

Percepción, terminología e interlocutores del discurso jurídico en el grado de especialización textual

Sergio Rodríguez-Tapia¹

Recibido: 21 de julio de 2020 / Aceptado: 4 de agosto de 2022

Resumen. Este trabajo analiza el grado de especialización textual del discurso jurídico. Para ello, a partir de los fundamentos discursivistas de la Teoría Comunicativa de la Terminología y de la teoría sociocognitiva de la percepción, y basándonos en la lingüística de corpus, se ha analizado una base de datos 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. Los resultados permiten dejar constancia de la complejidad del discurso jurídico, dada la dificultad de encontrar patrones claros de comportamiento en diversas variables, especialmente si se comparan con otras disciplinas como la medicina.

Palabras clave: Corpus; terminología; grado de especialización; percepción; reformulación, texto especializado.

[en] Perception, terminology and speakers of legal discourse in the text specialisation degree

Abstract. This paper analyses the text specialisation degree of legal discourse. To this end, based on the discursive foundations of the Communicative Theory of Terminology and the Socio-Cognitive Theory of Perception, and using corpus linguistics, a database containing information on six variables, both linguistic and extralinguistic (the type of text, lexical variety, reformulation, terminological density, the role of the interlocutors and the type of text), has been analysed using a correlational method with a quantitative-qualitative approach, which has been complemented with automatic learning. The results show the complexity of legal discourse, given the difficulty of finding clear patterns of behavior in various variables, especially when compared with other disciplines such as medicine.

Keywords: Corpus; terminology; specialisation degree; perception; reformulation, specialised text.

Sumario: 1. Motivación del estudio. 2. Objeto de estudio, hipótesis y objetivos. 3. Fundamentación teórica. 4. Metodología. 4.1. Criterios de compilación del corpus. 4.2. La teoría de la percepción. 4.3. Variables de análisis. 4.4. Aprendizaje automático. 5. Resultados y análisis. 5.1. Según la clase y los tipos textuales. 5.2. Según las relaciones discursivas. 5.3. Según la variedad léxica. 5.4. Según la terminología. 5.5. Según la reformulación. 5.6. Resumen de los valores más representativos del corpus. 5.7. Resultados según aprendizaje automático. 5.7.1. Probabilidad de clasificación según variables. 5.7.2. Árbol de decisión. 6. Conclusiones. Referencias bibliográficas.

Cómo citar: Rodríguez-Tapia, S. (2023). Percepción, terminología e interlocutores del discurso jurídico en el grado de especialización textual, *Círculo de Lingüística Aplicada a la Comunicación* 94, 137-152. <https://dx.doi.org/10.5209/clac.70744>

1. Motivación del estudio

En la actualidad, es ampliamente reconocido que el discurso del derecho no es exclusivo de un acto comunicativo entre especialistas, sino que es posible encontrar dudas y respuestas por parte de legos en foros, entradas de blogs de temas jurídicos escritas por profesionales no especialistas que no necesariamente tienen instrucción formal en derecho, y una larga lista de muchas otras situaciones. Estos contextos son solo algunos ejemplos que ilustran la recontextualización pragmática a la que se somete el prototipo de texto especializado sobre derecho, que permite identificar nuevos patrones en el uso de la terminología o la reformulación, no típicos de lo que podía entenderse como texto resultado de una relación entre especialistas (es decir, los profesionales del derecho). Este cambio en el

¹ Universidad de Córdoba (España)
Correo electrónico: sergio.rodriguez@uco.es
N.º ORCID: <http://orcid.org/0000-0002-3439-4891>

paradigma está relacionado con una motivación profesional y personal de diferentes receptores, que se traduce en expectativas informativas diferentes y en un grado de especialización diverso. Así, en este artículo consideraremos como integrador del discurso jurídico cualquier manifestación lingüística delimitada por un entorno jurídico o que aborda un campo o contenido relacionado con el derecho, atendiendo a la concepción que plantea Campos Pardillos (2007: 164), y que va más allá de la discursividad jurídica que existe en el «núcleo duro» de los textos jurídicos, como las sentencias, las leyes, las demandas, los recursos, etc.

2. Objeto de estudio, hipótesis y objetivos

El eje vertical o grado de especialización (Hoffmann, 1998 [1987]: 69), que constituye nuestro objeto de estudio, es actualizable de forma lingüística (Cabré, 2002: 89), por lo que puede defenderse que es posible describir el grado de especialización del discurso jurídico atendiendo a la interrelación de variables como la densidad terminológica, la reformulación, el tipo textual, la variedad léxica o la relación entre emisor y receptor. En este trabajo pretendemos demostrar que:

- la densidad terminológica (en concreto, la densidad de términos crípticos) es el principal factor que permite discriminar entre grados de especialización textual en el discurso del derecho; es decir, permite distinguir entre textos especializados, semiespecializados y no especializados;
- la relación emisor-receptor según su competencia cognitivo-comunicativa condiciona la configuración discursiva de los textos y el grado de especialización; y que
- la reformulación terminológica se asocia de forma directa con la densidad terminológica y con el grado de especialización: un texto estará sujeto a mayor reformulación cuanto más denso terminológicamente sea dicho texto.

En este marco, los objetivos de este estudio pretenden:

1. Identificar qué elementos es posible vincular al grado de especialización, especialmente a la hora de modificarlo o condicionarlo.
2. Poner a prueba los cimientos teóricos que desarrolla el trabajo de Rodríguez-Tapia (2019) y determinar una propuesta de marco teórico-metodológico de partida, con los que poder estudiar grandes masas de datos, muy superiores a las del presente trabajo. De esta forma, este estudio pretende constituirse como proyecto piloto de un estudio más amplio, para determinar la viabilidad de la herramienta teórica y del método de análisis y prever posibles problemas por la complejidad del objeto. Para ello, se compararán algunos resultados con las conclusiones del trabajo sobre discurso médico de Rodríguez-Tapia (2020). La finalidad ulterior de este estudio se justifica en la capacidad de predicción de características de los discursos sobre el campo del derecho, que servirían de herramienta a la redacción de textos similares (Loureda, 2002: 158), así como a la traducción, a la terminografía, a la estilística del campo jurídico o a contribuir a los estudios sobre lenguaje claro.

3. Fundamentación teórica

2.1. Fundamentos sobre la comunicación especializada

Este estudio adopta una perspectiva funcionalista de la comunicación y del texto (Bernárdez, 1982: 85), que percibimos como un objeto dinámico (Bernárdez, 1995: 137), y estructurado en tres niveles (Dijk, 1983), de los que tomamos especialmente el superestructural.

Concebimos nuestro objeto de estudio, el discurso especializado, como una variedad de lengua diafásica y diastrática, sujeta a las normas del sistema lingüístico (Rodríguez-Tapia, 2019: 89). Defendemos, pues, que está sujeto a un contínuum de variabilidad, relacionado con el principio de variación de la Teoría Comunicativa de la Terminología (Cabré, 1998: 72) y que está caracterizado por cuatro factores principalmente (Cabré, 2007: 90; Da Cunha, Cabré, San Juan *et al.* 2011: 266); que constituyen nuestras variables de análisis:

- superestructural: relacionado con los géneros o tipos textuales que selecciona específicamente cada texto (Cabré y Estopà, 2005: 36);
- pragmático, relacionado con la teoría del registro de Halliday;
- cognitivo, por representar y comunicar conocimiento especializado; y
- lingüístico; en el que destacan elementos gramaticales (predominio de sustantivos, presencia de la tercera persona y del sujeto impersonal, etc.) y, específicamente, el nivel léxico, reconocido por los investigadores por ser característico del texto especializado (Cabré, Da Cunha, San Juan *et al.*, 2014: 224).

A propósito del factor pragmático, la concepción del texto como producto social (Bernárdez, 1995), las propuestas de relación en la comunicación especializada (entre especialistas, y entre especialistas y sociedad; Díaz Rojo, 2000: 70-72), y del término como unidad semántico-pragmática de la TCT (Cabré y Estopà, 2005: 77), hacen posible identificar la relevancia del papel de los interlocutores en la percepción del valor terminológico de las unidades

léxicas (Cabré, 1999: 132). Así, ejemplos como el siguiente permiten ilustrar que la noción de término, la especificidad semántica vinculada a la forma, podrían depender de quién emite y quién recibe el texto:

Nuevamente el TS se pronuncia sobre la *cláusula suelo* y abre la puerta a posibles reclamaciones independientemente de que se trate de *prestatarios* con un perfil cualificado. En su origen, el Juzgado de lo Mercantil núm. 10 de Barcelona dictó sentencia estimando íntegramente las peticiones de los consumidores, y declaró nula la *cláusula suelo* que incluía la *escritura de subrogación* en el préstamo hipotecario del *promotor*, condenando además a la devolución íntegra de las cantidades indebidamente cobradas en virtud de la cláusula (Gonzalvo, 2020).

Las formas que hemos marcado en cursiva podrían ser consideradas terminológicas por la mayoría de los usuarios, ya sean especialistas o no. Además, otras formas, como *préstamo hipotecario*, *reclamaciones* o *Juzgado de lo Mercantil*, si bien banalizadas en el discurso común, podrían ser consideradas terminológicas por algún grupo de receptores, por desconocimiento de la forma o del concepto, por lo que el carácter relativo del término dependería de factores pragmáticos.

3.2. Marco sobre la investigación en torno al discurso jurídico

La bibliografía especializada ha centrado su atención en los últimos 20 años en describir las características del denominado lenguaje jurídico, lengua del derecho o discurso del derecho. El estudio de Campos Pardillos (2007) caracteriza el discurso jurídico teniendo en cuenta su léxico, estilística y morfosintaxis. Así, sintetiza algunas de las características ya recogidas a lo largo de la investigación previa: tendencia arcaizante, empleo de fórmulas y fraseología estereotipada, redundancia mediante dobles y tripletes, y uso de latinismos, verbos performativos y de oraciones de gran longitud o complejidad sintáctica (ibíd.: 156 y ss.).

La investigación ha puesto el foco en el léxico jurídico, como se comprueba por los trabajos dedicados a la terminología jurídica en el ámbito del español. Por ejemplo, Martí (2004: 170 y ss.) distingue entre léxico ordinario y terminológico (asimilable a *délfico* y *críptico* o *semicríptico* en nuestra clasificación). Alcaraz Varó *et al.* (2016) prestan especial atención a la descripción léxico-semántica del léxico jurídico. Por su parte, desde una perspectiva morfosemántica, García Pérez (2018) reflexiona sobre las relaciones de jerarquía etimológicas y la creación de nuevas acepciones que entran en conflicto con la univocidad perseguida en el discurso especializado. Y desde una perspectiva formal, también existen trabajos que analizan la fraseología jurídica (Tabares Plasencia y Batista Rodríguez, 2013; Hourani-Martín, 2020), muy numerosos en los últimos años.

Asimismo, es posible encontrar estudios que contrastan el discurso jurídico entre diferentes lenguas: véanse los trabajos de Cobos López (2012) o Tabares Plasencia y Batista Rodríguez (2013) para el alemán, Campos y Campos (2016) para el francés, González Vallejo (2020) para el italiano y parte de la producción de Alcaraz Varó (2007, 2012; *et al.*, 2016) para el inglés.

La investigación en torno al grado de especialización del discurso jurídico no es tan intensa como la dedicada a caracterizar lingüísticamente esta actividad comunicativa. Sin embargo, es posible hallar acercamientos muy variados en los trabajos dedicados al *plain language* o lenguaje claro. La investigación sobre el lenguaje claro desde la lingüística aplicada es especialmente relevante para el análisis del grado de especialización del discurso jurídico, dado que centra su atención en aquellos elementos característicos de este discurso que provocan problemas comunicativos en los potenciales receptores y que son responsables de la brecha comunicativa existente entre la Administración y la ciudadanía. En el caso del español de España, la literatura científica ha publicado trabajos que dejan constancia de la falta de claridad del lenguaje jurídico (Carretera González y Fuentes Gómez, 2019), la necesidad de «modernizar el lenguaje jurídico» (Campos Moreno, 2019) o de optimizarlo mediante técnicas que afectan a los niveles discursivo, morfosintáctico o léxico (Montolío Durán, 2019; Da Cunha y Escobar, 2021).

Esta investigación pretende contribuir a caracterizar tres estadios diferentes del discurso del derecho según el grado de especialización, atendiendo especialmente a la relación entre los tipos textuales, las relaciones discursivas, la variedad léxica, la densidad terminológica, o la reformulación. Consideramos que caracterizar el discurso jurídico desde estas perspectivas contribuirá a la labor de la lingüística aplicada en su desarrollo de estrategias en la investigación en torno al lenguaje claro.

4. Metodología

Este estudio adopta un enfoque cualitativo-cuantitativo para realizar un análisis correlacional que vincule las diferentes variables. Para ello, empleamos tres herramientas específicas: la lingüística de corpus, la teoría de la percepción y el aprendizaje automático.

4.1. Criterios de compilación del corpus

Hemos seguido los criterios esbozados por Torruella (2017: 41-42) de clasificación de corpus para compilar un corpus textual monolingüe sincrónico y no etiquetado de textos escritos y planificados sobre temas relacionados con el derecho en español, que cubre los años entre 2007 y 2017.

Con respecto de su representatividad cualitativa, el corpus está constituido por siete tipos textuales, que coinciden con las variables de análisis: 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) y tesis doctoral (TES). Estos siete tipos textuales son comunes a multitud de ámbitos de conocimiento y no son exclusivos del ámbito jurídico, como pueden ser las leyes, las sentencias, las demandas o los recursos, por nombrar solo algunos. En este sentido, en esta investigación nuestro foco se encuentra en analizar el comportamiento discursivo en torno al ámbito del derecho en tipos textuales no restringidos al mundo jurídico. Desde el punto de vista temático, el único requisito de los textos para formar parte del corpus es que aborden un tópico relacionado con el derecho. De esta manera, en el corpus es posible encontrar textos relacionados con el derecho administrativo, civil, económico, fiscal, internacional, laboral, mercantil, penal o procesal. Así, el corpus no está constituido por textos de una rama u subrama específica del derecho, puesto que los temas elegidos son aleatorios y pretenden constituir una muestra heterogénea y representativa del campo.

Atendiendo a su representatividad cuantitativa, han sido seleccionados 15 textos por cada tipo textual (contamos, pues, con 105 textos), que han sido analizados por 4 tipos de informantes (en total, 420 análisis). La estrategia seguida ha sido la de muestreo aleatorio estratificado, dado que los textos seleccionados atienden a las variables de siete tipos textuales (para tres temas diferentes originalmente, aunque en este estudio solo se presentan los estudios correspondientes al ámbito del derecho). Se ha perseguido el equilibrio interno en el número de textos principalmente y, de forma secundaria, en el número de palabras, puesto que la cantidad de palabras solo sirve para obtener valores sobre la densidad terminológica y léxica o la reformulación (en ningún momento la descompensación cuantitativa del número de palabras afecta a otra variable). Los textos, además, han sido sometidos a procedimientos de selección de material lingüístico en casos de extensión excesiva o que descompensaría el corpus (por ejemplo, la selección aleatoria por secciones de textos como tesis doctorales o artículos científicos frente al material de blogs o noticias (Torruella, 2017: 146).

	ART	ENT	BLOG	NOT	FOR	DIV	TES	TOTAL
n.º palabras	63 682	25 843	18 202	11 398	13 617	12 899	47 628	193 269
% total	32,95 %	13,37 %	9,42 %	5,90 %	7,05 %	6,67 %	24,64 %	100 %
n.º formas léxicas	32 427	13 175	9 085	6 733	6 733	6 831	24 052	99 036
% total	32,74 %	13,30 %	9,17 %	6,80 %	6,80 %	6,90 %	24,29 %	100 %

Tabla 1. Número de palabras y de formas léxicas por tipos textuales en el corpus

Este conjunto de textos constituye un subcorpus de un corpus mayor que compara el derecho, la medicina y la informática. Por tanto, debe subrayarse que el reducido corpus que estudiamos pretende servir como proyecto piloto de un análisis de mayor envergadura, de forma que sea posible identificar tanto problemas en el objeto de análisis como problemas en la representatividad del corpus.

4.2. La percepción como instrumento

A propósito de las herramientas para identificar términos en los textos, L'Homme (2020: 72-75) distingue diversas técnicas, como la vinculación de la unidad con un campo concreto, la naturaleza de los argumentos del verbo, las relaciones morfológicas o las relaciones paradigmáticas del candidato a término. También existen técnicas relacionadas con la estadística, como las técnicas aplicadas en el análisis contrastivo de palabras clave (Tercedor y Láinez, 2017: 32) o los grafos de coocurrencia léxica, es decir, las formas que suelen coaparecer (Nazar y Cabré, 2011: 53), entre otros. En este estudio, empleamos una técnica basada en la percepción humana.

Los seres humanos usamos la capacidad cognitiva de la percepción con cualquier elemento de la realidad, incluidos los lingüísticos, para discriminar valores según un conjunto de juicios que empleamos para evaluar dichos elementos (en nuestro caso, para discriminar entre valores terminológicos y no terminológicos de las formas léxicas). Así, la percepción opera por un desajuste de una relación de elementos y su valor prototípico, por lo que sirve como una forma de identificar la variación; en nuestro caso, la variación diastrático-diafásica que caracteriza al grado de especialización. Específicamente, a partir de los elementos lingüísticos, los hablantes son capaces de discriminar y clasificar fenómenos basándose en su percepción (Caravedo, 2014: 115-116), en concreto, basándose en los fenómenos de ocurrencia (repetición de patrones) y prominencia: elementos que sobresalen por su escasa frecuencia y que se vinculan a cierto número de condicionantes, por ejemplo, la aparición de unidades consideradas terminológicas en un tipo de texto considerado prototípicamente no especializado resulta un elemento prominente, que capta la atención del hablante, quien activa una serie de procesos subjetivos.

A pesar de este carácter subjetivo, Caravedo (2014: 173) considera que si un conjunto de rasgos subjetivos identificado en un hablante es aceptado colectivamente (es decir, identificados por otros hablantes simultáneamente) y presentan «comportamientos regulares y motivados», «no constituyen producto del azar, de la imaginación o de la preferencia idiolectal del individuo».

La utilidad del empleo de la teoría de la percepción se relaciona con los datos proporcionados por un conjunto de 73 informantes, divididos en cuatro grupos (usuario medio de la lengua, traductor profesional, estudiante de

lingüística y el investigador del objeto de estudio, en calidad de hablante-oyente ideal desde la perspectiva generativa), con la intención de contar con percepciones heterogéneas. La diversidad de informantes permite evitar el sesgo del estudio que podrían provocar los análisis de un tipo exclusivo de informante. El objetivo de esta investigación no es la distinción y contraste de las percepciones y de los grupos de informantes, sino que pretende caracterizar el discurso del derecho atendiendo al conjunto de valoraciones y clasificaciones realizadas por todos los informantes, por lo que todos los datos aquí planteados serán tratados y analizados como pertenecientes a un único grupo de informantes. El contraste de la percepción por grupos constituye un objetivo de análisis distinto que se plantea en futuros trabajos.

4.3. Variables de análisis

Basándose en el criterio de prominencia, los informantes proporcionan información cuantitativa (número de formas léxicas délficas, semicrípticas y crípticas) y cualitativa (clasificación del texto según el emisor, el receptor y la clase) sin conocer ninguna información del texto que analizan (superestructura, autor, formato, etc.). El interés de este estudio por que los grupos de informantes determinen la clase de cada texto se encuentra en que se pretende establecer una relación entre la clase identificada por los informantes (nuestra variable dependiente: texto especializado, semiespecializado o no especializado) y los elementos lingüísticos que identifican como relacionados o vinculados. Por el contrario, establecer la clase *a priori* por parte del investigador del estudio podría condicionar desde un primer momento los elementos lingüísticos atribuidos a las clases, lo que derivaría en una descripción y explicación erróneas del fenómeno. Por su parte, el investigador aporta los datos relacionados con el tipo de texto, la variedad léxica, la densidad terminológica y sus tipos, y el índice de reformulación y sus tipos.

- a) **Clase de texto.** Variable dependiente del estudio, pues se pretende vincular el conjunto de características a los diferentes valores de la clase de cada texto. Sus valores posibles son: texto especializado (TE), texto semiespecializado (TSE) y texto no especializado (TNE).
- b) **Tipo de texto.** Variable nominal según la noción de superestructura. Constituye uno de los criterios de compilación del corpus, por lo que viene determinado *a priori*. Sus valores posibles son: artículo científico, artículo de revista o web de divulgación especializada, consultas y respuestas en foro en línea, entrada de blog, entrevista, noticia y tesis doctoral.
- c) **Emisor/receptor.** Variables ordinales según el perfil cognitivo-comunicativo de los interlocutores de los textos. Sus valores son: especialista (Esp), instruido (Ins) y lego (L).
- d) **Ratio types/tokens estandarizada (TTR).** Variable cuantitativa (con valores mayores que 0) que relaciona tipos de formas léxicas (*types*) y formas léxicas (*tokens*) según Templin (1957). Permite analizar la diversidad léxica: cuanto mayor es el porcentaje, mayor es la diversidad léxica del texto. En este estudio se considera forma léxica toda aquella forma lingüística con lexema, sin debatir acerca de la noción de unidad lingüística en fraseología, puesto que nuestro interés es simplemente estadístico. Sobre el total de *tokens* de un texto se han filtrado las formas léxicas empleando una lista de exclusión de formas gramaticales.
- e) **Índice de densidad terminológica (IDT).** Variable cuantitativa (con valores entre 0 y 1) resultante de dividir el número total de formas léxicas terminológicas entre el número total de formas léxicas del texto. En esta variable se distinguen tres subtipos, similares a los de Alcaraz, Hughes y Gómez (2016: 56-64; vocabulario general, semitécnico y técnico):
 - IDT a) o IDT de términos délficos: el usuario conoce la denominación y el concepto, aunque no necesariamente de forma precisa. Por ejemplo, *sentencia*. En este trabajo se marcarán con el color azul.
 - 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 pobre. Por ejemplo, *dolo*. En este trabajo se marcarán con el color verde.
 - IDT c) o IDT de términos crípticos: el usuario desconoce la denominación y el concepto. Por ejemplo, *adehala*. En este trabajo se marcarán con el color amarillo.
- f) **Índice de reformulación (IR).** Variable cuantitativa (con valores normalizados entre 0 y 1) resultante de dividir el número de procedimientos de reformulación formal de un texto (como el empleo de formas denominativas alternativas, paréntesis o marcadores como «es decir», etc.) entre el IDT de dicho texto. Es una variable que codifica la reformulación a la que se somete la terminología relativa a dicho texto. Distingue tres tipos, según los planteados por Garcés (2008: 73), que relacionamos como ejemplos a continuación (aparecen en negrita el término reformulado y subrayado el procedimiento de reformulación):
 - IR a) o IR de procedimientos por expansión: «[...] una de las cuestiones que más interesa acerca de estos instrumentos de valoración del riesgo de violencia es la **capacidad predictiva** que tienen, esto es, con qué grado de acierto efectúan los pronósticos sobre la conducta futura de las personas» (Martínez Garay, 2016) ^[DER ART09].
 - IR b) o IR de procedimientos por reducción: «En otras palabras, **qué probabilidad hay de equivocarse cuando los operadores jurídicos toman decisiones basadas en estos pronósticos, ya sea por haber considerado de alto riesgo a un sujeto que en realidad luego no vuelve a delinquir (falsos positivos)**» (Martínez Garay, 2016) ^[DER ART09].
 - IR c) o IR de procedimientos por variación: «[...] entonces el Juez señalará el **lanzamiento**, (Desahucio) que conllevará el embargo de cuentas, nóminas, bienes etc.» (leonjbr, 2017) ^[DER FOR02].

4.4. Aprendizaje automático

Otra de las herramientas empleadas en el estudio, que ha facilitado identificar cómo se vinculan las variables, es el aprendizaje automático. Ese instrumento permite (1) identificar la probabilidad de que un texto sea clasificado según su grado de especialización según los valores del resto de variables; y (2) hallar qué combinaciones de dichas variables condicionan la clasificación de un texto en alguna de las tres clases. Para ello, hemos empleado el software WEKA con los resultados de los 420 análisis, en el que hemos empleado los algoritmos de red bayesiana y J48, con una configuración de árbol podado, una validación cruzada de 10 iteraciones y un intervalo de confianza de 0,0015, que resultan en un 79,05 % de tasa de éxito en la clasificación.

5. Resultados y análisis

Los gráficos y ejemplos siguientes representan una síntesis de las relaciones analizadas entre la clase textual y el resto de variables.

5.1. Según la clase y los tipos textuales

El Gráfico 1 relaciona el número de tipos de texto asociados a cada clase. A partir de dichos datos, se observan dos asociaciones claras: (a) el artículo científico y la tesis quedan vinculados casi exclusivamente al TE; (b) en menor medida, el foro queda vinculado al TNE. Los otros tipos de texto parecen presentar un equilibrio, orientado al TSE en el caso del blog y de la revista de divulgación; y orientado al TNE en la entrevista o la noticia.

Esta distribución es diferente a la que presenta el análisis del subcorpus sobre medicina, campo que muestra una distribución mucho más polarizada (Rodríguez-Tapia, 2020: 81), lo cual ya deja entrever que el discurso jurídico no muestra un comportamiento tan regular como otros campos.

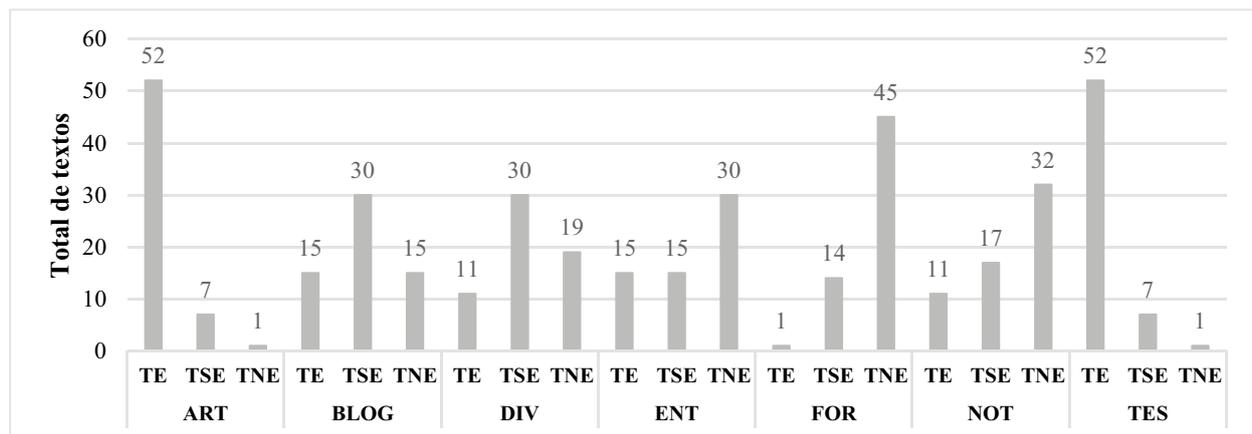


Gráfico 1. Clases de textos según tipo textual

La disparidad en las relaciones refleja la complejidad de la asociación de superestructura y clase de texto en los tipos textuales de la divulgación. Si bien en un primer momento podría considerarse que el tipo de organización formal de tesis o artículo (introducción, metodología, resultados, etc.) podría condicionar la clasificación del informante, debe recordarse que los informantes analizan fragmentos de secciones aleatorias, por lo que los argumentos que llevan a la clasificación no pueden hallarse en la distribución formal del texto.

Los datos anteriores parecen representados en el siguiente eje horizontal, que distingue límites discretos entre categorías. Así, se aprecia cómo ciertos tipos textuales se aproximan o alejan del límite que separa el TSE del TNE:

TE	TSE	TNE
ART		
TES		
	DIV	
	BLOG	
		ENT
		NOT
		FOR

Tabla 2. Distribución de tipos textuales según clases

5.2. Según las relaciones discursivas

El Gráfico 2 muestra la relación entre 3 variables en el eje horizontal: la primera desde arriba corresponde con el receptor; la siguiente, con el emisor; y la última, con la clase. Según estos datos, el TE nunca se dirige a un usuario lego y casi siempre está emitido por especialistas, lo que podría relacionarse con el hecho de que el emisor presuponga que el receptor domina necesariamente el conocimiento especializado (lo que explicaría que no existan receptores legos). Además, en contraste, el TNE destaca por la relación instruido-lego. En el caso del TSE, destaca el receptor instruido (tanto el interlocutor como la clase coinciden en su estatus de objeto de transición).

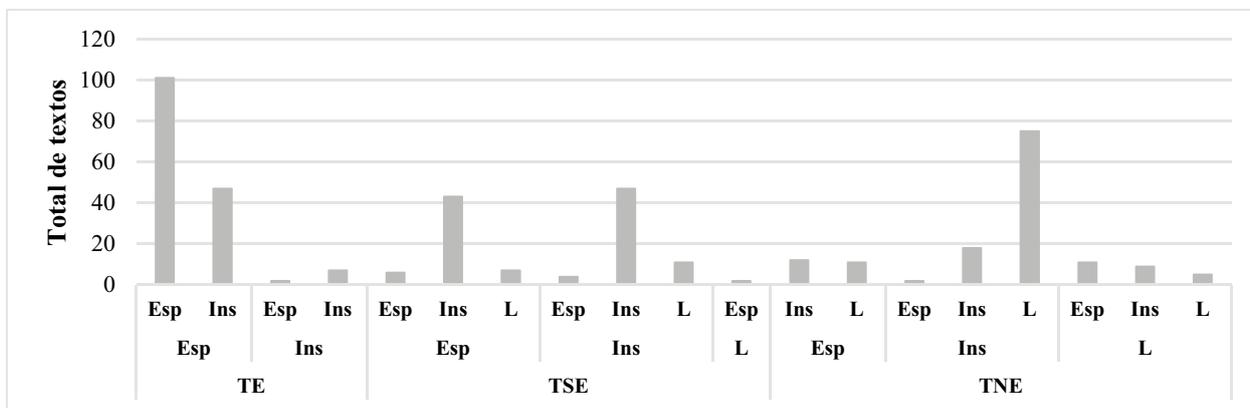


Gráfico 2. Relaciones discursivas según clase de texto

El proyecto de Ley de Contratos del Sector Público (artículo 32), en tramitación en este momento en las Cortes Generales, contiene una regulación más detallada de la encomienda a medios propios, adaptándose a lo establecido en la Directiva 2014/24/UE sobre contratación y precisando algunas cuestiones que estaban deficientemente reguladas. La nueva regulación de los denominados encargos de los poderes adjudicadores a medios propios personificados, cuando entre en vigor, tendrá incidencia en la constitución de comercializadoras municipales para el aprovisionamiento de energía eléctrica. Veamos cuales son los aspectos novedosos que contiene el artículo 32 del Proyecto de Ley, que deberán tenerse en cuenta al constituir una comercializadora que actúe como medio propio personificado de un Municipio (Corvinos, 2017) ^[DER BLOG10].

Tabla 3. Texto semiespecializado con emisor especialista y receptor instruido

Hacienda también remarca que las plazas creadas en virtud del Acuerdo de Estabilización no se detraen de las que se abrirán en los distintos sectores según la tasa de reposición que les atribuye el artículo 19.1 de la Ley de PGE: el 100% para los sectores prioritarios -Sanidad, Educación, Justicia, seguridad, control y lucha contra el fraude fiscal, laboral, de subvenciones públicas y contra la Seguridad Social, etc.- y el 50% en el resto. “Por eso la oferta de empleo es tan extraordinaria”, aseveran desde el Ministerio (Rosal, 2017) ^[DER DIV14].

Tabla 4. Texto semiespecializado con emisor y receptor instruidos

Las tablas 4 y 5 cuentan con el mismo tipo de emisor y diferente receptor a pesar de ser muy similares en el tipo de terminología que los informantes identifican. La clasificación en clases y receptores diferentes, en este caso, se deberá a otro tipo de elementos lingüísticos, como por ejemplo la reformulación o la densidad terminológica (no la cantidad de formas terminológicas).

Ahora bien, los pagos... el Estado se retrasa a la hora del pago, se cobra de una forma trimestral, y los baremos... Desde 1996 no han variado los honorarios [según el Reglamento de asistencia jurídica gratuita]. Son 21 años. El coste de la vida desde el año 2002 [el último que permite ver el INE] ha subido un 36,1%. Hablamos que el coste de la vida se ha encarecido aproximadamente un 40%, y los honorarios del Turno de Oficio, no. Es verdaderamente indigno. El Turno de Oficio se paga por debajo del salario mínimo interprofesional y es una verdadera vergüenza que tengamos estas retribuciones (González y Lafuente, 2017) ^[DER ENT04].

Tabla 5. Texto no especializado con instruido y receptor lego

5.3. Según la variedad léxica

De acuerdo con el Gráfico 3, la variedad léxica de un texto es inversamente proporcional a su grado de especialización y su densidad terminológica, si bien las diferencias no destacan entre TSE (TTR = 66,8 % e IDT = 0,222) y TNE (TTR = 67 % e IDT = 0,176), en contraste con el TE, que se perfila como la clase menos variada, debido posiblemente a la repetición de formas y estructuras que caracteriza a esta clase según los estudios en torno al discurso especializado.

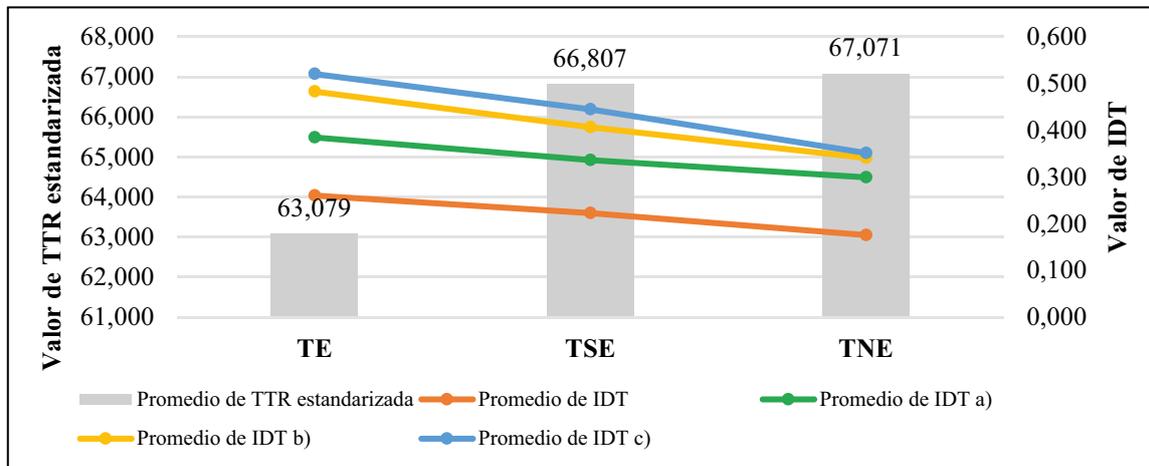


Gráfico 3. Relación de variedad léxica y densidad terminológica según clase de texto

A continuación, las tres tablas muestran en negrita las formas léxicas repetidas, lo que permite comprobar visualmente cómo dicha densidad (el carácter repetitivo de un texto) disminuye al reducirse su grado de especialización.

Considero que es **totalmente** incorrecta la **aplicación** del **principio** de **intervención mínima** y esto es, porque si leemos el artículo 65.5 apartado k) del **punto 5.º** del art. 65 de la Ley de tráfico ya nos está dando la pista, al **considerar** dicho comportamiento como falta muy grave, cuando no sea constitutiva de **delito** y lógicamente **conducir** sin **permiso** de **circulación** o tras la pérdida de vigencia del **permiso** de **circulación** o **conducir** con la pérdida de todos los **puntos** asignados legalmente, nos lleva a pensar que el sujeto que conduce en tal tesitura, ya a cometido todas las infracciones administrativas posibles y es plenamente conocedor de lo que está realizando, por lo que en el momento de cometer el ilícito, en la **conducta** se añade el plus de antijuridicidad exigido para **considerarlo delito**, al saber y conocer el infractor que su **conducta** es **totalmente** contraria a derecho, por lo que la **aplicación** del **principio de intervención mínima** no es la correcta para estos supuestos (Ruiz Rodríguez, 2017) ^[DER DIV07].

Tabla 6. Texto especializado con una variedad léxica del 50 %

J. E. V.: Sintéticamente, se podrían señalar los siguientes: la figura del **obispo** como **juez** inmediato y responsable de los procesos; ejecutividad de sentencia afirmativa de **primera instancia**; **juez único**, clérigo, bajo la responsabilidad del **obispo**, en **primera instancia**; **posibilidad** de dos **jueces** laicos dentro del colegio de **jueces**; ampliación del fuero competente; reducción de los trámites de la admisión de la demanda, citación y contestación de la lite; mayor valor probatorio de la declaración de las partes y del testigo **único**; **posibilidad** de presentar recurso de **apelación** a la sede del **obispo** metropolitano (o a la sede apostólica); carga económica de la Conferencia Episcopal, para asegurar la gratuidad del servicio, **posibilidad** de rechazo in limine de la **apelación** cuando sea dilatoria, etc. (Ámbito jurídico, 2015) ^[DER ENT14].

Tabla 7. Texto semiespecializado con una variedad léxica del 67 %

Por ello, el ponente de la sentencia, el magistrado del Moral García, estima que “no bastan por ello unos episodios, más o **menos** intensos o más o **menos** numerosos pero concentrados en pocos días y sin nítidos visos de continuidad, que además no comporten repercusiones en los hábitos de la **víctima**. La doctrina aprobada por el Pleno de la Sala establece que el delito de hostigamiento exige implícitamente una cierta prolongación en el tiempo o, al **menos**, que “quede patente, que sea apreciable, voluntad de perseverar en esas acciones intrusivas, que no se perciban como algo puramente episódico o coyuntural, pues en ese caso no serían idóneas para alterar las costumbres cotidianas de la **víctima** (Poder Judicial España, 2017) ^[DER NOT06].

Tabla 8. Texto no especializado con una variedad léxica del 78 %

Según los tipos textuales, los más variados son el foro, el artículo y la tesis (Gráfico 4), lo que contrasta con el Gráfico 3 y el Gráfico 1, puesto que la participación en foro suele considerarse un TNE, lo que parece constituir una excepción de la variedad léxica y su relación con el grado de especialización. En contraste, encontramos la entrevista y la noticia, que suelen adoptar la forma de TNE (Gráfico 1), seguidos del blog y la revista de divulgación, propios del TSE.

El discurso jurídico plantea un patrón de diversidad léxica según el grado de especialización irregular si se compara con el patrón del discurso médico, cuyos tipos textuales se relacionan de forma directa con el grado de especialización y de forma inversa con la diversidad léxica (Rodríguez-Tapia, 2020: 83).

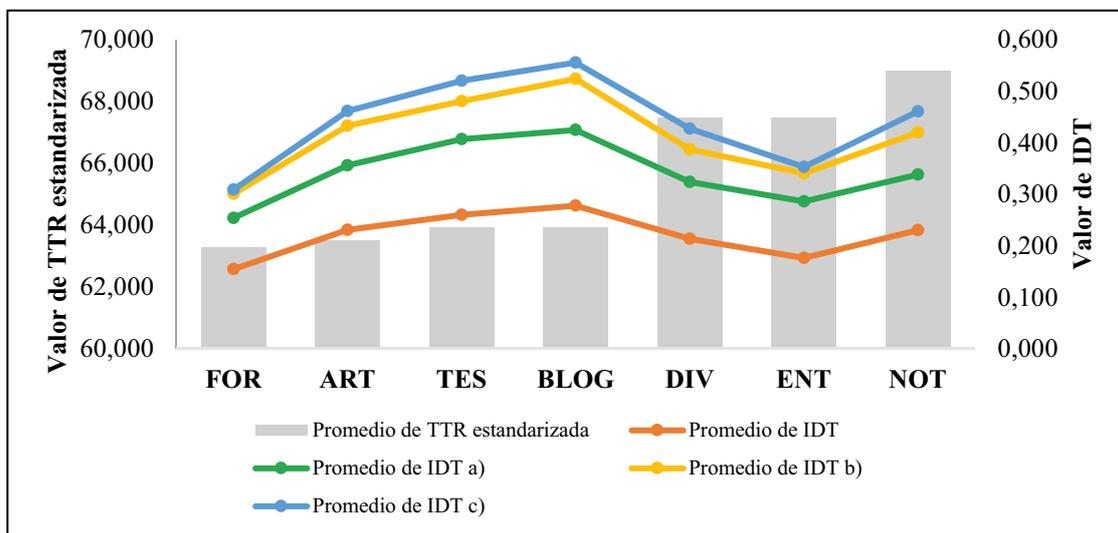


Gráfico 4. Relación de variedad léxica y densidad terminológica según tipo de texto

5.4. Según la terminología

Vinculado estrechamente al TE, el Gráfico 5 revela que el IDT y el grado de especialización establecen una relación directa, puesto que la clasificación en el grado aumenta conforme lo hace la densidad terminológica. No obstante, no todos los tipos de IDT permiten discriminar el grado: solo el IDT b) (de formas semicrípticas, es decir aquellas en las que el usuario conoce la denominación por su frecuencia, pero no el concepto o lo conoce de forma muy pobre) muestra las diferencias cuantitativas relevantes para distinguir el grado, por lo que podría defenderse que solo la densidad de términos semicrípticos (como *reenvío*, *artículo*, *acuerdo*, o *efecto*) permitiría contribuir a distinguir entre TE, TSE y TNE (si bien, como se apreciará en la tabla 13, no lo discrimina en todos los tipos textuales).



Gráfico 5. Relación de IDT según clases

Con respecto a la medicina, el derecho presenta diferencias menos notables en el IDT entre clases, lo que dificulta discriminar textos en la clasificación para el tema jurídico. Esta diferencia sutil puede justificar la dispersión que existe entre las tres clases en los diferentes tipos de texto para el tema jurídico (véase Gráfico 1).

A propósito de los tipos de densidad terminológica, se puede apreciar que solo el tipo (b), la densidad de términos semicríticos, es la única que muestra una diferencia destacable que permita distinguir las clases del grado de especialización. En las tablas 9 y 10 se ofrecen dos fragmentos de textos con diferente densidad terminológica (apreciable en la cantidad de formas marcadas en color) y en el tipo de terminología percibida por los informantes (existe mayor cantidad de formas semicríticas en verde en el TE de la tabla 9 que en el TNE de la tabla 10).

Además, por lo que se refiere a los matrimonios de las confesiones acatólicas, el reenvío que el artículo 7 de los Acuerdos hace a sus normativas propias a los efectos de regular la forma de celebración matrimonial, carece en realidad de trascendencia jurídica desde el momento en que el Estado establece otros requisitos formales sin los cuales estos matrimonios no desplegarán efectos: la tramitación de un expediente matrimonial previo, la intervención de un ministro religioso debidamente legitimado que autorice las nupcias y la participación de dos testigos mayores de edad. En base a esto, no puede decirse que se esté reconociendo la forma ritual religiosa de estas confesiones, sino que se está reconociendo una forma religiosa «pactada», que no viene a ser más que la traslación de la forma civil al ámbito religioso, en la que el ministro de culto hace las veces del juez o del funcionario civil que asiste el enlace junto con los testigos mayores de edad y certifica su celebración. Solo el respeto de los requisitos acordados conllevará la eficacia civil del matrimonio (Berenguer, 2015) ^[DER ART03].

Tabla 9. Artículo científico con alta densidad terminológica

Hola mi caso es el siguiente tengo tarjeta comunitaria mi esposo es español y me toca renovar porque ya se me caduco el 16 de nov. 2014 pedi cita y fuimos llevando los requisitos que pedian el x19 padron certificado d matrimonio y otras cosas mas entre ellas medios económicos pues el caso esq mi esposo no esta trabajando y yo trabajo echando horas en algunas casas de familia pero no tengo contrato cuando a depositar para renovar mi pregunta es pierdo yo mis documentos por no conseguir trabajo o no tener los medios económicos? ?? Xfa estoy todavia dentro del tiempo d los 3 meses del tiempo para renovar necesito ayuda urgente gracias (josias, 2015) ^[DER FOR12].

Tabla 10. Blog con baja densidad terminológica

5.5. Según la reformulación

El IR depende del total de formas terminológicas, por lo que sirve para medir cuántas veces la terminología se expresa con instrumentos formales diferentes con el objetivo de facilitar el acceso del receptor al conocimiento especializado.

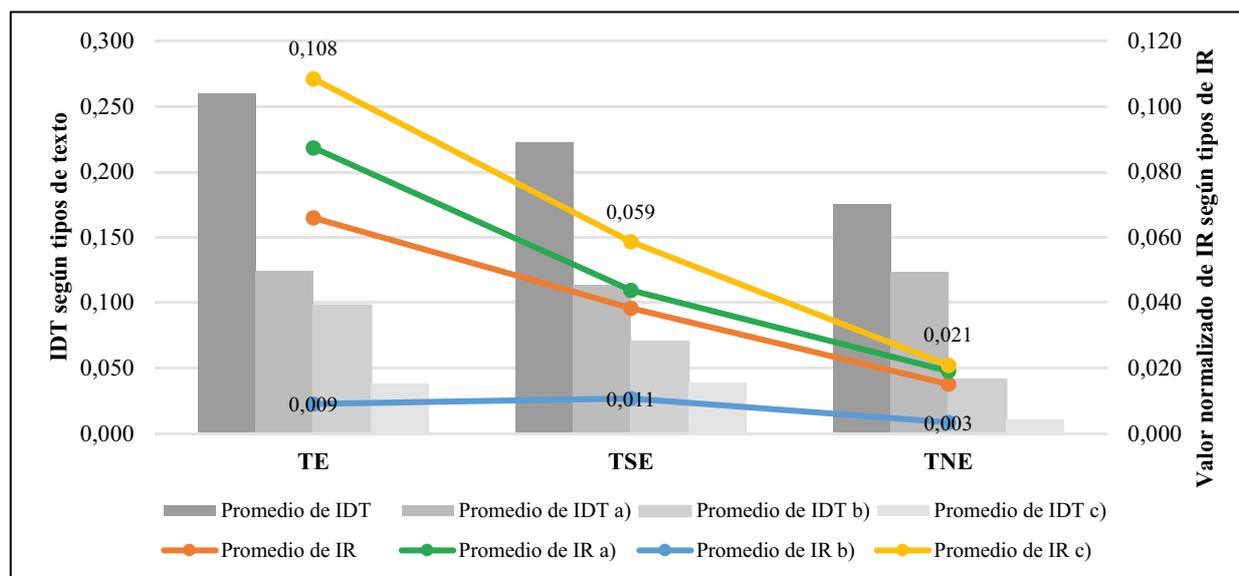


Gráfico 6. IR según clase de texto e IDT

En el Gráfico 6 se aprecia la relación directa entre IDT e IR, ya que la reformulación de un texto aumenta a medida que lo hace su densidad terminológica. En concreto, de los cuatro valores, el IR general, el IR por expansión y el IR por variación sí mantienen esta relación directa, por lo que una mayor cantidad de uso de explicaciones y variantes denominativas contribuye a discriminar el grado de especialización, puesto que aumentan la probabilidad

de que un texto sea considerado especializado. En el caso de la reformulación por reducción², las diferencias son tan poco notables que no permiten discriminar entre textos (este fenómeno se repite en el subcorpus de textos médicos; Rodríguez-Tapia, 2020: 86). Las siguientes tablas presentan ejemplos con diferente densidad terminológica y procedimientos de reformulación que se emplean (subrayado rojo para la expansión, negro para la reducción y morado para la variación).

Desde el punto de vista del **Derecho penal y penitenciario** – y con independencia de muchas otras cuestiones problemáticas que suscita el paradigma de la prevención en el que se inserta esta preocupación por la peligrosidad de los **delincuentes** (cfr., muy críticamente, Harcourt, 2007 o Albrecht, 2004) – una de las cuestiones que más interesa acerca de estos **instrumentos de valoración del riesgo de violencia** es la **capacidad predictiva** que tienen, esto es, con qué grado de acierto efectúan los pronósticos sobre la conducta futura de las personas. En otras palabras, qué probabilidad hay de equivocarse cuando los operadores jurídicos toman decisiones basadas en estos pronósticos, ya sea por haber considerado de alto riesgo a un sujeto que en realidad luego no vuelve a delinquir (**falsos positivos**), o por haber considerado de riesgo bajo a una persona que sin embargo más tarde sí comete nuevos delitos violentos (**falsos negativos**) (Martínez Garay, 2016) ^[DER ART09].

Tabla 11. Ejemplos de reformulación en texto especializado

Resumiendo : si el día que tengan establecido ustedes en el **contrato** como día final para el pago de la renta , por ejemplo es el 5 de cada mes, como ya esta vencida la renta del mes de Mayo, podrá **interponer** usted la **demandade Desahucio** por medio de un Abogado, y que el juzgado debe admitir y el Juez reclamara el pago en los siguientes 20 días y si sigue sin hacer el pago, entonces el Juez señalará el **lanzamiento, (Desahucio)** que conllevará el **embargo de cuentas, nóminas, bienes etc.** (leojbr, 2017) ^[DER FOR02].

Tabla 12. Ejemplos de reformulación en texto no especializado

5.6. Resumen de los valores más representativos del corpus

A partir de los 420 análisis, la tabla 13 indica la moda (en las variables cualitativas) y los promedios (en las variables cuantitativas) de los datos más representativos del corpus analizado para cada tipo de texto (filas) en relación con cada variable (columnas). En las variables cuantitativas los valores máximos han sido marcados en verde y los mínimos, en rojo.

Tema y tipo de texto	TTR	IDT				IR				Emisor	Receptor	Clase
		Global	a	b	c	Global	a	b	c			
ART	63,493	0,231	0,126	0,076	0,029	0,090	0,105	0,023	0,144	Esp	Esp	TE
TES	63,908	0,260	0,147	0,074	0,040	0,088	0,119	0,011	0,142	Esp	Esp	TE
DIV	67,467	0,214	0,110	0,063	0,040	0,025	0,027	0,006	0,050	Ins	Ins / L	TSE
ENT	67,471	0,176	0,110	0,054	0,012	0,042	0,059	0,007	0,049	Ins	Ins	TNE
BLOG	63,935	0,278	0,147	0,099	0,032	0,021	0,026	0,004	0,032	Esp	Ins	TSE
NOT	68,982	0,230	0,108	0,082	0,041	0,008	0,007	0,001	0,026	Ins	L	TNE
FOR	63,268	0,154	0,100	0,047	0,008	0,011	0,019	0,001	0,006	L / Ins	Ins	TNE
Promedio	65,503	0,220	0,121	0,071	0,029	0,041	0,052	0,008	0,064			

Tabla 13. Resumen global de valores más representativos para cada variable según tipo de texto

De todos los datos anteriores, los más llamativos se encuentran en el blog y la noticia. A propósito del blog, llama la atención que este tipo textual, junto con la tesis, sea el que cuente con mayor IDT a). Esto puede deberse a que se trata de terminología délfica, es decir, menos opaca al receptor lego. Lo mismo sucedería con la densidad de términos semicrípticos, elemento discriminante del grado en los textos sobre derecho. De hecho, en la tabla 2 el blog es el tipo textual que más cerca se encuentra del TE, sin incluirse dentro de él, lo cual es compatible con el hecho de que suele estar emitido por especialistas y dirigido a instruidos (véase tabla 13). Junto con el blog, destaca la noticia

² Cabe recordar que la densidad terminológica de formas semicrípticas (IDT b) y la reformulación por reducción (IR b) no guardan relación, puesto que miden fenómenos diferentes. Dicho de otra forma, el IR b) no mide la reformulación de los términos semicrípticos, sino las veces que se reformula cualquier tipo de término mediante procedimientos de reducción.

en IDT b), que es el tipo textual que encabeza el IDT c), muy cerca de la tesis o de la revista de divulgación. Por el contrario, el blog y la noticia apenas reflejan reformulación, lo que sugiere que los informantes son capaces de identificar el valor especializado de las formas léxicas pero que estas formas no suelen someterse a reformulación, bien porque se sobreentienden expectativas cognitivo-comunicativas en el receptor, bien porque se considera que la terminología es lo suficientemente transparente desde el punto de vista conceptual, bien porque la naturaleza de la terminología del derecho difiere de la de otros campos. De hecho, según la moda de los tipos de interlocutores de la tabla 13, queda patente el tipo de expectativas diferentes para el blog y para la noticia.

5.7. Resultados según aprendizaje automático

5.7.1. Probabilidad de clasificación según variables

Para presentar las probabilidades de clasificación se ha empleado el teorema de Bayes. Si usamos ejemplos para ilustrar el funcionamiento de la tabla, este instrumento permitiría (1) determinar la probabilidad de que un texto corresponda con una tesis doctoral si conocemos que se trata de un TE; y (2) indicar la probabilidad de que un texto sea un TE si conocemos que es una tesis doctoral. En (1) la probabilidad se discrimina verticalmente y en (2), horizontalmente.

	Valor del atributo	(1) Probabilidad del valor de la variable conociendo el valor de la clase			(2) Probabilidad del valor de la clase conociendo el valor de la variable		
		TE	TSE	TNE	TE	TSE	TNE
Tipo texto	ART	33,12 %	5,83 %	0,70 %	86,67 %	11,67 %	1,67 %
	BLOG	9,55 %	25,00 %	10,49 %	25,00 %	50,00 %	25,00 %
	DIV	7,01 %	25,00 %	13,29 %	18,33 %	50,00 %	31,67 %
	ENT	9,55 %	12,50 %	20,98 %	25,00 %	25,00 %	50,00 %
	FOR	0,64 %	11,67 %	31,47 %	1,67 %	23,33 %	75,00 %
	NOT	7,01 %	14,17 %	22,38 %	18,33 %	28,33 %	53,33 %
	TES	33,12 %	5,83 %	0,70 %	86,67 %	11,67 %	1,67 %
Emisor	Esp	94,27 %	46,67 %	16,08 %	65,20 %	24,67 %	10,13 %
	Ins	5,73 %	51,67 %	66,43 %	5,42 %	37,35 %	57,23 %
	L	0,00 %	1,67 %	17,48 %	0,00 %	7,41 %	92,59 %
Receptor	Esp	77,78 %	8,45 %	5,97 %	84,85 %	9,09 %	6,06 %
	Ins	22,22 %	69,72 %	28,36 %	18,93 %	58,58 %	22,49 %
	L	0,00 %	21,83 %	65,67 %	0,00 %	26,05 %	73,95 %
IDT global	<0,1025	6,37 %	5,83 %	28,67 %	17,24 %	12,07 %	70,69 %
	0,1025-0,1945	23,57 %	39,17 %	36,36 %	27,21 %	34,56 %	38,24 %
	0,1945-0,3935	56,69 %	52,50 %	31,47 %	45,18 %	31,98 %	22,84 %
	>0,3935	13,38 %	2,50 %	3,50 %	72,41 %	10,34 %	17,24 %
IDT b)	<0,0075	0,00 %	4,17 %	17,48 %	0,00 %	16,67 %	83,33 %
	0,0075-0,0525	30,57 %	31,67 %	58,04 %	28,40 %	22,49 %	49,11 %
	0,0525-0,1405	40,13 %	59,17 %	21,68 %	38,18 %	43,03 %	18,79 %
	>0,1405	29,30 %	5,00 %	2,80 %	82,14 %	10,71 %	7,14 %
IDT c)	<0,0015	19,11 %	9,17 %	48,95 %	27,03 %	9,91 %	63,06 %
	0,0015-0,0585	51,59 %	60,83 %	46,15 %	36,82 %	33,18 %	30,00 %
	0,0585-0,151	28,66 %	30,00 %	4,90 %	51,14 %	40,91 %	7,95 %
	>0,151	0,64 %	0,00 %	0,00 %	100,00 %	0,00 %	0,00 %
IR global	<0,060801	54,78 %	89,17 %	95,80 %	26,06 %	32,42 %	41,52 %
	>0,060801	45,22 %	10,83 %	4,20 %	78,89 %	14,44 %	6,67 %
IR a)	<0,075332	57,96 %	90,00 %	94,41 %	27,25 %	32,34 %	40,42 %
	>0,075332	42,04 %	10,00 %	5,59 %	76,74 %	13,95 %	9,30 %

	Valor del atributo	(1) Probabilidad del valor de la variable conociendo el valor de la clase			(2) Probabilidad del valor de la clase conociendo el valor de la variable		
		TE	TSE	TNE	TE	TSE	TNE
IR b)	<0,001595	56,05 %	71,67 %	79,72 %	30,56 %	29,86 %	39,58 %
	>0,001595	43,95 %	28,33 %	20,28 %	52,27 %	25,76 %	21,97 %
IR c)	<0,009085	29,30 %	53,33 %	71,33 %	21,70 %	30,19 %	48,11 %
	0,009085-0,022595	4,46 %	0,00 %	0,00 %	100,00 %	0,00 %	0,00 %
	0,022595-0,220056	51,59 %	42,50 %	28,67 %	46,82 %	29,48 %	23,70 %
	>0,220056	14,65 %	4,17 %	0,00 %	82,14 %	17,86 %	0,00 %

Tabla 14. Probabilidades de clasificación según el grado de especialización y las diferentes variables

Esta tabla contribuye a reflexionar sobre el valor condicionante de ciertas variables sobre otras y oponer aquellos casos poco claros (como el emisor instruido y el tipo textual BLOG) frente a casos cercanos a los valores binarios (como el tipo ART o TES, el lego como emisor, el especialista como receptor o el IDT c) o IR c), que facilitan la discriminación de clases).

5.7.2. Árbol de decisión

Por último, el árbol de decisión³ permite identificar qué combinación de valores condicionan cada clase. Un ejemplo observable es la Ilustración 1, que identifica el emisor como elemento condicionante principal: si el emisor es lego, la clase será TNE; si es instruido, la clase dependerá de su receptor en primera instancia, y de la densidad terminológica en segunda instancia. Los elementos que determinan la clase finalmente en los casos complejos son la variedad léxica (TRR) o la reformulación (IR), que, además, son variables que no discriminan siempre entre las mismas clases.

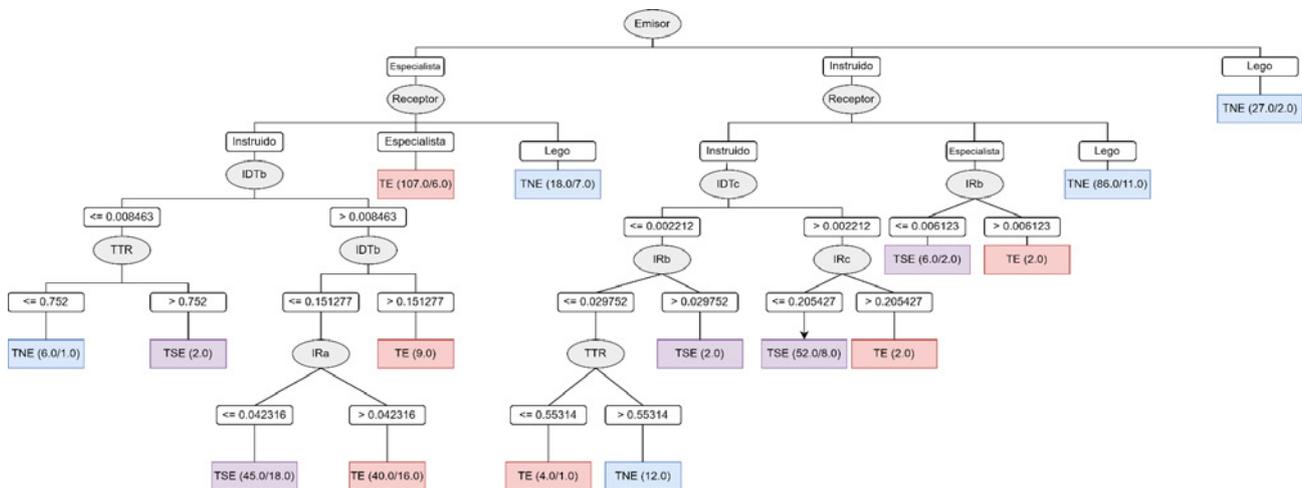


Ilustración 1. Árbol de decisión

El árbol de decisión refleja la complejidad del grado de especialización del discurso jurídico cuando el receptor de un texto es instruido, lo cual coincide con su carácter intermedio entre dos perfiles más o menos diferenciados.

6. Conclusiones

Este estudio ha pretendido identificar cómo se relacionan las diferentes variables que afectan a la clasificación del texto sobre derecho según su grado de especialización. Para ello, hemos empleado una propuesta teórico-metodológica, con el fin de medir su alcance y limitaciones a la hora de analizar el grado de especialización

³ Los números que aparecen en el árbol de decisión son los resultados exitosos y erróneos (a izquierda y derecha de la barra respectivamente) en la clasificación del algoritmo. La sobreesimplificación del esquema conlleva errores de clasificación al tratarse de un árbol podado, es decir, que omite información en pos de la sencillez del esquema.

textual del discurso jurídico. En este trabajo ha sido posible comprobar los siguientes fenómenos relacionados con nuestras hipótesis:

- Sobre la densidad terminológica, al contrario de lo planteado en las hipótesis, el elemento que contribuye a discriminar los textos según su grado de especialización es el IDT de formas semicripticas, es decir aquellas en las que el usuario conoce la denominación por su frecuencia, pero no el concepto, o lo conoce de forma muy pobre. No ha sido posible confirmar su carácter distintivo en el árbol de decisión (lo que sí fue posible al analizar los resultados del corpus de textos médicos; *cf.* Rodríguez-Tapia, 2020: 87-88). Sí se puede afirmar, no obstante, que cuanto mayor es la densidad terminológica general, mayor probabilidad existe de que un texto sea considerado especializado. En relación con esta cuestión, la variedad léxica se antoja inversamente proporcional a la densidad terminológica y no encuentra apenas diferencias significativas en los casos de TSE (TTR = 66,8 % e IDT = 0,222) y TNE (TTR = 67 % e IDT = 0,176), en contraste con el TE.
- Sobre la relación emisor-receptor, se ha podido comprobar que los tipos textuales suelen asociarse a cierto tipo de relaciones discursivas, lo que influye en la clasificación del grado de especialización. Por ejemplo, se comprueba que el texto especializado nunca se dirige a un usuario lego y casi siempre está emitido por especialistas, lo que podría relacionarse con el hecho de que el emisor presuponga que el receptor domina necesariamente el conocimiento especializado (lo que explicaría que no existan receptores legos). En concreto, el aprendizaje automático determina el emisor como elemento condicionante principal del grado, y no el receptor, como sí sucede con el corpus médico (*cf.* Rodríguez-Tapia, 2020: 87). Esta diferencia sugiere que es posible que el grado de especialización textual se comporte de forma diferente según el tema del texto (lo que no quiere decir que el tema imponga una clase de grado).
- Sobre la reformulación, ha sido posible validar nuestra hipótesis en parte, dado que la reformulación sí que está relacionada de forma directa con la densidad terminológica (y, por tanto, con el grado de especialización). No obstante, la reformulación por reducción no permite discriminar entre clases (este fenómeno se repite en el subcorpus de textos médicos). Además, de forma específica, ha sido posible identificar que los dos tipos textuales con mayor IDT b) e IDT c) (blog y noticia) son tipos textuales muy poco reformulados, lo que implicaría que el emisor suele reconocer cierto conocimiento previo sobre el tema en el receptor o considera que la terminología empleada no es opaca al receptor, aun sin contar con conocimientos sobre el tema.

En definitiva, teniendo en cuenta la información que relacionamos previamente y las conclusiones del estudio de Rodríguez-Tapia (2020) sobre el discurso médico, el discurso jurídico se percibe de forma mucho más compleja, al no existir elemento definitivo discriminante (en la medicina siempre es el IDT, pero en el derecho no se identifica claramente un criterio homogéneo, puesto que también se usan la variedad léxica y la reformulación). Asimismo, la distribución de clases por cada tipo textual (véase Gráfico 1) también es diferente, puesto que en el derecho no se muestra un comportamiento tan regular como en otros campos, sino indefinido y homogéneo.

El discurso jurídico también plantea un patrón de diversidad léxica según el grado de especialización irregular si se compara con el patrón del discurso médico, cuyos tipos textuales se relacionan de forma directa con el grado de especialización y de forma inversa con la diversidad léxica (Rodríguez-Tapia, 2020: 83). En cuanto al léxico, el hecho de que los tipos textuales de blog y noticia cuenten con valores tan elevados de densidad terminológica pero tan reducidos en reformulación, tiene implicaciones en la redacción de textos, dado que deja entrever que la naturaleza de la terminología del derecho pueda diferir de la de otros campos, lo cual es significativo de cara a futuros trabajos.

También debe tenerse en cuenta que los resultados aquí planteados no distinguen los grupos de informantes que analizaron los textos. Contrastar los datos que aportan cuatro grupos diferentes según su percepción es una línea de trabajo futura, al existir la posibilidad de que cierto tipo de usuarios identifiquen un tipo de terminología o de clases que otros no determinarían. Por ejemplo, hemos comprobado al comparar el Gráfico 3 y el Gráfico 1 que la participación en foro suele considerarse un TNE, lo que parece constituir una excepción de la variedad léxica y su relación con el grado de especialización (puede que esto se deba a ciertos grupos de usuarios y no que se defina como un patrón generalizado).

Debe recordarse que analizamos un corpus reducido porque este trabajo pretende constituirse como proyecto piloto de un estudio más amplio, para determinar la viabilidad de la herramienta de análisis y prever posibles problemas por la complejidad del objeto (lo cual queda de manifiesto al no descubrir ciertos factores que sí fue posible reconocer en el discurso médico empleando la misma herramienta). Los resultados sugieren que la información que proporcionan los datos extraídos del corpus permite validar que no existen problemas en la concepción del objeto de análisis ni en la representatividad del corpus. De esta forma, cabe defender la utilidad de este marco teórico y metodológico para analizar el grado de especialización del discurso jurídico, así como para identificar elementos que revelan la complejidad del discurso del derecho en aspectos como las relaciones discursivas o el léxico. En futuros trabajos, consideramos esencial seguir avanzando en perfilar estos elementos. En efecto, describir, en primera instancia, y explicar, en segunda, elementos característicos de cada una de las tres clases del grado de especialización en el discurso jurídico permitirá a labores como la redacción y revisión de textos, la traducción, la enseñanza de lenguas para fines específicos o la lexicografía y terminografía, entre otras, contar con modelos de análisis útiles para prever problemas teóricos y metodológicos.

Referencias bibliográficas

- Alcaraz Varó, Enrique, Hughes, Brian y Campos Pardillos, Miguel Ángel. (2012). *Diccionario de términos jurídicos: inglés-español. A dictionary of legal terms: Spanish-English*. Barcelona: Ariel.
- Alcaraz Varó, Enrique, Hughes, Brian, y Gómez, Adelina. (2016). *El español jurídico*. Barcelona: Ariel.
- Alcaraz Varó, Enrique. (2007). *El inglés jurídico: Textos y documentos*. Barcelona: Ariel.
- Bernárdez, Enrique. (1982). *Introducción a la lingüística del texto*. Madrid: Espasa-Calpe.
- Bernárdez, Enrique. (1995). *Teoría y epistemología del texto*. Madrid: Cátedra.
- Cabré, María Teresa y Estopà, Rosa (2005). Unidades de conocimiento especializado: caracterización y tipología. En María Teresa Cabré y Carme Bach (eds.), *Coneixement, llenguatge i discurs especialitzat* (p. 69-93). Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- Cabré, María Teresa, Da Cunha, Iria, San Juan, Eric *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* (p. 223-241). Berlin: Springer.
- Cabré, María Teresa. (1998). Elementos para una teoría de la terminología: hacia un paradigma alternativo. *El Lenguaraz. Revista Académica del Colegio de Traductores Públicos de la Ciudad de Buenos Aires*, I(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). 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* (p. 87-105). Castelló de la Plana: Publicacions de la Universitat Jaume I.
- 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* (p. 89-106). Arras: Artois Presses Université.
- Campo Moreno, Juan Carlos. (2019). [El futuro de la modernización del lenguaje jurídico](#). *Revista del Ministerio Fiscal*, 8, 58-71.
- Campos Martín, Natalia María y Campos Plaza, Nicolás. (2016). El lenguaje jurídico y administrativo español y francés aplicado al ámbito de la extranjería. Características terminológicas y discursivas. En María Mercedes Eurrutia Cavero (coord.), *El lenguaje jurídico y administrativo en el ámbito de la extranjería: Estudio multilingüe e implicaciones socioculturales* (p. 195-235). Bern: Peter Lang.
- Campos Pardillos, Miguel Ángel. (2007). El lenguaje de las ciencias jurídicas nuevos retos y nuevas visiones. En Enrique Alcaraz Varó, José Mateo Martínez y Francisco Yus Ramos (coords.), *Las lenguas profesionales y académicas* (p. 155-166). Barcelona: Ariel.
- Caravedo, Rocío. (2014). *Percepción y variación lingüística: enfoque sociocognitivo*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- Carretera González, Cristina y Fuentes Gómez, Julio Carlos. (2019). [La claridad del lenguaje jurídico](#). *Revista del Ministerio Fiscal*, 8, 7-40.
- Cobos López, Ingrid. (2012). [Análisis contrastivo del discurso jurídico español-alemán](#). *Alfinge: Revista de filología*, 24, 31-54.
- Da Cunha, Iria y Escobar, M.^a Ángeles. (2021). [Recomendaciones sobre lenguaje claro en español en el ámbito jurídico-administrativo: análisis y clasificación](#). *Pragmalingüística*, 29, 129-148.
- Da Cunha, Iria, Cabré, María Teresa, San Juan, Eric *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)* (p. 266-276). Berlin: Springer.
- 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.
- Dijk, Teun A. van. (1992 [1983]). *La ciencia del texto*. Barcelona: Paidós.
- Garcés Gómez, María del Pilar. (2008). *La organización del discurso: marcadores de ordenación y de reformulación*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- García Pérez, Rafael. (2018). [La construcción de la terminología especializada: un ejemplo del lenguaje jurídico-penal](#). *Dicenda*, 36, 209-226.
- González Vallejo, Rubén. (2020). [Un análisis lingüístico sobre las características del lenguaje jurídico italiano y español](#). *Revista de lenguas para fines específicos*, 26(2), 58-71.
- Hoffmann, Lothar. (1998 [1987]). *Llenguatges d'especialitat: selecció de textos* (Brumme, Jenny, ed.). Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- Hourani-Martín, Dunia. (2020). [Variación y transformaciones en unidades fraseológicas jurídicas: la importancia de su representación en herramientas para traductores](#). *Círculo de Lingüística Aplicada a la Comunicación*, 82, 55-68.
- 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* (p. 155-176). Barcelona: Ariel.
- Martí Sánchez, Manuel. (2004). [La compleja identidad del léxico jurídico](#). *Estudios de Lingüística. Universidad de Alicante*, 18, 169-189.
- Montolío Durán, Estrella. (2019). [La clarificación del discurso judicial. Estado de la cuestión y propuestas lingüísticas de optimización](#). *Revista del Ministerio Fiscal*, 8, 72-95.
- Nazar, Rogelio y Cabré, María Teresa. (2011). [Un experimento de extracción de terminología utilizando algoritmos estadísticos supervisados](#). *Debate Terminológico*, 7, pp. 36-55.
- 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.
- Rodríguez-Tapia, Sergio. (2020). [El grado de especialización textual del discurso médico en español: análisis basado en corpus desde la teoría de la percepción](#). *Panace@. Revista de medicina, lenguaje y traducción*, XXI(51), 75-92.

Tabares Plasencia, Encarnación y Batista Rodríguez, José Juan. (2013). Delimitación y análisis de las unidades fraseológicas jurídicas en español y alemán. En Carmen Mellado Blanco (coord.), *La fraseología del alemán y el español: lexicografía y traducción* (p. 75-87). München: Peniöpe.

Templin, Mildred C. (1957). *Certain language skills in children: Their development and interrelationships*. Wesport: Greenwood.

Tercedor Sánchez, Maribel y Láinez Ramos-Bossini, Antonio Jesús. (2017). *La comunicación en línea en el ámbito de la psiquiatría: caracterización terminológica de los foros de pacientes*. *Panace@* 18(45), 30-41.

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.

Referencias del corpus

Ámbito jurídico (30 de septiembre de 2015): [Jorge Ernesto Villa: ¿En qué consisten los cambios al proceso de nulidad del matrimonio católico?](#). *Ámbito jurídico*.

Berenguer Albadalejo, M.^a Cristina. (2015). [Los matrimonios en forma religiosa no canónica: celebración e inscripción en el registro civil de acuerdo con las reformas operadas por la ley 15/2015, de 2 de julio, de la jurisdicción voluntaria](#). *Derecho Privado y Constitución*, 29, 83-131.

Corvinos Baseca, Pedro (4 de mayo de 2017). [Comercializadoras municipales como medio propio para la adquisición de electricidad](#) [entrada de blog]. *Pedro Corvinos Abogado*.

González, Nuria y Lafuente, Eliseo (18 de febrero de 2017): [Eliseo Lafuente, presidente Asociación de abogados jóvenes de Soria, colaborador y fundador de Lafuente asociados](#). *Entrevista a Eliseo Lafuente para el periódico digital* [entrada de blog]. *Lafuente asociados*.

Gonzalvo, Irene. (11 de febrero de 2020): [La alta cualificación profesional de un prestatario no supe el deber de información del banco según el Tribunal Supremo](#). *Diario Jurídico*.

josias (25 de enero 01 de 2015): [renovar tarjeta comunitaria](#) [entrada de foro en línea]. *Pórtico legal*.

leonjbr (14 de 05 de 2017): [¿Puedo iniciar la denuncia de desahucio?](#) [entrada de foro en línea]. *Pórtico legal*.

Martínez Garay, Lucía. (2016). [Errores conceptuales en la estimación de riesgo de reincidencia. La importancia de diferenciar sensibilidad y valor predictivo, y estimaciones de riesgo absolutas y relativas](#). *Revista Española de Investigación Criminológica*, 3(14), 1-31.

Poder Judicial España (9 de mayo de 2017): [El Tribunal Supremo establece que el delito de ‘stalking’ requiere continuidad en el tiempo y alterar los hábitos de la víctima](#). *Poder judicial España*.

Rosal, Pedro del (11 de mayo de 2017): [Los contratos de los empleados públicos temporales estarán limitados a un máximo de tres años](#). *El Economista*.

Ruiz Rodríguez, M.^a Adoración (12 de mayo 05 de 2017): [Artículo 384 Código Penal: delito o sanción administrativa](#). *Noticias jurídicas*.