Láinez, 2017: 32), que necesitan de una selección previa de las unidades consideradas terminológicas. A primera vista, se aprecia la tendencia a la sintagmación de los términos marcados como crípticos (en amarillo) frente a unidades simples en los términos délficos (en

GRÁFICO 5. IR según clase de texto e IDT



azul). Además, cabe resaltar cómo los informantes distinguen tipos de términos según su percepción y son capaces de determinar el carácter terminológico de unidades del discurso común como *cirujano*, *pacientes* o *tratamientos quirúrgicos* (aunque desconocemos la motivación que les hace señalarlas como terminológicas) o incluso el carácter más críptico de *cálculos*, o totalmente opaco, como *depósito subepitelial de hidroxapatita*.

5.5. Según la reformulación

En este trabajo, el índice de reformulación es relativo al total de formas terminológicas, por lo tanto, mide el número de veces que la terminología se ve expresada por medios formales alternativos con el objetivo de facilitar el acceso del receptor al conocimiento especializado, como, por ejemplo, cuando se explica en qué consiste un término (expansión: IR a), se da nombre a un fenómeno específico (reducción: IR b) o se acompaña de un acrónimo (variación: IR c).

En el gráfico 5 se observa la relación directa entre IDT e IR: cuanto más elevada es la densidad terminológica de un texto, mayor es la reformulación a la que se ve sometido. No obstante, a pesar de esta relación directa, pueden defenderse las siguientes afirmaciones:

— La reformulación en general (IR) y la reformulación mediante procedimientos por variación (IR c) permiten identificar el texto especializado, pero no discriminar entre texto semiespecializado y no especializado. En otras palabras, un texto con numerosos acrónimos, variantes denominativas o sinónimos y explicaciones tendrá una probabilidad mayor de ser considerado especializado, y más aún si estas formas se repiten (lo cual encaja con la interpretación realizada sobre la escasa variedad léxica, lo que comentaremos posteriormente). Si no cuenta con dicha reformulación, deberemos emplear otro tipo de criterios para discriminar entre semiespecializado y no especializado, puesto que la reformulación resulta insuficiente.

— La reformulación por expansión (IR a) y la reformulación por reducción (IR b) no parecen permitir la discriminación entre textos, ya que varían de forma sutil entre textos

de diferente grado de especialización.

Debe recordarse que no existe relación entre la reformulación por reducción (IR b) y la densidad terminológica de unidades semicrípticas (IDT b), puesto que cada una de ellas mide fenómenos diferentes. En otras palabras, debe considerarse que el IR b) no mide la reformulación de las unidades semicrípticas (por ejemplo, *trastorno electrolítico*), sino la reformulación con procedimientos de reducción a la que se somete la terminología de un texto.

A continuación, se presentan tres ejemplos procedentes de tres clases diferentes en las que se observa tanto la densidad terminológica (cada color indica un tipo) como los procedimientos de reformulación que se emplean (subrayado en rojo para la expansión y subrayado en azul para la variación).

TABLA 13. Ejemplos de reformulación en texto especializado

«La **hiponatremia**, concentración de **sodio (Na) sérico** menor de 135 mmol/l, es un **trastorno electrolítico** frecuente que se asocia a una mayor **mortalidad** en la población general. Este peor pronóstico asociado a la **hiponatremia** es más acusado en ciertas **poblaciones**: mujeres, **pacientes posquirúrgicos**, **cirróticos** y **con insuficiencia cardiaca**, y se mantiene incluso en personas con **hiponatremia** considerada como moderada, 130-134 mEq/l². Ciertos **pacientes** con **enfermedad renal crónica, estadio 5D (ERC5D)**, en programa de **hemodiálisis (HD)**, presentan **hiponatremia**. El papel que juegan los riñones en la **prevención** de la **hiponatremia**, en la población general, en **pacientes** con **ERC5D**, tiende a desaparecer, en relación con la pérdida de la **función renal residual (FRR)**. En estos **enfermos**, el balance de **Na** y agua que se produce durante las sesiones de **diálisis** aparece como un nuevo factor en el desarrollo y mantenimiento de la **hiponatremia**» (Pérez-García *et al.*, 2016) [MED ART07].

TABLA 14. Ejemplos de reformulación en texto semiespecializado

«En este sentido, debemos poner el foco en el **paciente frágil**, aquel especialmente vulnerable a presentar **deterioro funcional** y **eventos adversos**, y no tanto en el **manejo** de **patologías concretas** de forma individual. Evidentemente el **manejo** de **patologías crónicas** que afectan a la **función** (fundamentalmente **osteomusculares**, **neurodegenerativas** y **cardiovasculares**) resulta muy importante para evitar el **deterioro funcional**. Además en el anciano siempre debemos estar atentos a los denominados **síndromes geriátricos** (**caídas**, **delirium** y **deterioro cognitivo**, **disfagia**, etc.) que condicionan de forma significativa la calidad de vida de nuestros **pacientes**» (Simón, 2017) [MED ENT06].

TABLA 15. Ejemplos de reformulación en texto no especializado

«Hace 20 años sufrí un **trauma craneoencefálico severo**, estuve en estado de coma durante 4 días, presente **pérdida temporal de memoria** y en el **periodo** de **convalecencia** abusé de **estimulantes** (**modafinil**, **caféina**, coca-cola, **aspirinas**) para mantenerme despierta en la noche, toda vez que si llegaba a dormir olvidaría lo que había estudiado. [...]

»El **cansancio** o **agotamiento** es un **síntoma** frecuente que toda persona tiene en algunas ocasiones ya sea por **sueño**, **por trabajo intenso**, **un viaje agotador** o **por la práctica excesiva de deporte**. [...]

»El **cansancio** es la sensación de falta de energía que disminuye la capacidad de trabajo y dificulta a los músculos realizar sus funciones con la facilidad habitual; es normal que este **síntoma** se presente después de largas jornadas de actividad física o mental, al no dormir lo suficiente, al padecer **anemia** o **depresión** o no alimentarse adecuadamente, entre otros factores» (A.A. V.V., 2014-2015) [MED FOR08].

Los procedimientos por variación en el texto especializado presentan, principalmente, formas acronímicas, por lo que no contribuyen a aumentar la variedad léxica de un texto: una vez presentado el acrónimo, es empleado de forma sistemática a lo largo del texto. Esto contrasta con el comportamiento de la variación en el texto no especializado (por ejemplo, el uso de «o agotamiento» en la Tabla 15. Ejemplos de reformulación en texto no especializado), que representa una variación denominativa (recuérdese que se ha vinculado el texto no especializado a la elevada variedad léxica). Por otra parte, la expansión emplea estrategias tanto de definición («es un trastorno electrolítico»), de explicación («aquel especialmente vulnerable a») o de ejemplificación («modafinil, caféina»).

5.6. Resumen de los valores más representativos del corpus

La tabla 16 presenta los datos más representativos del corpus para cada tipo de texto. En relación con los 420 análisis de los 105 textos, a continuación, se presentan los valores más frecuentes o moda (en el caso de las variables cualitativas, a la derecha) y los promedios (en el caso de las variables cuantitativas, a la izquierda). En el caso de las cuantitativas, se han marcado en color verde aquellos valores superiores, y en rojo, los inferiores.

Si interpretamos esta tabla atendiendo a las potenciales aplicaciones de los planteamientos teóricos que de aquí se derivan, se podría defender, por ejemplo, que:

— al redactar o traducir artículos científicos o tesis doctorales sobre medicina en español, la norma es la repetición del léxico, como avala su reducido TTR;

— existe una relación entre la clase de un texto y su superestructura, y esta se relaciona estrechamente con el receptor del texto, dado el objetivo del emisor de cubrir las expectativas de conocimiento especializado del receptor —disponer

TABLA 16. Resumen global de valores más representativos para cada variable según tipo de texto

Tema y tipo texto	TTR	IDT				IR				Emisor	Receptor	Clase
		Global	a	b	c	Global	a	b	c			
ART	54,421	0,266	0,121	0,067	0,079	0,076	0,094	0,019	0,110	Esp	Esp	TE
TES	56,649	0,373	0,115	0,093	0,164	0,079	0,086	0,010	0,205	Esp	Esp	TE
DIV	65,665	0,249	0,100	0,076	0,073	0,032	0,038	0,008	0,051	Esp / Ins	Ins	TSE
ENT	66,432	0,160	0,105	0,038	0,017	0,062	0,075	0,014	0,099	Ins	Ins	TSE
BLOG	71,555	0,236	0,087	0,063	0,085	0,032	0,046	0,007	0,029	Esp	L	TNE
NOT	74,037	0,190	0,096	0,061	0,033	0,032	0,034	0,012	0,047	Ins	Ins / L	TNE
FOR	72,094	0,140	0,066	0,033	0,041	0,053	0,082	0,006	0,053	Ins / L	L	TNE
MEDIA	65,836	0,230	0,099	0,062	0,070	0,053	0,065	0,011	0,085			

de esta relación permitiría a la lingüística de corpus o a la terminografía contar con criterios teóricos sólidos a la hora de documentar ciertas formas terminológicas en textos reales—;

—la reformulación por variación desempeña un papel relevante para distinguir el texto especializado, lo que entra en conflicto con la postura de escasa reformulación en el texto especializado que propone Ciapuscio (2003: 90), lo que sugiere la necesidad de llevar a cabo investigaciones futuras sobre la relevancia de la reformulación.

5.7. Resultados según aprendizaje automático

5.7.1. Probabilidad de clasificación según variables

El teorema de Bayes permite identificar probabilidades de clasificación. En la tabla 17 se presentan dos probabilidades diferentes, que se ilustran mediante ejemplos: (A) determina la probabilidad de que un texto corresponda a un artículo científico si sabemos que es especializado (la probabilidad se lee de arriba abajo, es decir, por columnas); y (B) determina la proba-

bilidad de que un texto sea especializado si sabemos que es un artículo científico (la probabilidad se lee de izquierda a derecha, es decir, por filas).

Las probabilidades que se relacionan aquí permiten profundizar más en el valor condicionante de las variables y valorar con pruebas objetivas aquellos casos poco claros (como el papel del instruido como emisor, el tipo textual de blog o de noticia, o la reformulación reducida) frente a casos con probabilidades muy polarizadas, casi binarias (como el tipo textual de artículo o tesis, el lego como emisor, el especialista como receptor, o un IDT c) o IR c) elevados, que casi garantizan un prototipo exclusivo de clase). Como marco de partida, esta sección presenta las ventajas de proyección teórica al aplicar este tipo de herramientas en el análisis de grandes masas de datos, como en corpus de varios millones de palabras o cientos de textos del discurso médico.

5.7.2. Árbol de decisión

En último lugar, el algoritmo J48 permite reunir las variables y conocer qué combinación de sus valores se asigna a cada clase.

FIGURA 1. Árbol de decisión

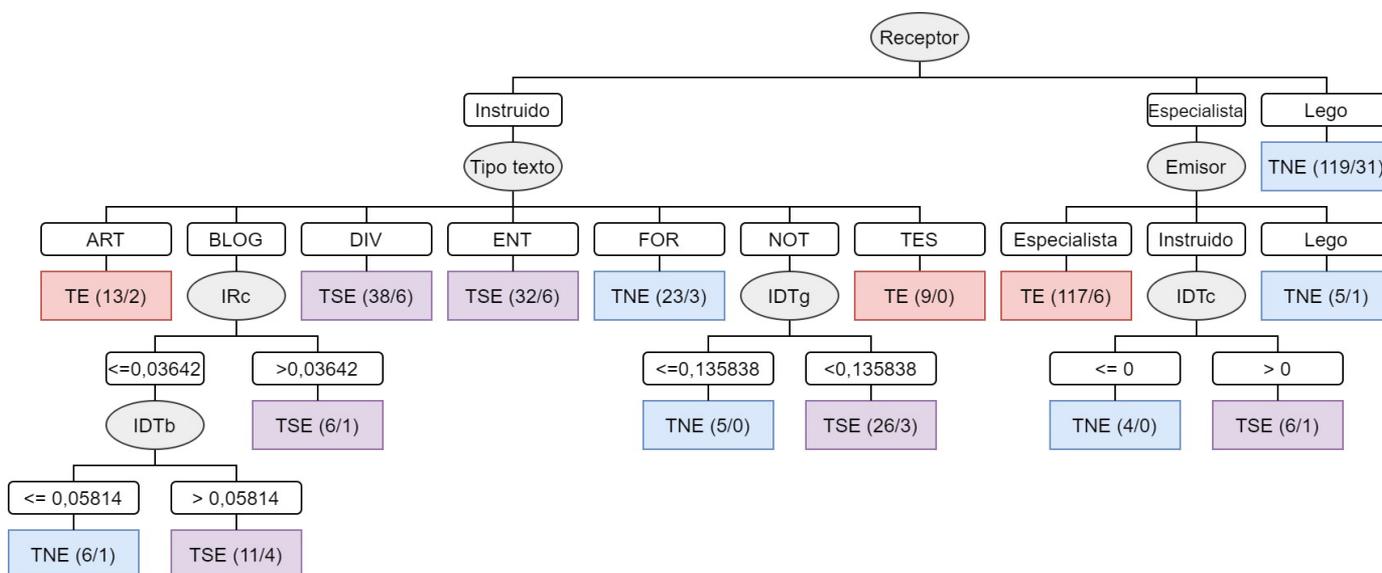


TABLA 17. Probabilidades de clasificación según el grado de especialización y las diferentes variables

Valor del atributo		(A) Probabilidad del valor de la variable conociendo el valor de la clase			(B) Probabilidad del valor de la clase conociendo el valor de la variable		
		TE	TSE	TNE	TE	TSE	TNE
Tipo de texto	ART	39,58 %	2,11 %	0,00 %	95,00 %	5,00 %	0,00 %
	BLOG	6,25	16,90	20,15	15,00	40,00	45,00
	DIV	9,03	29,58	3,73	21,67	70,00	8,33
	ENT	3,47	27,46	11,94	8,33	65,00	26,67
	FOR	0,00	4,23	40,30	0,00	10,00	90,00
	NOT	0,00	19,72	23,88	0,00	46,67	53,33
	TES	41,67	0,00	0,00	100,00	0,00	0,00
Emisor	Especialista	97,22	37,32	20,15	63,64	24,09	12,27
	Instruido	2,78	61,27	58,96	2,35	51,18	46,47
	Lego	0,00	1,41	20,90	0,00	6,67	93,33
Receptor	Especialista	77,78	8,45	5,97	84,85	9,09	6,06
	Instruido	22,22	69,72	28,36	18,93	58,58	22,49
	Lego	0,00	21,83	65,67	0,00	26,05	73,95
IDT global	< 0,1025	0,69	8,45	35,82	1,64	19,67	78,69
	0,1025-0,1945	12,50	40,85	35,07	14,63	47,15	38,21
	0,1945-0,3935	59,72	45,77	29,10	45,26	34,21	20,53
	> 0,3935	27,08	4,93	0,00	84,78	15,22	0,00
IDT b)	< 0,0075	0,69	6,34	13,43	3,57	32,14	64,29
	0,0075-0,0525	38,89	36,62	66,42	28,43	26,40	45,18
	0,0525-0,1405	42,36	54,23	17,16	37,89	47,83	14,29
	> 0,1405	18,06	2,82	2,99	76,47	11,76	11,76
IDT c)	< 0,0015	0,00	4,93	13,43	0,00	28,00	72,00
	0,0015-0,0585	26,39	59,86	69,40	17,59	39,35	43,06
	0,0585-0,151	33,33	32,39	17,16	41,03	39,32	19,66
	> 0,151	40,28	2,82	0,00	93,55	6,45	0,00
IR global	< 0,060801	54,17	73,24	77,61	27,27	36,36	36,36
	> 0,060801	45,83	26,76	22,39	49,25	28,36	22,39
IR a)	< 0,075332	62,50	75,35	79,10	29,70	35,31	34,98
	> 0,075332	37,50	24,65	20,90	46,15	29,91	23,93
IR b)	< 0,001595	39,58	47,89	58,96	27,94	33,33	38,73
	> 0,001595	60,42	52,11	41,04	40,28	34,26	25,46
IR c)	< 0,009085	20,83	44,37	61,94	17,05	35,80	47,16
	0,009085-0,022595	4,86	0,00	0,00	100,00	0,00	0,00
	0,022595-0,220056	51,39	52,11	35,82	37,76	37,76	24,49
	> 0,220056	22,92	3,52	2,24	80,49	12,20	7,32

Así, en la figura 1, aparece un árbol de decisión² cuya primera opción podría interpretarse como sigue: si un texto cuenta con un receptor especialista y su emisor es instruido, la clase dependerá de la densidad terminológica de términos crípticos (IDT c). Si es igual que cero, será considerado texto no especializado y, si es mayor, será considerado texto semiespecializado.

De este árbol se coligen dos ideas:

— El receptor se revela como el punto de partida del grado (la raíz del árbol), es decir, como condicionante de la combinación con el resto de las variables y como condicionante del grado de especialización, en definitiva (algo que

ya defendía Domènech desde el punto de vista teórico en su trabajo de 2006, p. 72). Desde el punto de vista estadístico, la percepción de los informantes, considerada de forma colectiva, valida esta relación condicionante del receptor sobre el grado de especialización.

— La densidad terminológica y sus tipos son discriminantes del grado de especialización, pues siempre sirven en última instancia para diferenciar entre clases. Esto no sucede con la reformulación, cuyo único valor discriminante es la reformulación mediante procedimientos de variación (IR c) (véase el blog en la figura 1), como ya se avanzó en el gráfico 5.

6. Conclusiones, aplicaciones y futuras líneas de trabajo

Para finalizar, puede afirmarse que los datos aportados permiten validar nuestras hipótesis:

— El receptor constituye el elemento que condiciona las características lingüísticas de los diferentes textos en el marco del grado de especialización, puesto que es el que fija la superestructura y la configuración léxica del texto. En definitiva, se comprueba la importancia del factor pragmático, no solo en la distinción del valor terminológico, sino en la propia configuración de la noción de grado de especialización.

— La densidad terminológica, en especial la densidad terminológica de términos crípticos (IDT c), es el principal factor discriminante del grado de especialización textual y la que permite distinguir entre textos especializados, semiespecializados y no especializados, si bien no es el único elemento que contribuye a su distinción. Esta comprobación ha sido posible a través de tres vías: el análisis de la relación IDT-clase, la determinación de probabilidades y el árbol de decisión.

— La relación entre emisor y receptor desde el punto de vista de su competencia cognitivo-comunicativa está vinculada a los tipos textuales seleccionados por el emisor para transmitir la información al receptor, lo que apoya las palabras de Cabré y Estopà (2005: 36), al afirmar que el texto especializado se caracteriza por seleccionar cierto tipo de formas discursivas especiales.

— La reformulación terminológica, al contrario que la idea tradicional, está asociada de forma directa al grado de especialización textual, de forma que cuanto mayor es el grado de especialización, mayor es el grado de reformulación al que se somete la terminología. En este caso, cabe detenerse y destacar el hecho de que la reformulación en general (IR) y la reformulación mediante procedimientos por variación (IR c) permiten identificar el texto especializado, pero no discriminar entre texto semiespecializado y no especializado, por lo que los investigadores deben emplear otro tipo de criterios para distinguir entre estos últimos, ya que la reformulación resulta insuficiente.

Este estudio presenta aplicaciones potenciales muy relacionadas con la redacción y traducción de originales, puesto que la identificación del grado de especialización permite prever las dificultades en la comprensión del usuario receptor por la falta de información conocida, de forma que los lingüistas pueden poner en marcha las estrategias discursivas y denominativas necesarias para que el acto comunicativo tenga éxito. Por ejemplo, el papel de la reformulación es especialmente relevante en aplicaciones como la redacción o la traducción de textos especializados, donde los acrónimos o sinónimos son un objeto de análisis esencial, al ser objetos de consulta documental, y que tienen repercusión en la creación de obras terminográficas (que satisfagan las necesidades en traducción). Como se puede apreciar, se trata de una relación donde se involucran simultáneamente varios campos. De esta forma, nos

aproximamos a explicar el texto como objeto dinámico (Bernárdez, 1995: 137). En el caso de la medicina, resultaría de gran utilidad aplicado a los estudios de comunicación entre especialista y paciente.

A partir de este estudio, se puede profundizar en el diseño de modelos de redacción basándonos en patrones caracterizados desde el punto de vista pragmático, en primer lugar, y lingüístico (específicamente léxico) en segundo lugar. De ahí que uno de los elementos por explorar deba ser la sintaxis. Otra de las aplicaciones de esta contribución tiene que ver con la lingüística contrastiva, puesto que desconocemos si el grado de especialización se comporta en el nivel léxico de la misma forma en todas las lenguas (¿presentan el alemán, el español, el francés y el inglés el mismo tipo y densidad de terminología en los tres grados?), lo cual es útil para la elaboración de corpus comparables. Esta precisión guarda relación con uno de nuestros objetivos, pues consistía en considerar la teoría de la percepción como una herramienta de análisis que permita identificar convenciones colectivas a partir de análisis subjetivos: a pesar de identificar estas convenciones, mediante promedios y modas, por ejemplo, nuestros resultados están limitados al análisis en español.

Por último, entre las futuras líneas que nos proponemos trabajar en relación con el grado de especialización, se encuentran dos principalmente:

— El estudio de la sintaxis, tanto interoracional como intraoracional, así como de la morfología (especialmente la oposición de categorías léxicas y gramaticales). A nuestro juicio, la densidad terminológica, aunque útil en la discriminación del grado, implica problemas metodológicos importantes desde la raíz, puesto que conlleva dificultades en la determinación del valor terminológico que, en nuestro caso, hemos tenido que determinar mediante percepción y no mediante otras técnicas (véase L'Homme, 2020). La comparación de la configuración sintáctica de los textos podría permitir una clasificación más ágil, especialmente si se descubren patrones identificables mediante inteligencia artificial que sean luego traspasables a memorias de traducción, traducción automática, asistencia automática para la redacción o corrección de estilo, etc.

— El estudio de los datos aportados por los informantes de forma desglosada, de manera que puedan analizarse las similitudes y divergencias en la percepción del léxico.

Notas

1. Los ejemplos que se relacionan a lo largo del trabajo son citas literales de los textos del corpus. Van acompañados del código que los identifica, junto con su fuente. Las citas incluyen todos los errores conceptuales y ortotipográficos de los originales. No han sido marcados con (*sic*) para evitar entorpecer la lectura de los ejemplos.
2. Los números que se relacionan en el árbol de decisión son los resultados exitosos (a la izquierda de la barra) y erróneos (a la derecha de la barra) en la clasificación del algoritmo.



mo. Al tratarse de un árbol podado, la sobresimplificación del esquema implica ciertos errores de clasificación.

Referencias bibliográficas

Adelstein, Andreína (2007): *Unidad léxica y significado especializado: modelo de representación a partir del nombre relacional madre* (Tesis doctoral). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Arntz, Reiner y Heribert Picht (1995): *Introducción a la terminología*. Madrid: Fundación Germán Sánchez Ruipérez.

Beaugrande, Robert-Alain y Wolfgang Dressler (1997): *Introducción a la lingüística del texto*. Barcelona: Ariel.

Bernárdez, Enrique (1982): *Introducción a la lingüística del texto*. Madrid: Espasa-Calpe.

Bernárdez, Enrique (1995): *Teoría y epistemología del texto*. Madrid: Cátedra.

Berns, Margie y Paul Kei Matsuda (2006): «Applied Linguistics: Overview and History», en Keith Brown (ed. in chief): *Encyclopedia of Language and Linguistics*. Amsterdam: Elsevier, pp. 394-405. <<http://dx.doi.org/10.1016/B0-08-044854-2/00599-X>> [consulta: 25.I.2020].

Bishop, Christopher M. (2010): *Pattern recognition and machine learning*. New York: Springer.

Cabré, María Teresa (1998): «Elementos para una teoría de la terminología: hacia un paradigma alternativo», *El Lenguaje*.

- raz. *Revista Académica del Colegio de Traductores Públicos de la Ciudad de Buenos Aires*, 1 (1): 59-78.
- Cabré, María Teresa (1999): *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Cabré, María Teresa (2002^a): «Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción», en Amparo Alcina y Silvia Gamero (eds.): *La traducción científico-técnica y la terminología en la sociedad de la información*. Castelló de la Plana: Publicacions de la Universitat Jaume I, pp. 87-105.
- Cabré, María Teresa (2002^b): «Textos especializados y unidades de conocimiento: metodología y tipologización», en Joaquín García y María Teresa Fuentes (eds.): *Texto, terminología y traducción*. Salamanca: Almar, pp. 15-36.
- Cabré, María Teresa (2007): «Constituir un corpus de textos de especialización: condiciones y posibilidades», en Michel Ballard y Carmen Pineira-Tresmontant (eds.): *Les corpus en linguistique et en traductologie*. Arras: Artois Presses Université, pp. 89-106.
- Cabré, María Teresa (2008): «El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en terminología», *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, 16: 9-36. <http://www.aelfe.org/documents/03_16_Cabre.pdf> [consulta: 25.1.2020].
- Cabré, María Teresa y Rosa Estopà (2005): «Unidades de conocimiento especializado: caracterización y tipología», en María Teresa Cabré y Carme Bach (eds.): *Coneixement, llenguatge i discurs especialitzat*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 69-93.
- Cabré, María Teresa et al. (2011): «Automatic specialized vs. non-specialized texts differentiation: a first approach», en Noa Talaván, Elena Martín Monje y Francisco Palazón (eds.): *Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10*. Madrid: Universidad Nacional de Educación a Distancia (UNED), pp. 301-310.
- Cabré, María Teresa et al. (2014): «Automatic specialized vs. non-specialized text differentiation: the usability of grammatical features in a Latin multilingual context», en Elena Bárcena, Timothy Read y Jorge Arús (eds.): *Languages for Specific Purposes in the Digital Era*. Berlin: Springer, pp. 223-241.
- Caravedo, Rocío (2014): *Percepción y variación lingüística: enfoque sociocognitivo*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- Chomsky, Noam (1965): *Aspects of a Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Ciapuscio, Guiomar Elena (2003): *Textos especializados y terminología*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Ciapuscio, Guiomar Elena e Inés Kuguel (2002): «Hacia una tipología del discurso especializado: aspectos teóricos y aplicados», en Joaquín García y María Teresa Fuentes (eds.): *Texto, terminología y traducción*. Salamanca: Almar, pp. 37-73.
- Ciapuscio, Guiomar Elena; Inés Kuguel e Isabel Otañi (2005): «El conocimiento especializado: el texto de especialización y los criterios para su tipologización», en María Teresa Cabré y Carme Bach (eds.): *Coneixement, llenguatge i discurs especialitzat*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 95-110.
- Da Cunha, Iría et al. (2011): «Automatic Specialized vs. Non-Specialized Sentence Differentiation», en Alexander Gelbukh (ed.): *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science (LNCS)*. Berlin: Springer, pp. 266-276.
- Díaz Rojo, José Antonio (2000): «El modelo ecolingüístico de comunicación especializada: investigación y divulgación», *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, 2: 69-84. <<http://www.aelfe.org/documents/text2-Diaz.pdf>> [consulta: 25.1.2020].
- Dijk, Teun A. van (1992 [1983]): *La ciencia del texto*. Barcelona: Paidós.
- Domènech, Ona (2006): *Textos especialitzats i variació vertical: la diversitat terminològica com a factor discriminant del nivell d'especialització d'un text* (Tesis doctoral). Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- Domènech, Ona (2007): «La noció de text especialitzat des de la perspectiva de la teoria comunicativa de la terminologia», en Mercè Lorente et al. (eds.): *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví*. Vol. 2. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 241-254.
- Garcés Gómez, María del Pilar (2008): *La organización del discurso: marcadores de ordenación y de reformulación*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- Halliday, Michael Alexander Kirkwood (1982 [1979]): *El lenguaje como semiótica social: la interpretación social del lenguaje y del significado*. México: Fondo de Cultura Económica.
- Hoffmann, Lothar (1998 [1987]): *Llenguatges d'especialitat: selecció de textos* (Brumme, Jenny, ed.). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- L'Homme, Marie-Claude (2020): *Lexical semantics for terminology: an introduction*. Amsterdam: John Benjamins.
- Loureda Lamas, Óscar (2002): «Los tipos de texto», en María Victoria Romero Gualda (coord.): *Lengua española y comunicación*. Barcelona: Ariel, pp. 155-176.
- McEnery, Tony y Andrew Hardie (2012). *Corpus Linguistics: method, theory and practice*. Cambridge: Cambridge University.
- Monterde Rey, Ana María (2002): *Ejercicios de introducción a la terminología*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- Montero, Silvia y Pamela Faber (2008): *Terminología para traductores e intérpretes*. Granada: Tragacanto.
- Nazar, Rogelio y María Teresa Cabré (2011): «Un experimento

- to de extracción de terminología utilizando algoritmos estadísticos supervisados», *Debate Terminológico*, 7: 36-55. <<http://seer.ufrgs.br/riterm/article/download/20712/11980>> [consulta: 20.IV.2020].
- Payrató, Lluís (1998): *De profesión, lingüista. Panorama de la lingüística aplicada*. Barcelona: Ariel.
- Prieto Velasco, Juan Antonio (2008): *Información gráfica y grados de especialización en el discurso científico: un estudio de corpus* (Tesis doctoral). Granada: Universidad de Granada.
- Rodríguez Tapia, Sergio (2019): *El grado de especialización textual: caracterización a partir de la percepción sociocognitiva de la terminología y las relaciones discursivas* (Tesis doctoral). Córdoba: Universidad de Córdoba.
- Temmerman, Rita (2002): «Sociocognitive Terminology Theory», en Pamela Faber y Catalina Jiménez (eds.): *Investigar en terminología*. Granada: Método, pp. 75-92.
- Templin, Mildred C. (1957): *Certain language skills in children: Their development and interrelationships*. Westport: Greenwood.
- Tercedor Sánchez, Maribel y Antonio Jesús Láinez Ramos-Bossini (2017): «La comunicación en línea en el ámbito de la psiquiatría: caracterización terminológica de los foros de pacientes», *Panace@*, 18 (45): 30-41. <https://www.tremedica.org/wp-content/uploads/n45_tribuna-MTece-dorAJLainez.pdf> [consulta: 20.IV.2020].
- Torruella Casañas, Joan (2017): *Lingüística de corpus: génesis y bases metodológicas de los corpus (históricos) para la investigación en lingüística*. Frankfurt am Main: Peter Lang.
- Zamorano Aguilar, Alfonso (2012): «Teorías del caos y lingüística: aproximación caológica a la comunicación verbal humana», *Signa*, 21: 679-705. <<https://doi.org/10.5944/signa.vol21.2012.6323>> [consulta: 25.I.2020].
- Referencias del corpus**
- AA. VV. (15 de octubre de 2014 – 1 de octubre de 2015): «cansancio», *Onmeda*. <<http://www.onmeda.es/foros/medicina-general/16219-cansancio>> [consulta: 25.I.2020].
- AA. VV. (25 de marzo de 2014 – 1 de abril de 2014): «¿Ganglios parte baja de la espalda?», *Onmeda*. <<http://www.onmeda.es/foros/medicina-general/4025-%C2%BFganglios-parte-baja-de-la-espalda>> [consulta: 25.I.2020].
- AA. VV. (6 de noviembre de 2006 – 10 de septiembre de 2014): «neuralgia intercostal idiopática», *PortalesMédicos.com*. <http://www.portalesmedicos.com/foros_medicina_salud_enfermeria/ubthreads.php/posts/15966> [consulta: 25.I.2020].
- Ediciones Mayo (13 de julio de 2015): «Plantas digestivas para uso terapéutico», *elfarmacéutico.es*. <<http://www.elfarmacéutico.es/index.php/fitoterapia/item/6156-plantas-digestivas-para-uso-terapeutico#.WR1GaNxFfc>> [consulta: 25.I.2020].
- EFE (22 de marzo de 2017): «Prueban quimioterapias no tóxicas para abatir el cáncer», *El Tiempo*. <<http://www.eltiempo.com/vida/salud/quimioterapia-no-toxica-para-combatir-cancer-70518>> [consulta: 25.I.2020].
- Grilo Bensusan, Israel *et al.* (2016): «Estudio prospectivo de la ansiedad del paciente previa a la realización de una colonoscopia ambulatoria», *Revista Española de Enfermedades Digestivas*, 108 (12). <<http://dx.doi.org/10.17235/reed.2016.4104/2015>> [consulta: 25.I.2020].
- Güell, Oriol (26 de enero de 2020): «Lo que se sabe del nuevo virus de China que ha puesto en alerta al mundo», *El País*. <https://elpais.com/sociedad/2020/01/23/actualidad/1579777128_930268.html> [consulta: 25.I.2020].
- Junta Editorial de Cancer.Net (julio de 2016): «Tratar el dolor con medicamentos», *Cancer.Net*. <<https://www.cancer.net/es/asimilaci%C3%B3n-con-c%C3%A1ncer/efectos-f%C3%ADsicos-emocionales-y-sociales-del-c%C3%A1ncer/manejo-de-los-efectos-secundarios-f%C3%ADsicos/dolor/tratar-el-dolor-con-medicamentos>> [consulta: 25.I.2020].
- Paricio, Jorge (10 de mayo de 2017): «En España, uno de cada diez ciudadanos de más de 40 años padece EPOC, pero el 73 % de las personas que la sufren no lo sabe», *Salud y Medicina*. <<http://www.saludymedicina.org/post/en-espana-uno-de-cada-diez-espanoles-de-mas-de-40-anos-padece-epoc-pero-el-73-de-las-personas-que-la-sufre-no-lo-sabe>> [consulta: 25.I.2020].
- Pérez-García, Rafael *et al.* (2016): «Hiponatremia, mortalidad y hemodiálisis: una asociación no explicada», *Nefrología*, 36 (1). <<http://dx.doi.org/10.1016/j.nefro.2015.10.005>> [consulta: 25.I.2020].
- Ruiz Plazas, Xavier (2017): *Estudio sobre los factores implicados en la litiasis renal bilateral* (Tesis doctoral). Illes Balears: Universitat de les Illes Balears. <<http://www.tdx.cat/bitstream/handle/10803/402784/txrpx1de1.pdf?sequence=1&isAllowed=y>> [consulta: 25.I.2020].
- Salud y Medicina* (12 de mayo de 2017): «Azurion, último avance en terapias guiadas por la imagen», *Salud y Medicina*. <<http://www.saludymedicina.org/post/azurion-ultimo-avance-en-terapiasguiadas-por-la-imagen>> [consulta: 25.I.2020].
- Simón Vázquez, Clara (3 de abril de 2017): «El envejecimiento saludable tiene mucho que ver con el concepto de envejecimiento activo», *El Médico Interactivo*. <<http://www.elmedicointeractivo.com/articulo/entrevistas/envejecimiento-saludable-tiene-mucho-ver-concepto-envejecimiento-activo/20170403124909111162.html>> [consulta: 25.I.2020].
- Taberner, Rosa (19 de abril de 2017): «Herpes zóster: también en niños», *Dermaixel*. <<http://www.dermapixel.com/search?updated-max=2017-04-22T20:00:00%2B02:00&max-results=7>> [consulta: 25.I.2020].
- Villamor, Ángel (19 de octubre de 2015): «Reparación quirúrgica de la fractura de clavícula; segura y justificada», *El Blog del Doctor Villamor*. <<https://www.angelvillamor.com/2015/10/reparacion-quirurgica-de-la-fractura-de-la-html>> [consulta: 25.I.2020].