Facultad de Ciencias Económicas y Empresariales

Aplicación de técnicas de Text Mining para evaluar el impacto de las reformas legales en las sentencias del Tribunal Constitucional

Autor: Inés Vinuesa Armada

Director: Lucía Barcos Redín

MADRID | Abril 2024

Resumen

Este trabajo analiza el impacto de las reformas del requisito de especial trascendencia constitucional sobre los temas tratados en las sentencias del Tribunal Constitucional. A través de una metodología empírica basada en técnicas de *Text Mining*, y concretamente en el algoritmo *Latent Dirichlet Allocation* (LDA), se realiza un análisis comparado de tres conjuntos de sentencias obtenidas mediante *web scraping*. El estudio muestra cómo el análisis textual avanzado permite evaluar los efectos prácticos de las reformas e identificar posibles orientaciones estratégicas del Tribunal, con implicaciones para la protección de los derechos fundamentales.

Palabras clave: Topic Modeling, LDA, Tribunal Constitucional, Sentencias, Especial Trascendencia Constitucional, Web scraping, derechos fundamentales.

Abstract

This study analyzes the impact of the reforms to the requirement of *special constitutional significance* on the topics addressed in the rulings of the Spanish Constitutional Court. Using an empirical methodology based on *Text Mining* techniques, specifically the *Latent Dirichlet Allocation* (LDA) algorithm, a comparative analysis is carried out on three sets of rulings obtained through *web scraping*. The study demonstrates how advanced textual analysis enables the evaluation of the practical effects of the reforms and the identification of possible strategic orientations of the Court, with implications for the protection of fundamental rights.

Keywords: Topic Modeling, LDA, Constitutional Court, Sentences, Special Constitutional Significance, Web Scraping, Fundamental Rights.

Índice

Índice de figuras

Índice de tablas

1. INTRODUCCIÓN

1.1. Justificación y motivación del tema

En el panorama jurídico actual, el debate doctrinal en torno al requisito de "especial trascendencia constitucional" cobra una relevancia creciente. Este concepto, introducido en nuestro ordenamiento por la Ley Orgánica 6/2007, que modificó la Ley Orgánica 2/1979 del Tribunal Constitucional, tuvo como objetivo, entre otros, reducir el volumen de recursos de amparo pendientes de admisión (González, 2018). Esta reforma implicó una serie de decisiones de política legislativa orientadas a optimizar el funcionamiento del Tribunal Constitucional.

Esta primera reforma ha sido objeto de críticas. El magistrado Eugeni Gay Montalvo, en un voto particular emitido en un auto de interpretación del requisito, destacó que la finalidad principal del recurso de amparo es la protección de los derechos fundamentales. En su argumentación, expresó su desacuerdo con la nueva exigencia impuesta a los recurrentes, quienes deben justificar la especial trascendencia constitucional de su demanda sin posibilidad de subsanación en caso de defecto formal (Matia, 2009).

En este sentido, la eficacia de dichas medidas ha sido cuestionada y nuevas restricciones fueron introducidas mediante un Acuerdo publicado en el BOE el 23 de marzo de 2023, lo que ha intensificado el debate sobre el acceso al recurso de amparo (Rodríguez, 2023). Parte de la doctrina sostiene que los requisitos actuales vulneran derechos fundamentales al restringir de manera excesiva la capacidad de los ciudadanos de recurrir en amparo (Valiente, 2024).

Además, la falta de motivación concreta en las providencias de inadmisión, conocida como providencia no motivada, ha generado críticas al considerar que el Tribunal podría estar seleccionando los casos de acuerdo a intereses propios, más que bajo criterios objetivos de relevancia constitucional.

Frente a este contexto jurídico y doctrinal, resulta imperioso un análisis empírico sobre la aplicación práctica de los criterios establecidos en las reformas de 2007 y 2023. El uso de técnicas de text mining se presenta como una herramienta innovadora para analizar la jurisprudencia del Tribunal Constitucional, permitiendo identificar si los casos admitidos tienden a concentrarse en ciertos temas o si se observa una tendencia al aumento o disminución de los recursos de amparo estimados. En particular, el topic modeling, y concretamente el algoritmo Latent Dirichlet Allocation (LDA), permite detectar de forma automatizada los temas más recurrentes en un gran volumen de resoluciones, facilitando así el estudio de patrones temáticos y su evolución a lo largo del tiempo.

En definitiva, un estudio basado en la jurisprudencia del Tribunal Constitucional, utilizando enfoques innovadores y técnicas de topic modeling, permitirá no solo evaluar con mayor precisión el impacto real de las reformas en el acceso al recurso de amparo, sino también ofrecer propuestas de mejora orientadas a la protección efectiva de los derechos fundamentales en nuestro sistema jurídico.

1.2. Objetivo general y específicos

Este trabajo combina el análisis jurídico con herramientas de Text Mining, con el objetivo de ofrecer una nueva aportación a la estadística jurisdiccional en el ámbito del Derecho Constitucional. En particular, se centra en estudiar el impacto que ha tenido la introducción y evolución del requisito de especial trascendencia constitucional en la admisión de recursos de amparo ante el Tribunal Constitucional.

El objetivo general del estudio es determinar si las reformas que han afectado a dicho requisito han provocado cambios relevantes en las materias tratadas por las sentencias del Tribunal. Para ello, se aborda un análisis estructurado de los cambios observados en distintos niveles.

En primer lugar, se examina la eficiencia del Tribunal en términos cuantitativos, es decir, si las modificaciones normativas han influido en el número de sentencias dictadas. Este análisis permite valorar si el nuevo sistema ha contribuido a agilizar la carga de trabajo del Tribunal o, por el contrario, ha derivado en una reducción de su actividad resolutiva.

En segundo lugar, se analiza el contenido material de las sentencias, con especial atención a los temas jurídicos abordados. Se pretende identificar si el Tribunal mantiene una línea estable en los derechos objeto de análisis o si, por el contrario, utiliza el nuevo marco normativo para concentrarse en determinados temas, posiblemente más complejos, innovadores o socialmente relevantes.

El objetivo final es aportar una explicación al patrón observado: si existe una constancia temática en las resoluciones, o si se aprecia una variación significativa a lo largo del tiempo. En este sentido, se reflexiona sobre si el Tribunal Constitucional está utilizando el requisito de especial trascendencia constitucional como una herramienta para seleccionar aquellos asuntos que generan mayor controversia, impacto público o acumulación de recursos, o si, por el contrario, prefiere centrarse en materias más específicas. Esta aproximación permitirá comprender no solo los efectos prácticos de la reforma, sino también las posibles orientaciones estratégicas que adopta el Tribunal en el ejercicio de su función jurisdiccional.

1.3. Metodología

Para alcanzar los objetivos definidos se ha diseñado una metodología comparativa basada en el análisis de tres bases de datos distintas, construidas específicamente para reflejar el impacto del requisito de especial trascendencia constitucional en distintos contextos normativos.

Previamente al análisis empírico, se ha realizado un estudio del contexto jurídico que enmarca el caso de análisis, revisando la evolución normativa y jurisprudencial del recurso de amparo

y del propio requisito de admisión, con base en legislación, doctrina especializada y resoluciones del Tribunal Constitucional.

Asimismo, se ha llevado a cabo una revisión del marco conceptual relativo al text mining, prestando especial atención al topic modeling como técnica de referencia para la identificación de patrones temáticos en grandes volúmenes de texto y en particular al LDA.

La primera base de datos recogerá el conjunto de sentencias dictadas por el Tribunal Constitucional en el año 2005, representando así la situación previa a la introducción del citado requisito. La segunda base corresponde al año 2009, permitiendo analizar el escenario una vez que la reforma de 2007 ya había sido incorporada y comenzaban a observarse sus primeros efectos interpretativos. Finalmente, la tercera base de datos se centrará en las sentencias emitidas en el año 2024, con el fin de valorar los posibles efectos derivados de la última reforma, aprobada en 2023.

La recopilación de datos se ha llevado a cabo mediante técnicas de webscraping, accediendo a los enlaces disponibles en la página oficial del Tribunal Constitucional y extrayendo el contenido completo de las sentencias correspondientes a cada uno de los tres periodos seleccionados.

Una vez recopiladas, las sentencias se someten a un proceso de preprocesamiento textual para posteriormente calcular la coherencia temática con el objetivo de determinar el número óptimo de temas a extraer en cada caso.

Finalmente, se aplicará el algoritmo de modelado de temas LDA, y se realiza una comparativa que permita extraer conclusiones sobre la evolución de los temas tratados por el Tribunal Constitucional.

1.4. Estructura del trabajo

A partir de aquí, el trabajo se estructura como sigue. El capítulo 2 ofrece un análisis del funcionamiento del Tribunal Constitucional, con especial atención al régimen jurídico y características de los recursos de amparo. El capítulo 3 presenta los fundamentos técnicos del Text Mining y del topic modeling, precedidos por una breve revisión de la literatura académica sobre legal tech. A continuación, el capítulo 4 expone la metodología empleada en el estudio, y el capítulo 5 recoge y analiza los principales resultados obtenidos. Finalmente, en el capítulo 6, se extraen las conclusiones del trabajo y se apuntan posibles líneas futuras de investigación.

2. CUESTIONES LEGALES PREVIAS

2.1. El Tribunal Constitucional Español y sus sentencias

El Tribunal Constitucional es una institución clave en el sistema jurídico español, encargada de garantizar la primacía de la Constitución (CE) y el respeto al Estado de Derecho (García-Pelayo, 2014). A diferencia del resto de tribunales, no forma parte del Poder Judicial, sino que actúa como un órgano independiente con competencias específicas en materia de control de las normas. Su regulación se encuentra en la Ley Orgánica 2/1979, de 3 de octubre, del Tribunal Constitucional (en adelante, LOTC).

Entre sus atribuciones se encuentra el control previo de determinadas normas, como los tratados internacionales y algunas leyes orgánicas. Además, realiza un control posterior de la adecuación de las leyes a la Constitución a través de los recursos de inconstitucionalidad. En este sentido, su labor puede tener efectos legislativos indirectos, dado que sus resoluciones pueden suponer la expulsión de una norma aprobada por las Cortes del ordenamiento jurídico (Rubio Llorente, 1988). Por ejemplo, el Tribunal Constitucional puede enjuiciar la falta de competencia del Gobierno de regular materias reservadas a la ley por medio del Real-Decreto. Esta norma carece de formalidades legales al emanar del poder ejecutivo y no de las Cortes Generales.

Además, el Tribunal Constitucional se encarga de la protección de los derechos y libertades fundamentales recogidos en la Constitución (García-Pelayo, 2014). El derecho más tratado en las sentencias es el derecho de los individuos a defender sus intereses ante los tribunales, esto es, el derecho a la tutela judicial efectiva (art. 24 CE) en sus diferentes expresiones (Tribunal Constitucional, 2025). Este derecho abarca, entre otros, el derecho a la no indefensión, el derecho a un proceso con todas las garantías, el derecho a utilizar los medios de prueba pertenecientes o el derecho a la presunción de inocencia.

El Tribunal Constitucional tiene una gran cantidad de casos pendientes, lo que hace que resolverlos tome mucho tiempo. En el caso de los recursos de amparo, este retraso puede ser de entre tres y cinco años, lo que dificulta la protección rápida y efectiva de los derechos fundamentales. Como resultado, la gran mayoría de estos recursos ni siquiera llegan a ser admitidos a trámite (Pérez Tremps, 2004).

Además, el sistema para proteger estos derechos tiene sus propias limitaciones. Para que el Tribunal Constitucional intervenga, primero es necesario haber agotado todas las opciones ante los tribunales ordinarios. Sin embargo, los retrasos en ambas instancias pueden hacer que esta protección llegue demasiado tarde, poniendo en riesgo su efectividad (García de Enterría, 2014).

De ello se desprende la relevancia del Tribunal Constitucional para la ciudadanía, ya que su funcionamiento no solo contribuye al desarrollo de la sociedad, sino que sus sentencias desempeñan un papel fundamental tanto en la evaluación de la eficacia de determinadas medidas como en la identificación de los problemas más recurrentes en materia de derechos fundamentales.

2.2. Los recursos de amparo

Desde 1994, algunos expertos (Diez-Picazo, 1994) ya advertían que el Tribunal Constitucional rechazaba demasiados recursos de amparo usando un motivo de inadmisión muy general, establecido en el artículo 50.1.c) de la LOTC, que hoy ya no está vigente. En esos casos, el Tribunal argumentaba que las demandas no tenían un contenido lo suficientemente relevante como para justificar una decisión, sin explicar en detalle por qué llegaba a esa conclusión. Estas demandas no recibían una sentencia que resolviera el proceso, sino que eran directamente desestimadas.

En este contexto, el número de recursos de amparo presentados ante el Tribunal Constitucional aumentaba exponencialmente. Para hacer frente a esta situación, la Ley Orgánica 6/1988 incorporó un nuevo motivo para rechazar estos recursos: si ya existía una sentencia previa del Tribunal sobre un caso similar y que hubiera sido desestimada, el nuevo recurso no se admitiría. Ante esta reforma, algunos expertos señalaron la importancia de analizar con datos concretos cómo el Tribunal conoce sus precedentes y desestima demandas basándose en ellos (Diez-Picazo, 1994).

Posteriormente, la Ley Orgánica 6/2007, del 24 de mayo, modificó el proceso de admisión de los recursos de amparo. A partir de esta reforma, el artículo 50 de la LOTC establece dos motivos para rechazar un recurso: errores en el procedimiento y la falta de especial trascendencia constitucional, que es el único motivo de inadmisión relacionado con el fondo del asunto (Cabañas García, 2010).

Este nuevo requisito implica que quien presenta un recurso de amparo debe explicar por qué su caso tiene una especial trascendencia constitucional. Sobre este punto, el Auto 252/2009 del Tribunal Constitucional aclara que no basta con decir que el caso la tiene o darlo por hecho.

Por su parte, la interpretación del significado de la especial trascendencia constitucional vino fundamentalmente en una sentencia del año 2009, la STC 155/2009, de 25 de junio. El Tribunal considera imperativo avanzar en la interpretación del concepto y establece una serie de supuestos, numerus apertus, en los que concurre la especial trascendencia constitucional.

Entre esos supuestos se encuentran algunos como que la demanda trate un problema sobre un derecho fundamental que aún no ha sido analizado, o si es necesario aclarar o cambiar el criterio establecido en sentencias anteriores debido a nuevos contextos sociales o cambios en las leyes. También se admite si la vulneración del derecho proviene de una ley, de una interpretación

repetida de los tribunales que afecta negativamente al derecho o si hay fallos judiciales contradictorios sobre el mismo tema (Tribunal Constitucional, 2009).

El Tribunal Europeo de Derechos Humanos dictaminó que la reforma mencionada no se ajustaba al derecho europeo en el caso Arribas Antón vs. España. Además, advirtió al Tribunal Constitucional de que no estaba aplicando correctamente su propia doctrina, lo que afectaba al principio de buena administración de justicia (Hernández Ramos, 2016).

Finalmente, en 2023, el Tribunal Constitucional seguía siendo consciente de que muchos recursos de amparo eran rechazados y no llegaban a obtener una sentencia firme. A pesar de las medidas tomadas no había una disminución significativa en el número de demandas presentadas ante el Tribunal.

Por ello, aprobó el Acuerdo de 15 de marzo de 2023, que regula la presentación de estos recursos a través de su sede electrónica. Según los expertos, una de las principales novedades es la obligación de completar un formulario donde se explique, entre otros aspectos, por qué el recurso tiene especial trascendencia constitucional. Además, se establecen ciertas reglas de formato, como un límite de 50.000 caracteres, el uso de Times New Roman en tamaño 12 y un interlineado de 1,5 (Torres Díaz, 2023).

2.2.1. Estructura

Todas las sentencias del Tribunal Constitucional siguen una estructura uniforme. A continuación, se muestra la estructura de las sentencias para después extraer una parte concreta de las mismas y formar la base de datos con la que se aplicarán los algoritmos de machine learning.

Como ejemplo, se toma la Sentencia del Tribunal Constitucional (STC) 107/1984, de 23 de noviembre, conocida como el caso del Conserje Uruguayo (López Guerra, 2008). Esta

resolución fue una de las primeras en reconocer los derechos fundamentales de los extranjeros en España.

En primer lugar, la sentencia incluye un contexto introductorio que identifica la Sala que resuelve el recurso y la composición de sus magistrados. En este caso, la resolución fue dictada por la Sala Segunda, integrada por el presidente, don Jerónimo Arozamena Sierra, y los magistrados don Francisco Rubio Llorente, don Luis Díez-Picazo y Ponce de León, don Francisco Tomás y Valiente, don Antonio Truyol Serra y don Francisco Pera Verdaguer.

A continuación, se mencionan las partes del proceso, en este caso, el demandante, don Leonardo Leyes Rosano, junto con su representación legal. También se especifican las resoluciones impugnadas, que en este supuesto corresponden a la Sentencia de la Magistratura de Trabajo núm. 3 de Barcelona, dictada en el procedimiento 533/1982, en relación con una cuestión de naturaleza laboral.

Seguidamente, la sentencia desarrolla los antecedentes del caso bajo el epígrafe "I. Antecedentes". En esta sección, se expone un resumen de los hechos relevantes que motivaron la interposición y admisión del recurso de amparo. Se detalla el paso por los diferentes tribunales y las fechas en las que esto tuvo lugar.

Posteriormente, en la sección II. Fundamentos Jurídicos, se realiza un análisis normativo del caso, aplicando e interpretando las disposiciones legales pertinentes. En el caso de los recursos de amparo, se comienza con la referencia al artículo constitucional que consagra el derecho fundamental invocado.

Finalmente, la sentencia concluye con el Fallo, que recoge la decisión del Tribunal. En el caso de las sentencias, esta decisión solo puede consistir en la estimación o desestimación del recurso. En este supuesto, la Sala Segunda decidió desestimar el recurso de amparo. Como es preceptivo, la resolución se ordena publicar en el Boletín Oficial del Estado (BOE).

3. DERECHO Y TEXT MINING

3.1. Legal Tech en la actualidad

Legal Tech hace referencia a la aplicación de software y tecnología para mejorar los servicios legales. Esta idea surge frente un constante paradigma que sitúa a los abogados y al resto de profesiones jurídicas fuera de los ámbitos de innovación (Corrales, Fenwick y Haapio, 2019).

Según Bues y Matthaei (2017) dentro de Legal Tech pueden identificarse tres categorías:

La primera de ellas consiste en incorporar tecnologías para la gestión documental en la nube (por ejemplo, la base de datos Net Documents) junto con herramientas de ciberseguridad.

La segunda consiste en adoptar tecnologías que permitan la realización de tareas jurídicas. Desde la búsqueda de doctrina o jurisprudencia hasta la redacción de un contrato se puede aplicar IA y otros sistemas de machine learning.

La tercera, en la que podría enmarcarse el NLP (Natural Lenguaje Processing), permite complementar las habilidades del abogado con los conocimientos en programación. Dentro de esta categoría, se encuentran aquellas tecnologías que facilitan la labor del abogado y promueven la efectividad en el trabajo realizado. Existen diferentes campos como la automatización de contratos, que se realiza mediante aplicaciones como Contract Express o Document Drafter que combinan código con conocimiento jurídico, el blockchain y los Smart contracts.

En términos generales, la tecnología blockchain se define como un sistema tecnológico que permite a un conjunto de usuarios llevar a cabo acciones que son verificadas y registradas de manera inmutable en una red de servidores interconectados, denominados nodos. Estos nodos operan bajo el principio de registros distribuidos, conocido como Distributed Ledger Technology (DLT) (Finck, 2019).

Esta tecnología se ha utilizado en ámbito jurídico para votar en Juntas Generales de Accionistas (Santander S.A. y Broadridge Inc., 2018). Asimismo, la tecnología blockchain permite cumplir con normativas relativas al derecho a la privacidad, como el Reglamento general de protección de datos, e implementar contratos con ejecución automática o smart contracts (Kulhari, 2018).

Los expertos defienden ampliamente la aplicación de las nuevas tecnologías en el campo jurídico (Dale, 2018). Particularmente, el NLP no es una excepción. El derecho es una ciencia basada en el lenguaje, por lo que no debe sorprender que aquellos algoritmos basados en el lenguaje tengan una presencia cada vez mayor en los despachos.

Aunque el NLP no es ninguna novedad, hemos visto un crecimiento exponencial del número de empresas de nuevo crecimiento que dicen aplicar técnicas de Deep learning en el ámbito jurídico, es decir, técnicas que recurren a varias capas de procesamiento para extraer información (Shinde y Shah, 2018). También hemos visto un crecimiento en las áreas en las que el text mining y concretamente el NLP, pueden aportar en la búsqueda de información legal relevante para nuevos casos (Dale, 2018).

Concretamente, Mielnik y Altszyler (2023) destacan la capacidad del LDA para completar en poco tiempo tareas como la identificación de los temas presentes en cientos de miles de fallos judiciales de Argentina, que a los seres humanos nos requerirían varios años.

Otro estudio que ilustra la aplicación del LDA se centra en analizar 4.356 sentencias judiciales sobre casos narcóticos de un Tribunal de Indonesia (Sari, Kosasih e Indarti, 2023). Este trabajo muestra que el topic modeling permite no solo detectar patrones en el contenido de las sentencias, sino también establecer relaciones útiles para entender tendencias judiciales y estructurar decisiones jurídicas complejas.

Sin embargo, la literatura sigue siendo limitada en comparación con otras áreas del análisis de textos. Existen pocos estudios centrados específicamente en sentencias judiciales, lo que sugiere una oportunidad relevante para la investigación.

3.2. Técnicas avanzadas de análisis de texto para el estudio de sentencias

El aprendizaje automático o machine learning consiste en un conjunto de algoritmos diseñados para identificar patrones en los datos ajustando sus hiperparámetros y mejorando su desempeño a partir de la información con la que son entrenados (Mielnik, 2022).

Dentro del machine learning, el aprendizaje no supervisado se emplea para analizar datos no etiquetados, con el objetivo de agrupar observaciones en función de su similitud sin respuestas previamente definidas (Taboada Villamarín, 2024). En esta categoría se enmarca el topic modeling, una técnica de text mining que será aplicada en el presente estudio.

3.2.1. Text Mining

El crecimiento exponencial de los datos digitalizados en plataformas online—desde noticias y publicaciones en redes sociales hasta reseñas de clientes, artículos científicos y comunicados de prensa—ha generado un gran volumen de datos no estructurados. En este contexto, la minería de texto o text mining ha emergido como una herramienta clave para el análisis de estos datos, permitiendo a los investigadores extraer conocimiento y desarrollar teorías. Su aplicación no solo facilita la identificación de patrones y tendencias, sino que también está cobrando cada vez más relevancia en la generación de nuevos marcos teóricos y en la comprensión de fenómenos complejos. Por ejemplo, Kumar (2021), ilustra la utilidad de estos algoritmos para encontrar las palabras clave de los artículos de la base de datos de Scopus.

Cada documento está compuesto por tokens, que hacen referencia a cada una de las palabras del documento (Sakthi Vel, 2021). A su vez, dentro de un documento podemos encontrar

enunciados, párrafos, espacios y caracteres. Estos últimos pueden definirse como la unidad más básica de texto que representa cada símbolo o letra de un token. Los documentos se unen formando un corpus.

Para implementar modelos analíticos en grandes corpus de texto es imprescindible realizar un preprocesamiento de los datos para reducir la presencia de ruido y resaltar los patrones y características más relevantes del corpus (Hernández y Rodríguez, 2008).

Este proceso se desarrolla en varias fases:

Figura 1. Fases del preprocesamiento. Fuente: Elaboración Propia.

En primer lugar, la construcción de la base de datos puede llevarse a cabo de diferentes formas en relación con el tipo de acceso (público, privado o semipúblico).

Una Interfaz de Programación de Aplicaciones (API) es un método de acceso a software, generalmente empleado en entornos semipúblicos que permite a usuarios externos extraer información específica. Por otro lado, la práctica conocida como web scraping consiste en la

extracción de datos de páginas web en formato HTML mediante código, sin necesidad de una solicitud de acceso, lo que facilita su aplicación en diversos sitios web. Finalmente, también es posible obtener un corpus a través de la descarga directa de archivos (Taboada Villamarín, 2024).

Una vez recopilados los textos, el siguiente paso es la tokenización, un procedimiento que convierte el texto legible por humanos en unidades procesables por máquina. La forma más común de tokenización es la segmentación en palabras (Gil Pascual, 2021). También se convierten todas las letras en minúsculas para evitar distinciones irrelevantes entre los tokens.

Posteriormente, se lleva a cabo un proceso de filtrado, cuya finalidad es eliminar términos irrelevantes o palabras vacías (stopwords) con el objetivo de reducir el ruido en el análisis de texto y centrarse en los términos más significativos (Alaminos Fernández, 2023).

La siguiente fase consiste en la aplicación de lematización, aunque también puede recurrirse al stemming. Por su parte, la lematización busca identificar la forma base de cada palabra obtenida en la tokenización. Comparte ciertas similitudes con el stemming pues esta técnica se caracteriza por eliminar los sufijos para encontrar la raíz o lexema de las palabras. Sin embargo, para obtener el lema de una palabra se emplean diccionarios lingüísticos para reemplazar cada término por su forma canónica, lo que permite un análisis del lenguaje más preciso (Gil Pascual, 2021).

Después, la Figura 1 hace referencia a la identificación de colocaciones mediante el uso de n-gramas, que consisten en secuencias continuas de n caracteres o palabras extraídas de un texto. Este método se presenta como una forma de paliar los efectos generados por el Bag of Words (BoW) en el que el orden de las palabras no se tiene en cuenta a la hora de analizar los datos, siendo irrelevante tanto la asociación semántica como la gramática (Grootendorst, 2022).

Los n-gramas permiten capturar relaciones contextuales dentro del texto. Su importancia radica en que la combinación de dos o más palabras puede tener un significado distinto al de cada palabra por separado.

Finalmente, se construye la matriz DTM, cuya creación es fundamental dentro de un enfoque de BoW. En este proceso, cada texto se representa como un vector de palabras, y la matriz DTM ofrece información sobre la frecuencia de aparición de cada término (Gurcan y Cagiltay, 2019).

La métrica Term Frequency-Inverse Document Frequency (TF-IDF) es un método de ponderación cuyo cálculo se basa en dos componentes: la frecuencia de aparición de la palabra en un documento (Term Frequency, TF) y la importancia de la palabra en todo el corpus, medida a través de su rareza en los documentos (Inverse Document Frequency, IDF), lo que permite dar mayor peso a los términos distintivos y reducir la influencia de palabras demasiado frecuentes (Gil Pascual, 2021).

De este modo, cada documento se representa como un vector en el que los elementos reflejan la importancia relativa de los tokens que lo componen.

La fórmula para calcular esta métrica de forma simplificada consiste en:

$$TF.IDF_{(t,d)=} TF * IDF$$

$$\text{TF (t)} = \frac{\text{Number of occurrences of term t in a document}}{\text{All terms contained within the document}}$$

$$\text{IDF(t)} = \frac{\text{Total number of documents}}{\text{Number of documents with term t}}$$

Figura 2: Fórmulas de la métrica TF-IDF. Fuente: Jain, S., Jain, S. K., y Vasal, S. (2024). "An Effective TF-IDF Model to Improve the Text Classification Performance". *IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*.

Además, para calcular el peso de una palabra se suele suavizar el cálculo de la IDF mediante un logaritmo. Existen muchas variantes del cálculo de TF-IDF mostrado en la Figura 2, entre los que encontramos el TF-IDF Mejorado, el TF-IDF-CF o el R-TF-IDF-CF, que incluyen variaciones en la fórmula para mejorar su rendimiento (Jain et. al, 2024).

En definitiva, el uso de técnicas de preprocesamiento es fundamental para la correcta aplicación de algoritmos de text mining, ya que permite estructurar la información de manera eficiente y mejorar la calidad del análisis.

3.2.2. Topic Modeling

El modelado de temas o topic modeling es una técnica de aprendizaje no supervisado de machine learning que consiste en descubrir temas ocultos dentro de una colección de documentos. En este trabajo, el topic modeling se ha llevado a cabo haciendo uso de LDA. Se trata de una de las herramientas más poderosas para explorar la estructura temática de un dataset, sin previa información sobre el mismo (Wasiq et al.,2024).

También existen otros algoritmos para el análisis de texto que persiguen el modelado de temas en documentos como Bert Topic, que supera la representación BoW (Grootendorst, 2022). En este sentido, en el artículo de Rawat, Ghildiyal y Dixit (2022) se utiliza este modelo para extraer temas relevantes de un corpus formado por sentencias relacionadas con la ley de Matrimonio Hindú.

Entre las aplicaciones del LDA se encuentra la exploración de datos, la organización documental, la obtención de información o la elaboración de algoritmos de recomendación (Abay Mersha et al, 2024).

Además, un aspecto distintivo del método LDA es que se basa en el mencionado enfoque BoW para extraer los temas más relevantes. Tal y como se explica en el apartado 3.2.1, en aras de

paliar este defecto se hace uso de colocaciones de n-gramas o términos que suelen aparecer unidos.

En este sentido, no es infrecuente afirmar que LDA no es eficiente en grandes vocabularios. Para evitar esto, se eliminan tanto las palabras más frecuentes como las menos frecuentes y se busca obtener modelos interpretables. Sin embargo, persiste el riesgo de eliminar aquellas palabras más importantes, limitando así la aplicación del modelo (Dieng et al., 2019).

Se trata de un modelo probabilístico que debe su nombre a la distribución de Dirichlet. Parte de que los documentos se presentan como combinaciones aleatorias de temas latentes, donde cada tema se caracteriza por una distribución sobre palabras (Blei et al., 2003).

El modelo parte de los siguientes hiperparámetros (Xie S. y Feng Y., 2015):

- Alfa (α): Parámetro relativo a la distribución de tópicos en cada documento
- Beta (β): Parámetro relativo a la distribución de palabras en cada tópico

Con ellos se obtienen las siguientes matrices de distribución:

- Theta (θ): Matriz de distribución de tópicos en los documentos del corpus generada a partir de la distribución de Dirichlet como parámetro α
- Phi (ϕ): Matriz de distribución de palabras por tópico generada a partir de la distribución de Dirichlet como parámetro β

La selección de hiperparámetros puede ser un proceso complejo. A medida que el valor de alfa aumenta, los tópicos tienden a distribuirse de manera más uniforme en el corpus. Por el contrario, un valor bajo de alfa implica que un menor número de tópicos domina el conjunto de datos. En cuanto a beta, un valor elevado favorece la diversidad de palabras dentro de cada tópico.

Otro de los principales desafíos en el topic modeling es la determinación del número óptimo de tópicos. Para abordar esta cuestión, en el presente trabajo se emplea el modelo de coherencia

propuesto por Mimno et al. (2011). La coherencia UMass se centra en medir la consistencia semántica de los temas mediante el análisis de la ocurrencia de las palabras, lo que permite identificar temas que carecen de sentido o son demasiado generales, sin necesidad de recurrir a evaluaciones humanas costosas o a corpus de referencia externos.

El modelo UMass utiliza la probabilidad condicional para medir la relación entre palabras dentro de un tópico sin requerir un corpus externo ni llevar a cabo una división de los datos en training y test. La suma de la coherencia UMass para cada número de tópicos tiene en cuenta el orden entre las M (M = 5, 10, 15, 20) palabras más probables (Oman M. et al, 2015).

En síntesis, la aplicación del modelo LDA permite obtener, por un lado, la distribución de palabras dentro de cada tópico y, por otro, la asignación de pesos que reflejan la importancia relativa de los tópicos en cada uno de los documentos del tópico, pero para ello es preciso acudir a las métricas de coherencia para conocer el número óptimo de tópicos.

4. DESCRIPCIÓN DE LA APLICACIÓN Y FASES DE LA METODOLOGÍA

4.1. Descripción del caso práctico y de las fases de análisis

Con el fin de comprender el impacto del requisito de especial trascendencia constitucional, este capítulo presenta un análisis del conjunto de sentencias del Tribunal Constitucional en tres momentos clave, marcados por los hitos más relevantes en la evolución normativa de los recursos de amparo.

En primer lugar, se examina un corpus de resoluciones que reflejan la situación previa a la entrada en vigor de la primera norma que introdujo dicho requisito. Como esta reforma comenzó su tramitación parlamentaria el 22 de noviembre de 2005 y entró en vigor el 25 de mayo de 2007, se han seleccionado las sentencias dictadas en 2005, con el objetivo de evitar cualquier posible influencia derivada del debate legislativo.

En segundo lugar, se analiza el impacto inicial de la reforma de 2007 mediante el estudio de las sentencias emitidas en 2009. Esta elección permite observar los primeros efectos de su aplicación una vez que el Tribunal dispuso de cierto margen para interpretar el nuevo requisito, lo cual ocurrió con la Sentencia 155/2009, de 25 de junio.

Finalmente, se estudian las resoluciones correspondientes al año 2024 con el objetivo de detectar posibles cambios derivados de la reforma de 2023, aprobada mediante Acuerdo publicado en el BOE el 23 de marzo de ese mismo año.

El propósito de este análisis es determinar, a través de los datos, si el Tribunal Constitucional está utilizando los cambios normativos para seleccionar y admitir a trámite recursos relacionados con materias concretas. Si bien uno de los objetivos fundamentales de las reformas ha sido reducir la carga de recursos de amparo, parte de la doctrina ha cuestionado si

estos cambios también han servido como instrumento para que el Tribunal seleccione aquellos asuntos que considera de mayor relevancia.

Por tanto, no se trata únicamente de evaluar si ha habido un aumento o una reducción en el número de sentencias dictadas, sino también de analizar cómo ha variado el contenido de las mismas. Las preguntas clave a las que se busca dar respuesta son: ¿Ha aumentado la eficiencia del Tribunal, resolviendo un mayor número de casos? ¿O, por el contrario, ha reducido su actividad decisoria? ¿Se mantienen los temas tratados a lo largo del tiempo o se observan variaciones significativas? Y en este último caso, ¿cuáles podrían ser las razones de dichos cambios?

4.2. Construcción de las bases de datos

Dado que el número de sentencias es voluminoso, se ha optado por la automatización del proceso de creación de la base de datos. Así, como afirman Shamshiri et. al (2024), uno de los obstáculos fundamentales en la construcción de bases de datos es la dificultad para encontrar datasets reales y actualizados. Por ello, mediante la técnica de webscraping, se extraen y organizan datos de internet de manera sistemática.

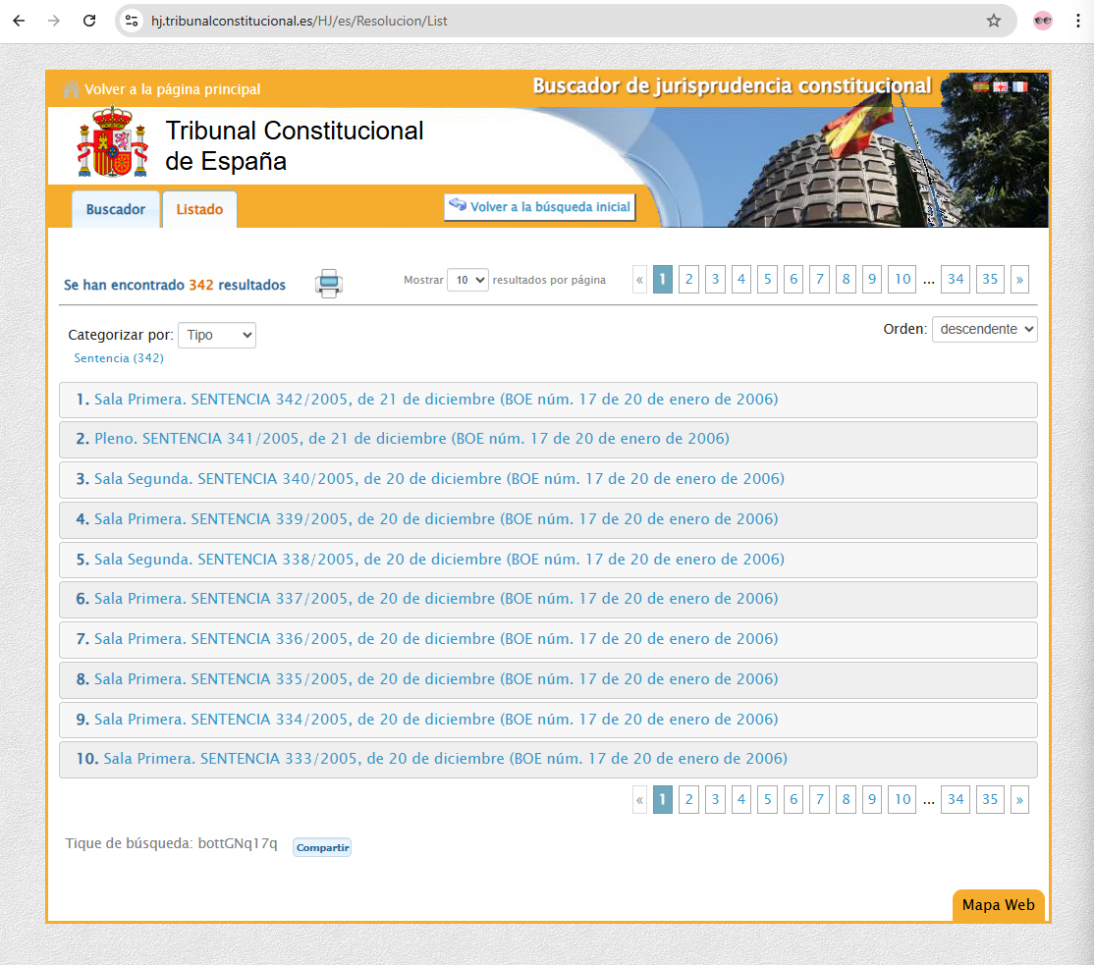

Figura 3: Página Web de la búsqueda de jurisprudencia del Tribunal Constitucional. Fuente: Elaboración propia

La Figura 3 muestra el resultado de una búsqueda de jurisprudencia en la página web del Tribunal Constitucional correspondiente al año 2005. El sistema ha encontrado un total de 342 sentencias, distribuidas en 35 páginas, mostrándose 10 resultados por página.

Los resultados están ordenados en orden descendente, por lo que la primera sentencia que aparece (Sentencia 342/2005) es la última dictada en ese año, con fecha de 21 de diciembre de 2005. A medida que se avanza por las páginas, se accede a sentencias anteriores, hasta llegar a la Sentencia 1/2005, que se encontrará al final de la última página (página 35).

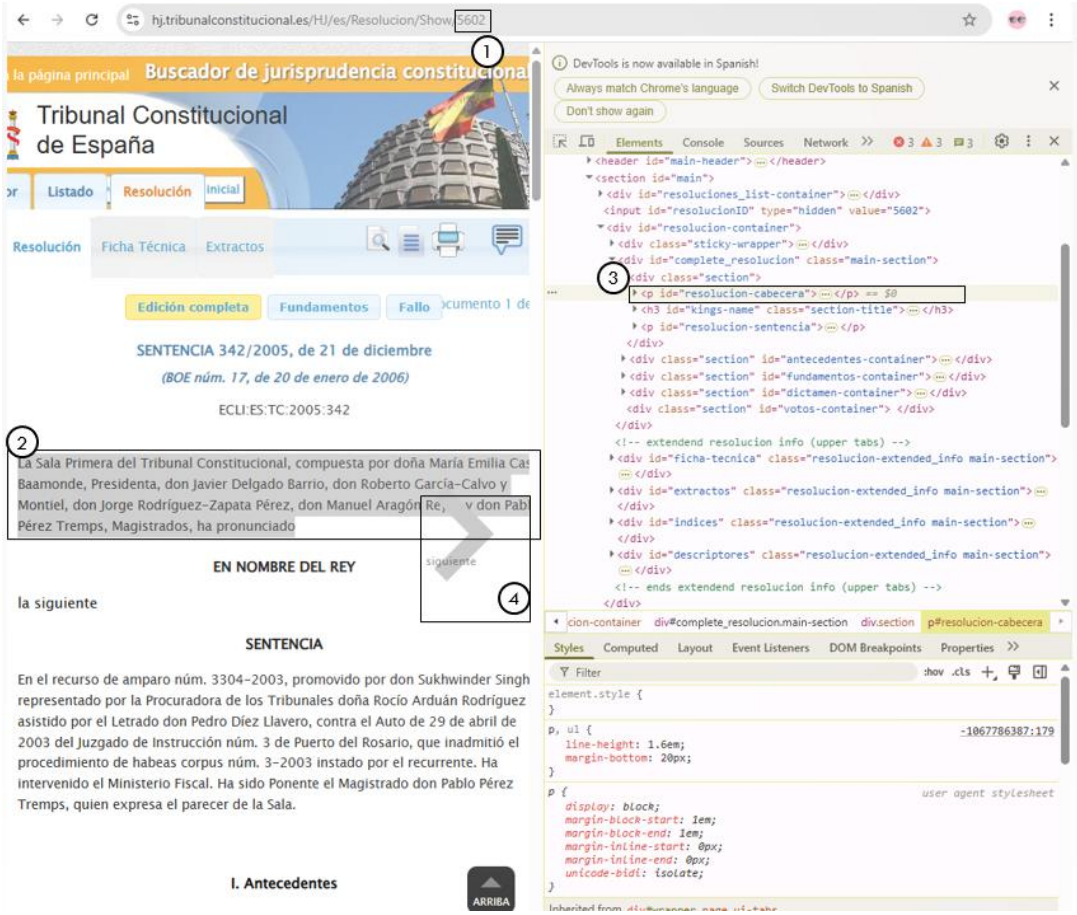

Figura 4. Inspección del texto de la página web del Tribunal Constitucional. Fuente: Elaboración propia

Al acceder a cualquiera de las sentencias listadas en la Figura 3, el usuario es redirigido a una nueva página, tal como se observa en la parte izquierda de la Figura 4 . En ella se presenta el contenido completo de la resolución, incluyendo la identificación del órgano judicial, los antecedentes, fundamentos jurídicos y el fallo.

Con el objetivo de recopilar todas las sentencias dictadas por el Tribunal Constitucional durante un año específico, se identificó un patrón estructural en los enlaces URL de su página oficial. En concreto, el número situado al final de cada enlace —señalado con el número 1 en la Figura 4— sigue una secuencia numérica consecutiva. Por ejemplo, la Sentencia 342/2005 corresponde a un enlace que finaliza en 5602; al modificar manualmente este número por 5601 o pulsando la flecha numerada con el 4, se accede a la Sentencia 341/2005, y así sucesivamente.

No obstante, debido a una modificación en la configuración de los enlaces del Tribunal Constitucional, la obtención de las sentencias correspondientes al año 2024 resultó algo más tediosa, según se observa en el Anexo III: Código completo utilizado en RStudio relativo al año 2024. La numeración de los enlaces ahora incluye tanto las sentencias como los autos dictados, lo que generaba un total de 1.303 elementos. Para filtrar los autos, se modifica el bucle de extracción de sentencias, de manera que verifique si la primera línea de la sentencia comenzaba con el formato "ECLI:ES:TC:2024:NUM.A", ya que se observó que las sentencias no contenían una "A" al final de dicha línea. Así se pasa de un total de 1.303 resultados iniciales a 155 sentencias.

Para determinar qué secciones del texto debían ser extraídas, se procedió a inspeccionar el código HTML de la página web. Como se observa en la Figura 4, el contenido de la sentencia —marcado con el número 2— se encuentra encapsulado principalmente en etiquetas <p> (párrafos), como queda reflejado en el recuadro número 3. Asimismo, se consideraron las etiquetas <h4>, dado que aportan estructura y encabezados relevantes para la interpretación del texto.

Además, se optó por desechar la parte de los fundamentos de hecho porque, tal y como se explica en el apartado 2.2.1, la situación fáctica de los individuos que recurren al Tribunal Constitucional no ayuda a entender los temas que se tratan en las sentencias.

Con base en estos hallazgos, se desarrolló en R una función denominada get_online_text, programada para acceder automáticamente a las resoluciones disponibles en la web del Tribunal Constitucional, recogida en el código de los ANEXOS.

A este respecto, cabe recordar que este órgano no solo resuelve recursos de amparo destinados a la protección de los derechos de los ciudadanos, sino que también dicta otras resoluciones,

principalmente recursos de inconstitucionalidad y cuestiones de competencia, dedicados a examinar la legalidad de las normas y a enjuiciar la competencia entre órganos.

En consecuencia, el análisis incluirá temas derivados de estos otros tipos de sentencias. No obstante, en términos generales, el número de resoluciones que no corresponden a recursos de amparo constituye una proporción significativamente menor dentro del conjunto de sentencias Tribunal Constitucional, por lo que esta precisión no afectará a nuestro análisis general.

Nombre de la variable	**Descripción**
Identificador	Número de la sentencia ordenado por fecha ascendiente.
Fundamentos de hecho	*Contenido de las vulneraciones de derecho*
Fallo	Resolución final del Tribunal y votos particulares (si existieran)

Tabla 1: Estructura y variables de los datasets. Fuente: Elaboración propia.

Para la construcción de las bases de datos, se consideró fundamental que, al menos, se incluyeran tres columnas: una que contuviera el identificador de cada sentencia, otra con los fundamentos jurídicos, y una última con el fallo, como se muestra en la Tabla 1.

Además, se genera un dataframe que puede exportarse a un archivo Excel para su visualización.

En el año 2005, el corpus está compuesto por 337 sentencias, en 2009 se extraen un total de 220 sentencias y el corpus del año 2024 está compuesto por 155 sentencias.

4.3. Preprocesamiento de los datos

Una vez recopilados los tres conjuntos de datos, se lleva a cabo la tokenización y el preprocesamiento de los fundamentos jurídicos de las sentencias. Tal y como se indica en el punto 3.2.1, este procedimiento tiene como finalidad estructurar y depurar la información, simplificando el texto para optimizar la aplicación del modelo LDA.

En este estudio, se han seguido las etapas recomendadas en la literatura académica, conforme a lo señalado por Velilla (2023). Se presentan las siguientes técnicas:

- Tokenización
- Conversión del texto en minúsculas
- Eliminación de stopwords
- Lematización
- Identificación de colocaciones

En primer lugar, se emplea la librería *udpipe* para el etiquetado morfosintáctico permitiendo identificar y clasificar las distintas categorías gramaticales, como sustantivos, verbos y adjetivos, en los textos de las sentencias (Wijffels, 2022). Como parte de este proceso de tokenización, se eliminan los símbolos, números y signos de puntuación que carecen de significado para las sentencias. Asimismo, se utiliza la misma librería para convertir todo el texto a minúsculas, evitando así que términos como "Igualdad" e "igualdad" sean tratados como palabras distintas.

El siguiente paso consiste en la eliminación de las stopwords con el objetivo de centrar el análisis en los términos que aportan mayor valor a las sentencias por su carácter distintivo. En esta fase, se emplea la librería *snowball* (Reyes-Ortiz et al., 2017) que ofrece soporte para el español a diferencia de otras librerías como *smart.*

Ejemplos de estos términos son "el", "sea", "tuvieron" y "las". Además, se amplía esta lista con palabras específicas del ámbito jurídico que, tras un análisis del corpus, se han identificado como recurrentes, pero poco informativas. Entre estos términos se encuentran "recurso", "amparo", "vulneración", "juzgado", "sentencia", "jurídico", "derecho", "procedimiento", "instancia" y "fundamento".

Por su parte, se eliminan los términos con menos de tres caracteres y se filtran aquellas palabras que no pertenezcan a las categorías gramaticales de sustantivos (NOUN), adjetivos (ADJ) o nombres propios (PROPN), dado que estos elementos suelen contener la mayor carga informativa dentro del corpus.

Ejemplos de lematización realizados son la conversión de tokens como “resueltas” o “resolvió” en “resuelto”: El inconveniente de esta técnica es la tardanza del procesamiento puesto que se trata de un corpus formado por sentencias cuyos fundamentos jurídicos son extensos. Además, aunque la literatura suele aplicar ambas técnicas (Figuerola et al., 2004), la lematización suele derivar en mejores resultados.

Posteriormente se procede a la identificación de colocaciones como bigramas, es decir, combinaciones de palabras que aparecen juntas con una frecuencia significativa. En este caso, tras evaluar diferentes valores, se ha determinado un umbral mínimo de aparición de 10.

Para seleccionar los bigramas más relevantes, se emplea la medida estadística Pointwise Mutual Information (PMI), que permite determinar qué combinaciones de palabras aparecen juntas con mayor probabilidad de lo esperado por azar. Se han seleccionado aquellos bigramas con un PMI igual o superior a 3, lo que indica una fuerte asociación entre los términos.

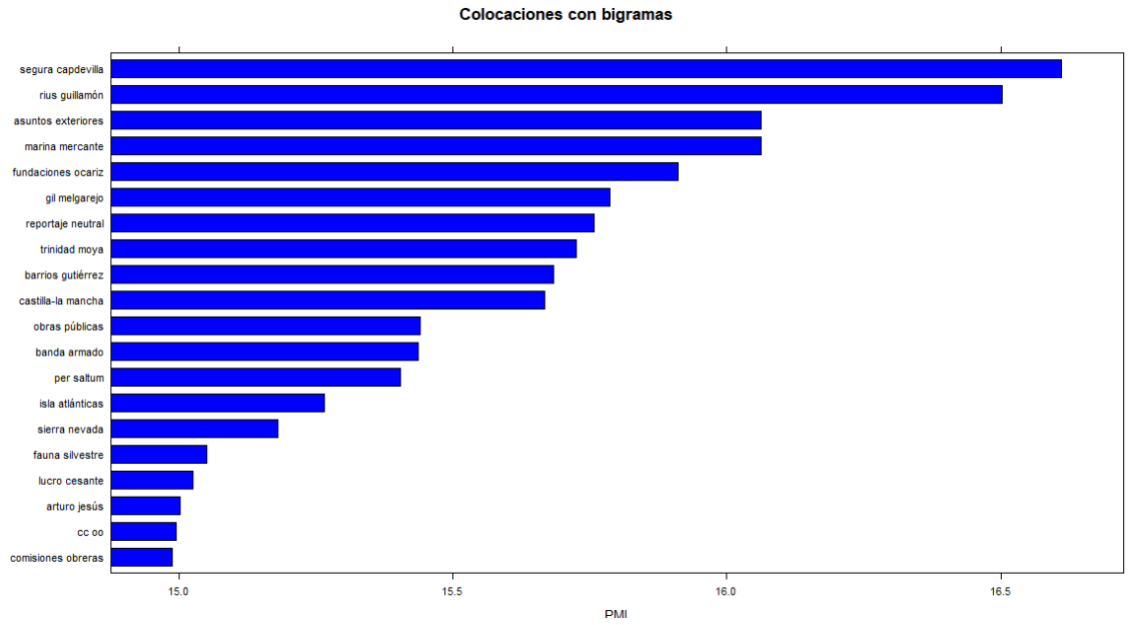

Figura 5: Ejemplo de colocaciones en el corpus. Fuente: elaboración propia a partir de los datos del caso de estudio.

La Figura 5 muestra los bigramas más frecuentes mediante un gráfico de barras utilizando la biblioteca *lattice*. Si bien algunos bigramas incluyen nombres de magistrados sin relevancia para el análisis del texto, otros, como “asuntos exteriores”, “obras públicas” o “comisiones obreras”, resultan significativos, ya que permiten captar mejor el significado del texto. De este modo, se distingue entre un “asunto” y los “asuntos exteriores”, entre una “obra” y una “obra pública”, o entre una “comisión” y las “comisiones obreras”.

Finalmente, los bigramas identificados se incorporan al corpus con el objetivo de mejorar la representación de los términos en el análisis posterior. Para ello, cada término se sustituye por su correspondiente bigrama si forma parte de las combinaciones detectadas, lo que permite una representación más precisa de las relaciones semánticas en el texto.

4.4. Análisis exploratorio de los datos

Antes de la aplicación del LDA se realiza una serie de técnicas para obtener una primera aproximación del contenido de los tres corpus de sentencias a través de la métrica TF-IDF.

Previamente a la aplicación de esta métrica, se analiza la frecuencia de términos por documento.

Frecuencia de términos por documento

Documento	Término	Frecuencia
doc1	proceso	1
doc1	auto abril	2
doc1	instrucción	2
doc1	puerto rosario	3
doc1	petición	1
doc1	habea corpus	3
doc1	sukhwinder	4
doc1	singh	4
doc1	cuestión aquí	1
doc1	resuelto noviembre	1
doc1	virtud aplicación	1
doc1	libertad extranjero	1
doc1	vigente sazón	1
doc1	medida ingreso	1
doc1	expresa	1
doc1	dicción	1
doc1	interesado	1
doc1	juez instrucción	1
doc1	competente	1
doc1	decisión judicial	2

Figura 6. Frecuencia de términos por documento. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 6 presenta los primeros 20 términos de la primera sentencia (doc1) junto con su frecuencia de aparición para el corpus del año 2005. También se incluyen bigramas, como "Puerto Rosario" o "habeas corpus" (este último término hace referencia a un derecho fundamental).

A continuación, se calcula la métrica TF-IDF para cada uno de los corpus con el objetivo de identificar preliminarmente los términos más relevantes presentes en cada conjunto de

documentos. Para ello, se agrupan los términos a partir de la matriz TF-IDF y se suma el valor correspondiente para cada palabra en el corpus, obteniendo así una medida agregada de su importancia. Los resultados se visualizan mediante un gráfico de barras que muestra los términos con mayor peso y una nube de palabras que permite representar un volumen más amplio de vocabulario relevante.

4.5. Obtención de la coherencia

Para determinar el número de tópicos, se utiliza la métrica de coherencia UMass y se obtiene el siguiente gráfico, que muestra una tendencia decreciente en la coherencia a medida que aumenta el número de tópicos.

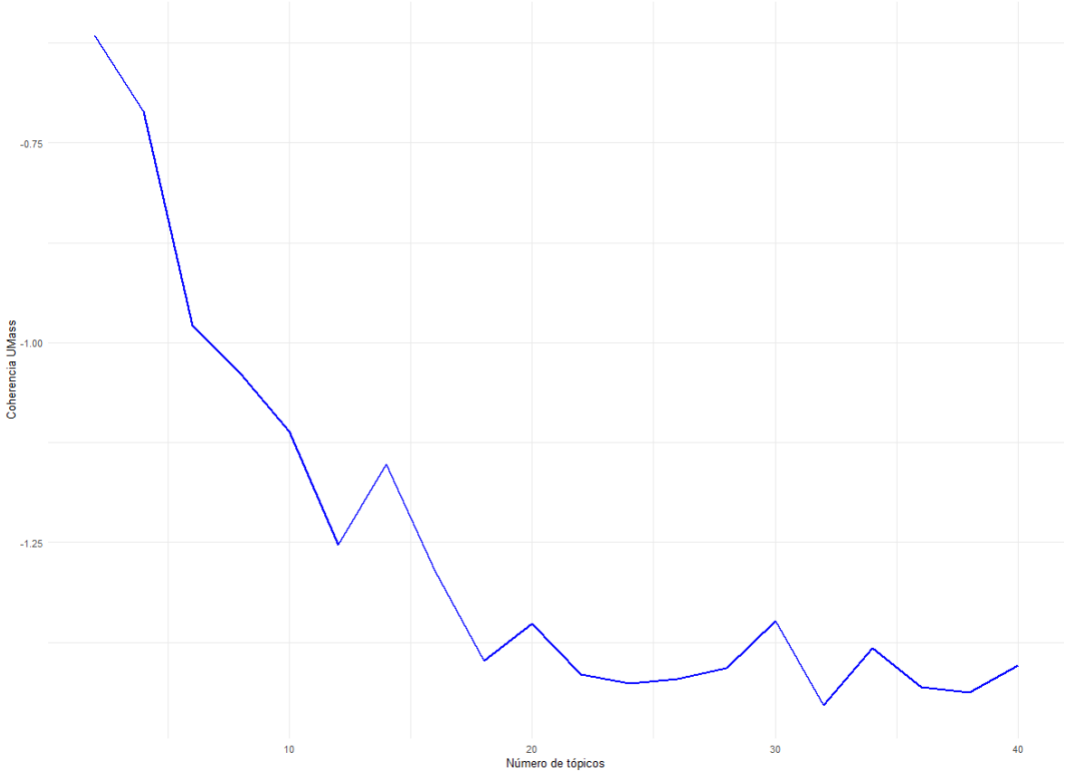

Figura 7. Cálculo de la coherencia para el modelado de tópicos. Fuente: elaboración propia a partir de los datos del caso de estudio

En la Figura 7 se muestra el gráfico para el modelado de tópicos. En el eje de las X se representa el número de tópicos (desde 0 hasta 40), y en el eje de las ordenadas (Y), se muestra la coherencia UMass. En el análisis de coherencia de los modelos de tópicos, para cada valor de k (número de tópicos) dentro del rango especificado, se realizan cinco ejecuciones independientes del modelo LDA utilizando el método de Gibbs. En cada iteración, se extraen las diez palabras más representativas de cada tópico y se calcula la coherencia. Posteriormente, se obtiene el promedio de estas cinco mediciones de coherencia para cada k, asegurando una estimación más robusta y reduciendo la variabilidad inherente al proceso de inicialización aleatoria del modelo.

Como se aprecia en la Figura 7, el valor de coherencia disminuye de forma exponencial hasta aproximadamente k = 12, momento a partir del cual se estabiliza con un ligero repunte, alcanzando un máximo local en k = 14. Si bien los valores de coherencia más elevados se observan en k = 10 y k = 14, una revisión manual de los resultados revela que, con k = 10, desaparecen algunos temas relevantes identificados en el corpus, lo que compromete la interpretabilidad del modelo. Por otro lado, aunque k = 14 presenta una coherencia ligeramente superior, introduce tópicos redundantes, lo que también afecta negativamente al análisis.

Teniendo en cuenta que la elección óptima de k debe equilibrar tanto la coherencia como la interpretabilidad de los temas, se opta finalmente por un modelo con k = 12. Este valor ofrece un punto medio adecuado: mantiene una buena coherencia temática y permite la aparición de tópicos claramente diferenciados y no redundantes, asegurando así la riqueza del análisis sin sacrificar calidad interpretativa.

Por motivos de eficiencia computacional, dado que el cálculo de la coherencia implica obtener repetidamente las frecuencias de coaparición de términos para distintos valores de k, la evaluación de coherencia se ha realizado únicamente sobre el corpus correspondiente al año

2024, que es el que contiene el menor número de sentencias. Sin embargo, se ha llevado a cabo una revisión manual de varios niveles de k y resulta conveniente mantener un mismo valor de k para los tres corpus.

Para la inferencia, se emplea el muestreo de Gibbs, tanto para calcular la coherencia del modelo como para identificar los tópicos más recurrentes. En cuanto a los hiperparámetros del modelo, se asigna un valor de α igual a 50 dividido por el número de temas, lo que influye en la distribución de los términos dentro de los temas, y un δ de 0.1, que regula la dispersión de los documentos dentro de los temas. Ello se obtiene de la literatura existente citada en Velilla (2023)[1]. Además, para obtener los tópicos más recurrentes se fija una semilla aleatoria para garantizar la reproducibilidad de los resultados, ello permitirá obtener los mismos resultados cada vez que se ejecuta el código.

4.6. Modelado de tópicos y comparación de los años

La aplicación de LDA aporta dos salidas fundamentales: la distribución de términos en cada tópico (recogido en la matriz phi) y, por otro lado, la distribución de tópicos para cada uno de los documentos (recogido en la matriz theta).

Una vez identificados los temas mediante su visualización, se procede a la interpretación de los tópicos obtenidos para cada año analizado. Esta interpretación se basa en los términos con mayor peso dentro de cada tópico, a partir de los cuales se asigna un nombre representativo que facilite su comprensión. En aquellos casos en los que los términos resultaban ambiguos, se consultaron directamente los documentos con mayor carga temática para contextualizar adecuadamente el contenido.

[1](Griffiths y Steyvers, 2004; Ponweiser, 2012; Yan et al., 2009)

En cuanto a la visualización de resultados, se utilizan distintos gráficos interactivos. En primer lugar, un gráfico de distancia intertópica que permite evaluar la calidad y la diferenciación entre los tópicos. En este gráfico, la distancia entre círculos representa la similitud temática, mientras que el tamaño de cada círculo indica la importancia relativa del tópico dentro del corpus.

En segundo lugar, se calcula la proporción general de cada tópico dentro del corpus mediante un gráfico de barras horizontales. Esta visualización permite observar de forma clara el peso relativo de cada tema en un año concreto, contribuyendo así a la comparación temática entre los diferentes periodos analizados.

Por último, se genera un mapa de calor que muestra cómo se distribuyen los distintos tópicos a lo largo de los documentos del corpus. Estos documentos se organizan cronológicamente en orden descendente, lo que facilita el análisis de la evolución temática a lo largo del tiempo.

5. RESULTADOS OBTENIDOS Y SU DISCURSIÓN

A continuación, se presentan los resultados obtenidos para cada uno de los tres corpus, comenzando con un análisis exploratorio que facilita una primera aproximación a los datos y permite identificar patrones generales a través de la métrica TF-IDF.

Con el objetivo de comparar las diferencias entre los años analizados, se ha optado por estructurar la información de manera paralela para cada corpus, es decir, para cada año. Finalmente, se ofrecerá una interpretación conjunta que sintetice las principales conclusiones del análisis.

5.1. Resultados del análisis exploratorio

Uno de los aspectos más destacables identificados antes de realizar el modelado de tópicos es la disminución progresiva en el número de sentencias dictadas a lo largo de los períodos analizados. En concreto, en el año 2009 ya se aprecia una reducción aproximada del 36 % respecto al volumen de sentencias de 2005. Esta tendencia descendente se acentúa aún más en el año 2024, lo que constituye un primer indicio del posible impacto de las medidas adoptadas durante estos periodos.

Figura 8. Nube de las palabras más relevantes para el año 2005. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 8 pone de manifiesto la presencia destacada de términos como “sindical” y “trabajador”, los cuales se vinculan con el derecho a la libertad sindical recogido en el artículo 28 de la Constitución. Asimismo, aparecen palabras relacionadas entre sí, como “parque” y “nacional”, que aluden a cuestiones de carácter medioambiental.

Junto a estos, se identifican términos de naturaleza procesal, como “prueba”, que resultan recurrentes en la mayoría de los procedimientos judiciales. También destacan conceptos vinculados a derechos fundamentales, como “libertad”, o a competencias propias del Tribunal Constitucional, como “autonómico”.

Por otro lado, se observan términos de menor frecuencia que resultan igualmente significativos. Es el caso de “enseñanza”, asociado al derecho a la educación (artículo 27 de la Constitución), así como abreviaturas legislativas como “ljca” (Ley de la Jurisdicción Contencioso-Administrativa) o “lopj” (Ley Orgánica del Poder Judicial), utilizadas comúnmente para referirse a textos legales concretos.

Figura 9. Nube de las palabras más relevantes para el año 2009. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 9 revela, en una primera lectura, la existencia de una categoría predominante de términos relevantes. Sobresalen palabras como “partido”, “electoral” y “candidatura”, lo que

sugiere una concentración temática en torno a procesos electorales. Asimismo, se identifican términos procesales como “prueba” y “pena”, que reflejan aspectos comunes a múltiples procedimientos judiciales.

Si bien términos como “trabajador” continúan presentes, su relevancia disminuye en comparación con otros tópicos predominantes en este corpus.

En cuanto a los términos de menor frecuencia, se observan palabras como “mujer”, “varón” y “menor”, que aparecen con un tamaño reducido indicado una menor relevancia estadística en el análisis. También se identifican expresiones vinculadas a derechos específicos, como “médico” o “salud”, relacionadas con el derecho a la integridad física, que si bien no destacan en términos globales, mantienen un interés temático relevante.

Figura 10. Nube de las palabras más relevantes para el año 2024. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 10 muestra los términos más representativos del año 2024, evidenciando palabras como “ayuda” y “vivienda”. Además, resulta especialmente significativo el crecimiento de la palabra “mujer”, que puede considerarse una de las más relevantes del corpus. Esta aparece vinculada a otros términos clave como “embarazo”, “violencia”, “sexual” o “menor”, lo que

sugiere una mayor atención a cuestiones relacionadas con derechos de género y protección de colectivos vulnerables. Esto podría estar relacionado con la evolución del marco normativo en materia de igualdad y derechos familiares.

En cuanto a los términos de menor importancia relativa, predominan palabras genéricas como “asunto” o “civil”, las cuales suelen estar asociadas a aspectos procedimentales secundarios y no aportan información sustancial en términos temáticos al quedar subsumidas en conceptos más amplios del corpus.

Se ofrece además otra forma adicional de visualizar las palabras más relevantes de cada corpus mediante la métrica TF-IDF a través de gráficos de barras disponibles en el Anexo IV: Gráficos complementarios a las nubes de palabras. La Figura 23, Figura 24 y Figura 25 muestran los 20 términos con mayor TF-IDF para cada corpus de sentencias.

5.2. Año 2005

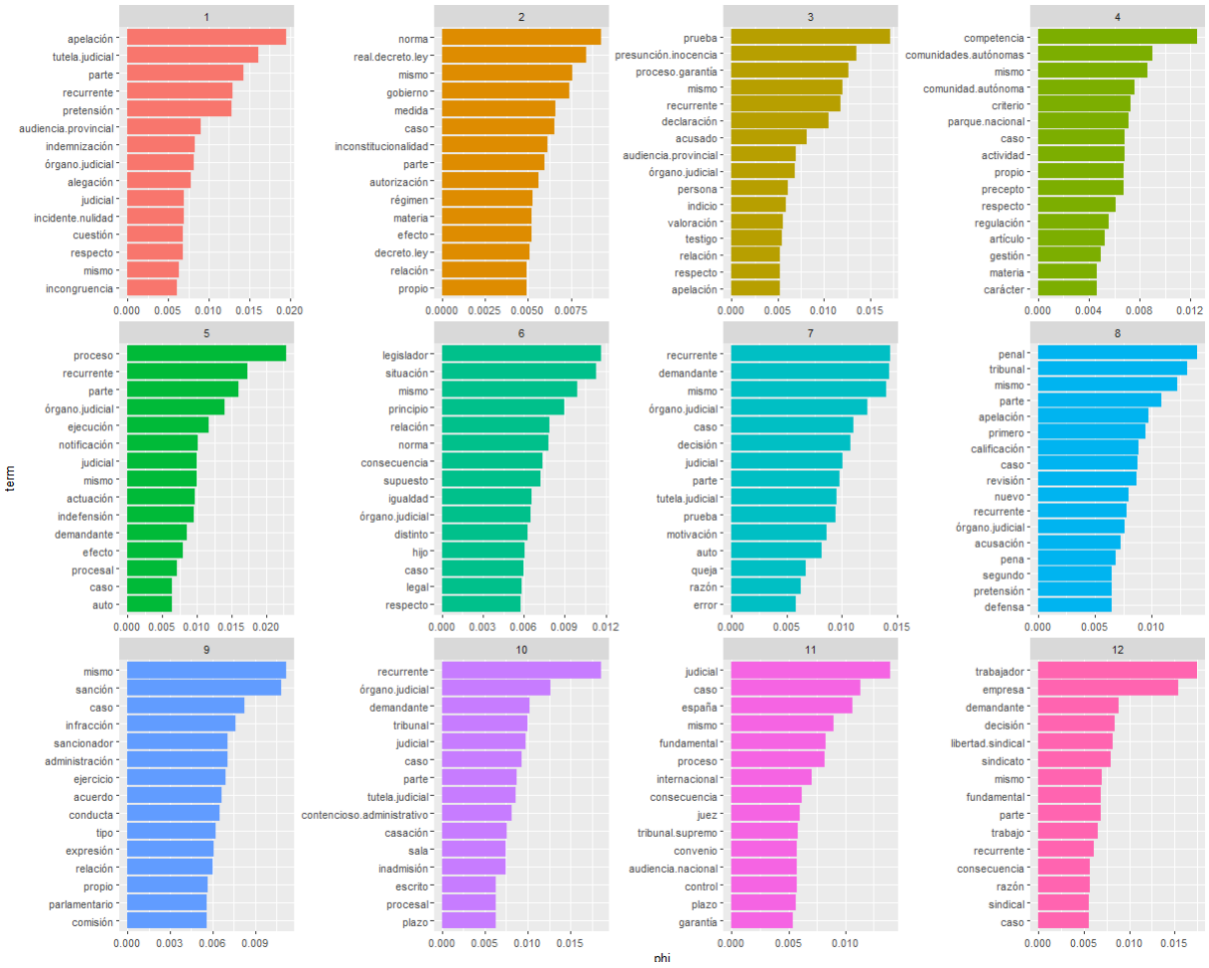

Figura 11: Visualización de los tópicos para el año 2005. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 11 identifica los 15 términos más representativos de cada uno de los tópicos en función de su peso en el modelo. A cada tópico se le ha asignado un nombre interpretativo, que será utilizado en las visualizaciones posteriores correspondientes a este corpus para facilitar su lectura y análisis.

- **Tópico 1 – Indemnización**: las palabras "incidente nulidad", "incongruencia" e "indemnización" hacen alusión a las reclamaciones de los particulares para conseguir

una indemnización superior a la establecida por la Audiencia Provincial, que suele ser el órgano judicial anterior.

- **Tópico 2 – Regulación por Real-Decreto Ley:** Este tópico trata de aquellos asuntos en los que el Tribunal Constitucional enjuicia la falta de competencia del Gobierno de regular por Real-Decreto Ley materias reservadas a la ley. Destacan términos como "inconstitucionalidad" pues es el nombre del recurso utilizado para ello, y "Real-Decreto Ley".

  Esta situación responde a que, a finales de la década de 1990 y en los años siguientes, la legislación mediante esta vía alcanzó niveles históricos, lo que motivó que los recurrentes alegaran la posible inconstitucionalidad de estas disposiciones al ser promulgadas por el poder ejecutivo en lugar del legislativo.

- **Tópico 3 – Tutela judicial efectiva: Presunción de inocencia:** Términos como "proceso garantía", "prueba" o "presunción de inocencia" conducen al tópico que recoge las sentencias en los que los particulares recaban el derecho a un proceso con todas las garantías de legalidad. Este derecho forma parte del derecho a la tutela judicial efectiva.
- **Tópico 4 – Competencias Autonómicas:** El reparto de competencias entre Estado y las Autonomías gobierna este tópico, en el que destacan términos como "competencia", "comunidades autónomas" y "parque nacional". Este último término destaca porque en materia de parques nacionales hay muchas competencias compartidas, que con carácter general dan lugar a recursos de competencia.
- **Tópico 5 – Tutela judicial efectiva: Indefensión:** Términos como "ejecución", "proceso" e "indefensión" conducen inevitablemente a otra modalidad del derecho a la tutela judicial efectiva: la indefensión.

- **Tópico 6 – Igualdad:** Términos como “situación”, “mismo”, “norma” e “igualdad” presumen un tópico dedicado al derecho a la igualdad de todos los individuos ante la ley.
- **Tópico 7 – Tutela judicial efectiva: Prueba:** Otra vertiente del derecho a la tutela judicial efectiva es el derecho a una correcta obtención y valoración de las pruebas. Este tópico muestra la gran relación de los términos “tutela judicial” y “prueba”.
- **Tópico 8 – Penal:** Términos como “pretensión”, “defensa” y “pena” aluden a casos en los que se impugna la sentencia proveniente de órganos penales, pues suelen ser aquellos que tratan más materias relacionadas con derechos fundamentales.
- **Tópico 9 – Función política:** El conjunto de términos como “sanción”, “administración” y “parlamentario” pueden parecer extraños. Para acudir a la denominación del tópico se procedió a examinar la matriz theta. Se buscaron los documentos con mayor presencia en el tópico 9 y se obtuvo como resultado las sentencias 92/2005 y 91/2005, cuyo contenido versa sobre funcionarios a los que se les sanciona o se les niega una indemnización. Con carácter general, se trata del ejercicio de la función política.
- **Tópico 10 – Tutela judicial efectiva: Inadmisión:** La última vertiente del derecho a la tutela judicial efectiva que encontramos en este corpus es el derecho que invocan los particulares a obtener una sentencia en plazo. El exceso de trabajo del contencioso-administrativo genera que en ocasiones los particulares no obtengan resoluciones firmes, y que como consecuencia se tengan por inadmitidas sus peticiones.
- **Tópico 11 – Tratados y Convenios Internacionales:** Términos como “España”, “internacional” y “proceso” conducen a la competencia del Tribunal Constitucional de declarar la conformidad de estas normas con la constitución o aplicar derechos derivados de las mismas a los particulares. Para su denominación también se ha

procedido a un examen de la matriz theta, obteniendo como sentencia definidora del tópico la STC 303/2005, que trata del derecho del habeas corpus en materia de extranjería.

- **Tópico 12 – Libertad sindical:** Términos como "trabajador" o "empresa" conducen al derecho a la libertad sindical, término expresamente destacado también dentro del último tópico.

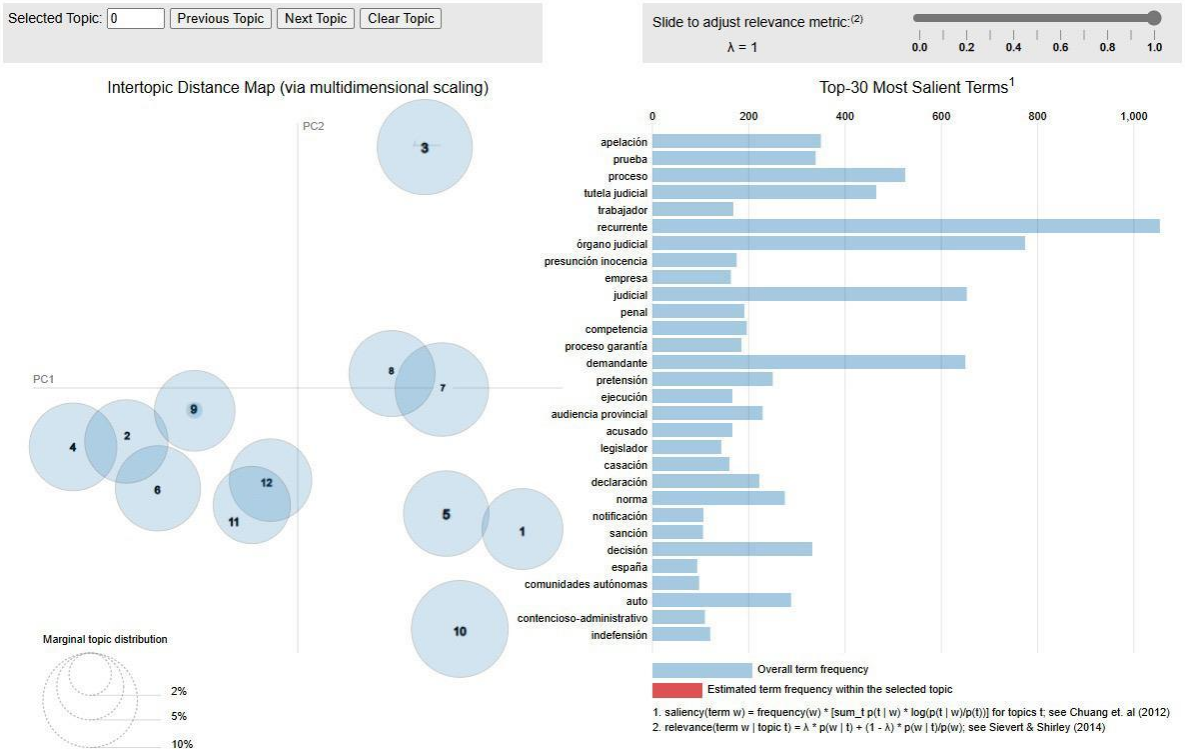

Figura 12: Mapa de distancia intertópica para el año 2005 con LDAVis. Fuente: elaboración propia a partir de los datos del caso de estudio

Aunque la Figura 12 consta de dos secciones diferenciadas, en este análisis nos centraremos en el mapa de distancia intertópica situado en el lado izquierdo, puesto que los términos más relevantes ya se han ilustrado en el apartado 5.1. a través de las nubes de palabras.

En la Figura 12 se observa que predomina una superposición entre los temas 7 y 8, correspondientes a la obtención de la prueba y los procedimientos penales. Esto tiene sentido

puesto que uno de los elementos definitorios de los procedimientos penales es el establecimiento de la pena. Todavía mayor es la superposición de los temas 11 y 12, relativos a cuestiones internacionales y libertad sindical. Ello puede deberse a que tanto en materia de extranjería como en el ámbito de los trabajadores, se frecuenta la invocación a la discriminación. En ambos casos el vocabulario utilizado es similar.

Asimismo, el tema relativo a la presunción de inocencia (3) parece configurarse de forma alejada a los demás, lo cual resulta sorprendente pues debería guardar similitud con el resto de vertientes del derecho a la tutela judicial efectiva. Una posible explicación es que las resoluciones centradas en la presunción de inocencia emplean un vocabulario más técnico y específico que otras vertientes de este derecho, lo que podría haber llevado al modelo a identificarlo como un tema diferenciado.

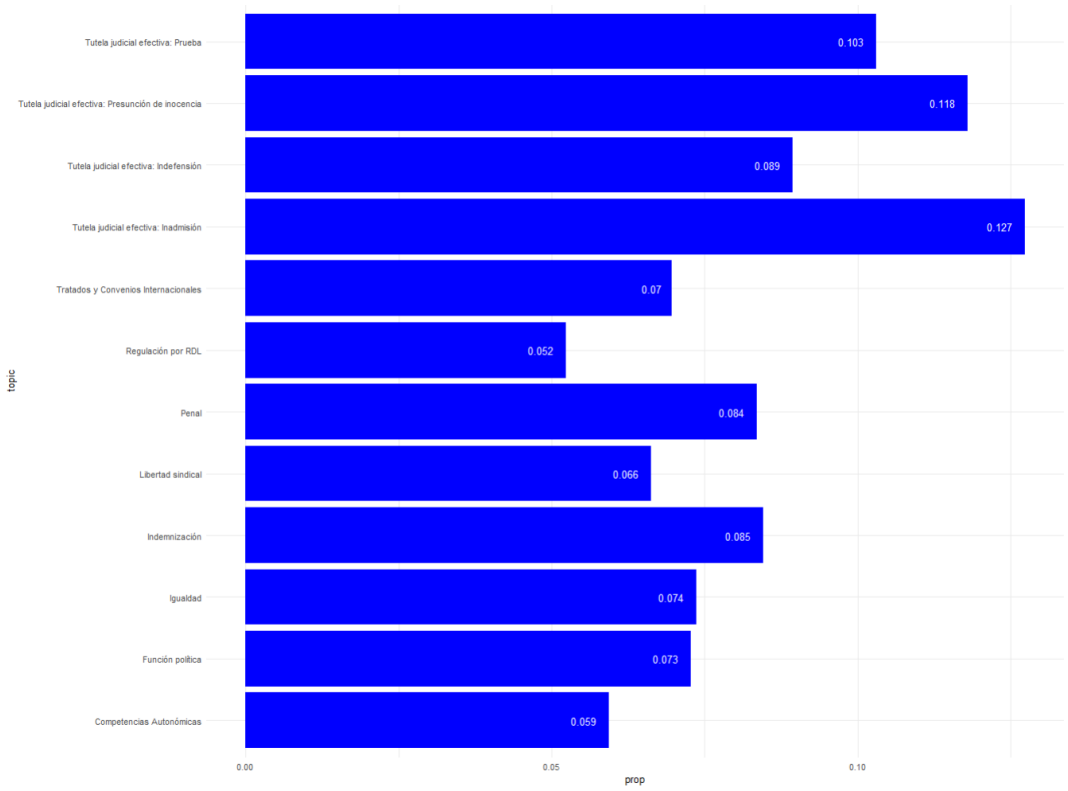

Figura 13: Peso relativo de cada tópico en el año 2005. Fuente: elaboración propia a partir de los datos del caso de estudio.

De acuerdo con la tendencia observada en la Figura 13, el tema predominante es el de la tutela judicial efectiva, manifestado en sus diversas facetas. Por otro lado, los conflictos de competencia (a nivel de regulación por Real-Decreto y a nivel autonómico) son los menos frecuentes. Esto es lógico pues la gran mayoría de sentencias resueltas por el Tribunal Constitucional son recursos de amparo.

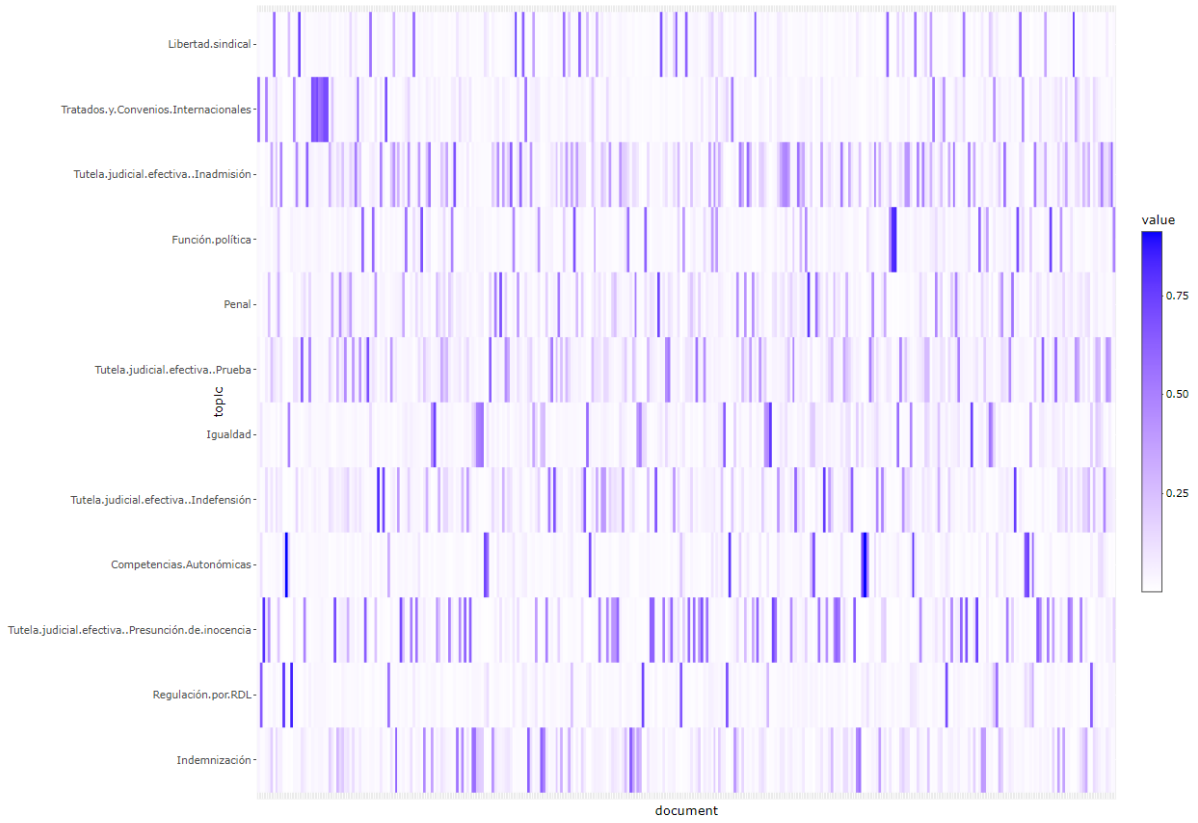

Figura 14: Mapa de calor para el año 2005. Fuente: elaboración propia a partir de los datos del caso de estudio

En la Figura 14, el eje horizontal (X) representa los distintos documentos analizados, mientras que el eje vertical (Y) corresponde a los diferentes tópicos identificados. La intensidad del color en cada celda indica la proporción relativa de cada tema dentro del documento correspondiente: los tonos más claros (blanco) reflejan una presencia baja o nula del tópico, mientras que los tonos más oscuros (azul) indican una mayor relevancia temática.

Del análisis de la Figura 14 se desprende que algunos temas, como las indemnizaciones, están distribuidos de manera uniforme a lo largo del año, mientras que otros, como los conflictos que devienen de normas con carácter internacional, tienden a concentrarse en determinados períodos, especialmente hacia finales del año. Esto sugiere que ciertos asuntos, como la invocación de la inadmisión como motivo de amparo o la solicitud de modificación de indemnizaciones por parte de los recurrentes, son frecuentes a lo largo del año, mientras que otros tienen un carácter más puntual y permiten caracterizar con mayor precisión el contenido de las sentencias emitidas en 2005.

5.3. Año 2009

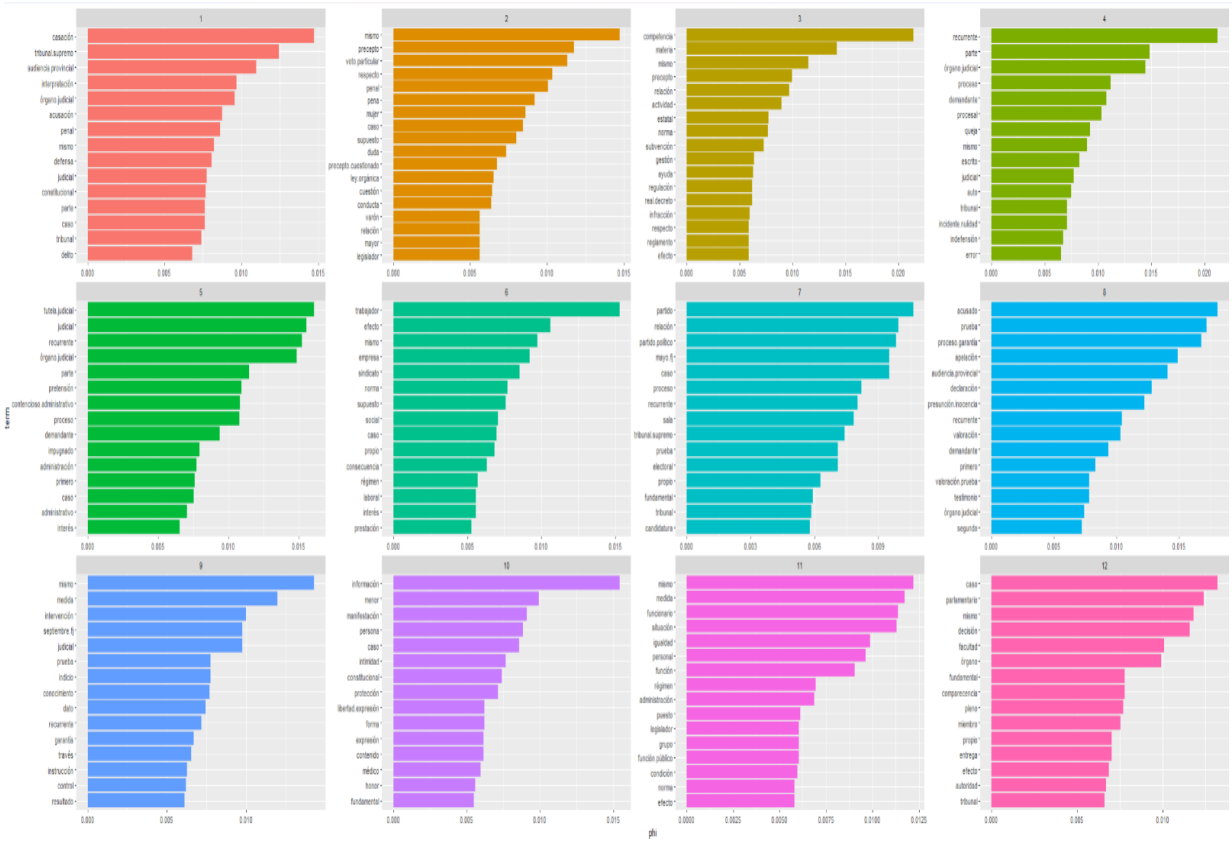

Figura 15: Visualización de los tópicos para el año 2009. Fuente: elaboración propia a partir de los datos del caso de estudio

Se identifican los siguientes tópicos de acuerdo con la Figura 15:

- **Tópico 1 – Recursos de casación:** Términos como “tribunal supremo” y “audiencia provincial”, junto con “recurso de casación” y otros términos procesales indican las impugnaciones que llegan al Tribunal Constitucional son recursos de casación, cuyo recurso previo ha sido dictado por las Audiencias Provinciales.
- **Tópico 2 – Igualdad: Mujer:** Esté tópico se caracteriza por términos como “mujer”, “varón” y “relación”, que lleva a pensar en el derecho de igualdad entre el hombre y la mujer. También destacan las palabras de “penal” y “pena”, puesto la invocación de este derecho tiene lugar en procedimientos de orden penal.
- **Tópico 3 – Regulación por Real Decreto Ley:** Términos como “competencia”, “real decreto” y “estatal” confluyen a la regulación del Gobierno por medio del Real-Decreto Ley. Además, destacan términos más específicos como “subvención” o “gestión”.
- **Tópico 4 – Tutela judicial efectiva: Indefensión:** Términos como “incidente nulidad” y “error”, junto con la propia “indefensión”, conducen a la modalidad de derecho a la tutela judicial efectiva relativa a la indefensión.
- **Tópico 5 – Tutela judicial efectiva: Inadmisión:** Este tópico ya presente en el corpus de sentencias de 2005 hace referencia a las invocaciones de los particulares a su derecho a la tutela judicial efectiva como consecuencia del procedimiento de la Administración. Destacan términos como “tutela judicial”, “impugnado” y “administración”.
- **Tópico 6 – Libertad sindical:** Términos como “trabajador”, “sindicato” y “laboral” conducen al tópico relativo al derecho a la libertad sindical.
- **Tópico 7 – Función política:** Este tópico trata derechos relativos a la función política a través de términos como “partido político”, “prueba” y “electoral”.
- **Tópico 8 – Tutela judicial efectiva: Presunción de inocencia:** Términos como “prueba”, “presunción de inocencia” y “valoración” hacen alusión a la variante del derecho a la tutela judicial efectiva relativa a la presunción de inocencia.

- **Tópico 9 – Tutela judicial efectiva: Prueba:** Términos como “pruebas”, “garantía” y “control” se refieren al derecho a la tutela judicial efectiva, que abarca el derecho a una correcta obtención y valoración de las pruebas.
- **Tópico 10 – Libertad de expresión, honor e intimidad:** Términos como “intimidad”, “libertad de expresión” y “honor” hace referencia a dos artículos de la Constitución. Uno de ellos trata el derecho al honor y a la intimidad y otro el derecho a la libertad de expresión.
- **Tópico 11 – Igualdad: Función política:** Términos como “administración” y “función” recuerdan al tópico del corpus de 2005 relativo a la función política. Sin embargo, destaca el término “igualdad” como especificidad al mismo.
- **Tópico 12 – Competencias del Congreso:** Términos como “comparecencia”, “pleno” y “facultad” conducen a el enjuiciamiento de las competencias políticas del Congreso.

En cuanto a la identificación de los temas más recurrentes, se observa que, si bien la mayoría de ellos se mantienen frente al corpus del año 2005, algunos presentan matices nuevos. Un ejemplo de ello es la evolución del derecho de igualdad (tópicos 2 y 11).

En 2005, los términos asociados a esta materia eran de carácter general, como “legislador”, “norma” o “legal”. En cambio, en 2009 este tema ofrece dos variantes. En la primera destaca la presencia del término “mujer” y, por otro lado, se vincula la igualdad con la función política. Ello puede deberse a que la reducción del número total de sentencias en 2009 permite que emerjan características más definitorias de las resoluciones analizadas.

Por otro lado, en 2009 surgen nuevos temas que no estaban presentes en 2005 como el derecho a libertad de expresión, el honor y la intimidad. En contraste, desaparecen cuestiones más generales, como la indemnización, dando paso a una mayor especificidad en los temas tratados.

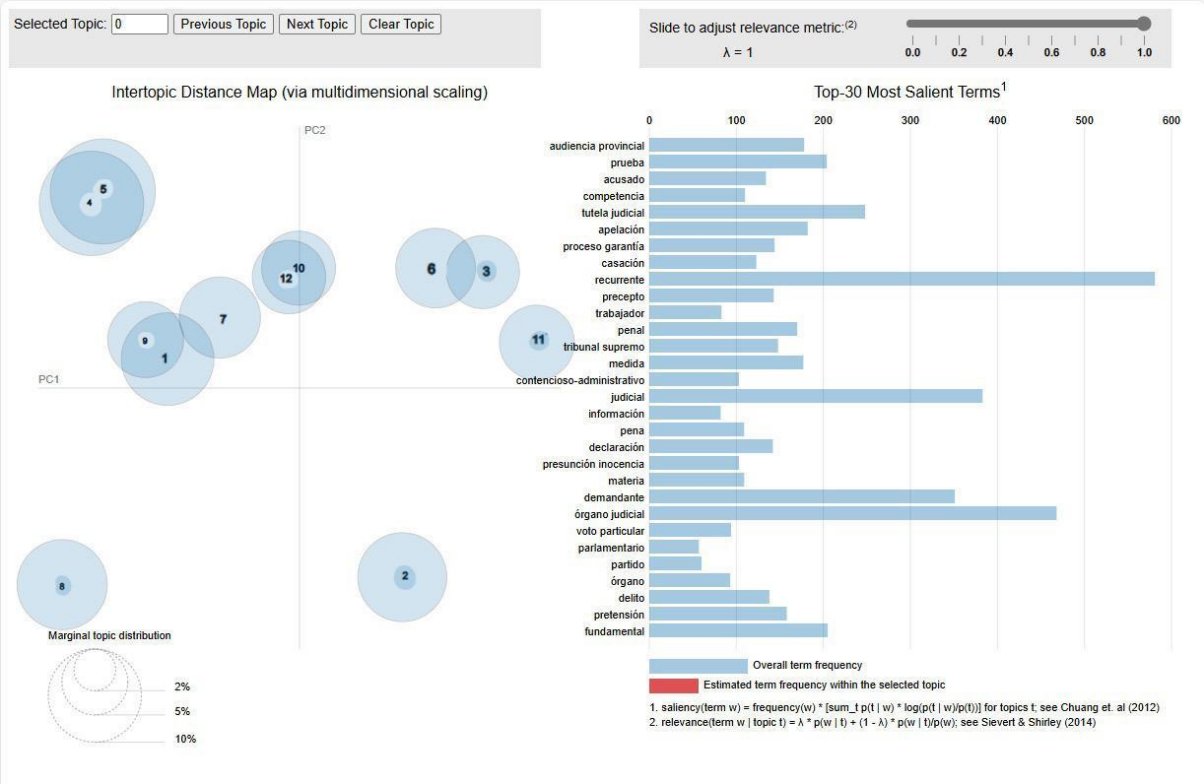

Figura 16: Mapa de distancia intertópica para el año 2009 con LDAVis. Fuente: elaboración propia a partir de los datos del caso de estudio

De acuerdo con la Figura 16 los temas abordados en 2009 presentan una mayor superposición que en la Figura 12. Este fenómeno podría indicar que, en lugar de abarcar una diversidad más amplia de cuestiones, el Tribunal Constitucional está aplicando el requisito de especial trascendencia constitucional de manera más selectiva, admitiendo a trámite aquellos asuntos que considera prioritarios para su resolución.

En particular, se observa una superposición casi total entre los temas 10 y 12, que se refieren a la libertad de expresión y a las competencias del Congreso, respectivamente. Esta coincidencia resulta comprensible, dado que en el tópico relacionado con las competencias del Congreso se incluyen sentencias en las que los diputados recurren en amparo para la protección de sus derechos. De manera general, el derecho más invocado por los diputados es su capacidad para expresarse libremente en el seno de las Cortes.

Asimismo, el tamaño de los círculos no es tan uniforme como ocurría en la Figura 12, lo cual indica un mayor número de sentencias recogiendo ciertos temas.

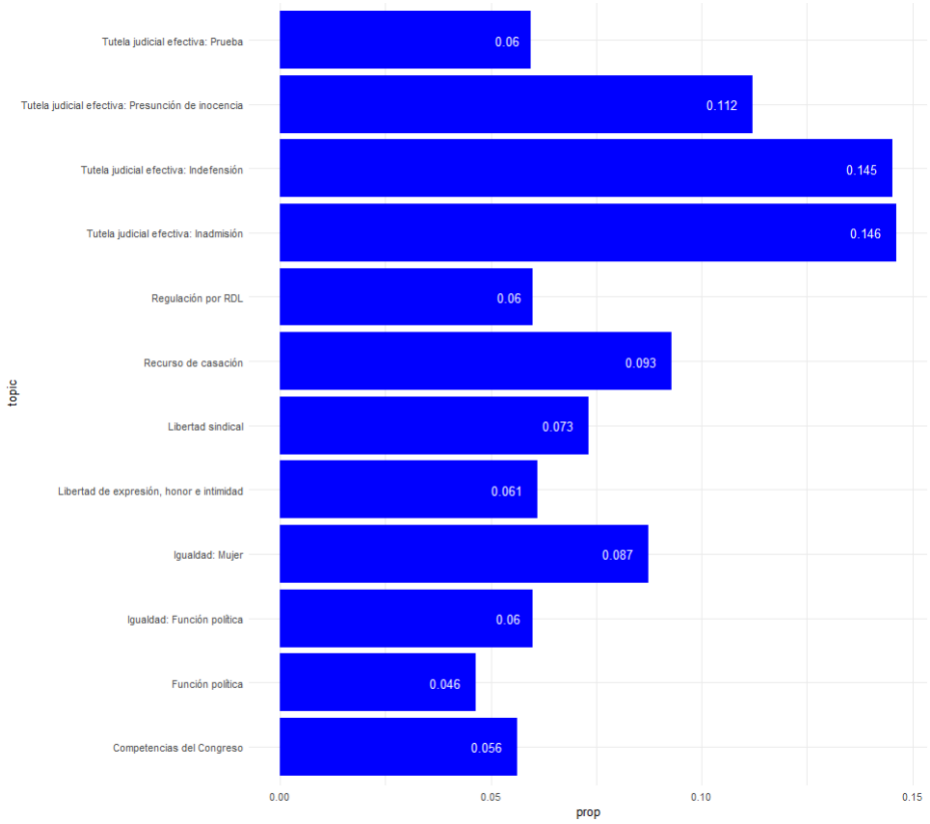

Figura 17. Peso relativo de cada tópico en el año 2009. Fuente: elaboración propia a partir de los datos del caso de estudio.

En relación con lo anterior, la Figura 17 confirma una mayor concentración de los temas en torno a la tutela judicial efectiva. Por otro lado, al igual que en la Figura 13, los temas más frecuentes en las sentencias son los relativos a la tutela judicial efectiva. También persiste la escasa proporción de sentencias relativas al tópico de competencias. Sin embargo, los tópicos relativos a la libertad sindical y la igualdad aumentan su prevalencia frente al corpus anterior.

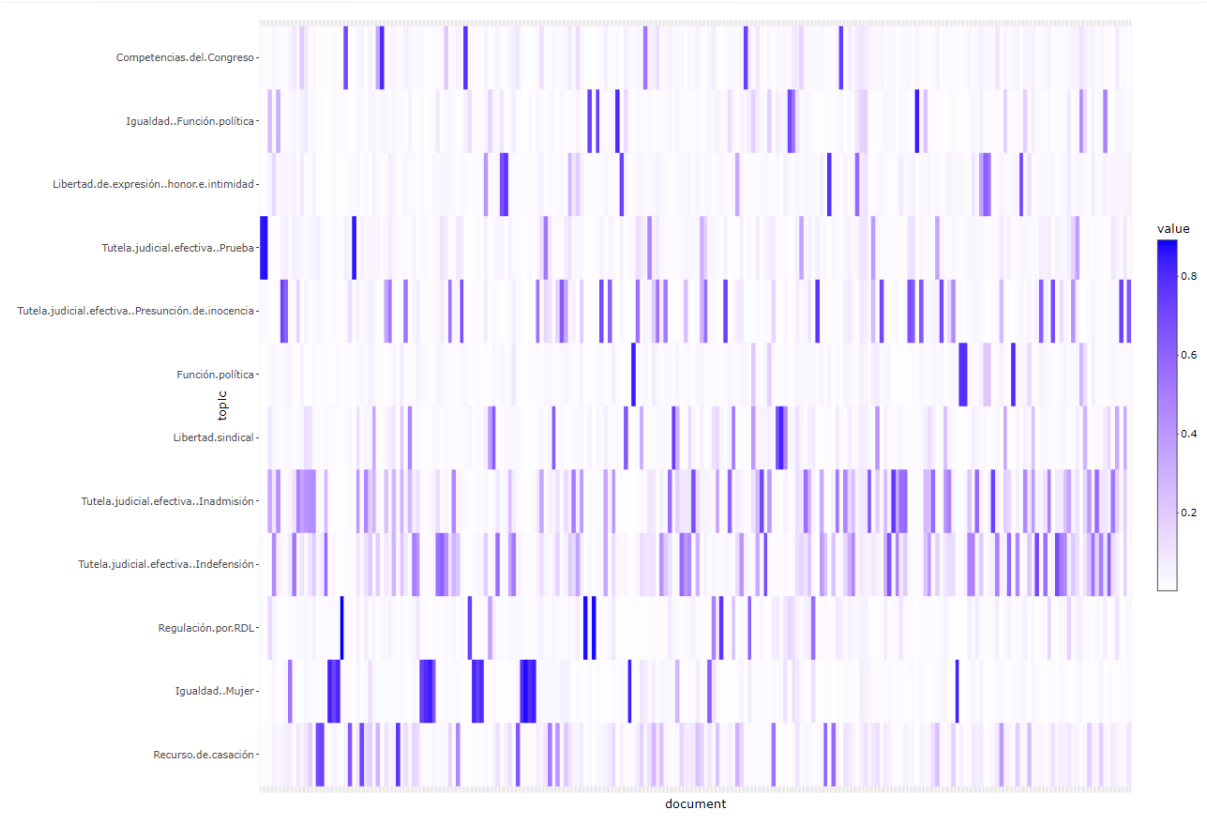

Figura 18: Mapa de calor para el año 2009. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 18 muestra que los temas más genéricos son abordados de forma continuada por el Tribunal. Por ejemplo, cuestiones vinculadas a la tutela judicial efectiva se tratan de manera constante a lo largo del año, al igual que ocurre con los recursos de casación. En cambio, las sentencias relacionadas con el derecho a la igualdad de la mujer presentan una presencia especialmente destacada hacia finales del año. Esto puede deberse a que el Tribunal tiende a concentrar el estudio de un tema determinado y resuelve de forma conjunta los recursos relacionados con esa materia.

5.4. Año 2024

En cuanto al análisis del corpus de 2024, se encuentran los siguientes tópicos.

Figura 19: Visualización de los tópicos para el año 2024. Fuente: elaboración propia a partir de los datos del caso de estudio

- **Tópico 1 – Tutela judicial efectiva: Extradición:** Términos como "tutela judicial", "extradición" y "orden público" hacen referencia a aquellas sentencias que versan sobre cómo el derecho de un Estado a solicitar la extradición de un ciudadano se contrapone al derecho a la tutela judicial efectiva.
- **Tópico 2 – Competencias Autonómicas:** Términos como "competencia", "dominio público" y "protección" infieren competencias compartidas entre el Estado y las Comunidades Autónomas. Destacan términos específicos como "litoral" o "suelo rústico".
- **Tópico 3 – Interés superior del menor:** Términos como "menor", "padre", "madre" y "interés superior" se refieren al derecho fundamental de los menores a que sus intereses sean considerados prioritarios. No se trata de un derecho incluido directamente en los artículos que conforman los derechos fundamentales, sino que vino introducido por la Convención de los Derechos del Niño.

- **Tópico 4 – Regulación por Real Decreto Ley:** Nos encontramos ante un tópico recurrente en el que destacan “decreto ley”, “infracción” y “control”.
- **Tópico 5 – Tutela judicial efectiva: Prueba:** Términos como “penal” y “prueba” sugieren una de las principales manifestaciones del derecho a la tutela judicial efectiva relativa a la valoración de las pruebas como fundamental para la determinación de la pena.
- **Tópico 6 – Función política:** Este tópico trata derechos relativos a la función política. Destacan términos como “diputado”, “acuerdo” y “proposición de ley”.
- **Tópico 7 – Presupuestos:** Este tópico hace referencia a los derechos invocados relacionados con la falta de regulación presupuestaria. Destacan términos como “ley presupuesto”, “acto” y “gobierno”.
- **Tópico 8 – Vivienda:** Términos como “vivienda”, “regulación”, “comunidad autónoma” y “competencia” hacen referencia a los conflictos de competencia en materia de vivienda como consecuencia de los derechos que los particulares invocan.
- **Tópico 9 – Tutela judicial Efectiva: Proceso:** Términos como “tutela judicial” y “proceso” hacen referencia a vulneraciones del derecho a la tutela judicial efectiva en sede del procedimiento judicial.
- **Tópico 10 – Libertad: Mujer:** En este tópico destacan palabras como “libertad”, “mujer” e “interrupción voluntaria”. Se trata de sentencias en las que se invoca al derecho a la libertad de la mujer.
- **Tópico 11 – Tutela judicial efectiva: Presunción de inocencia:** Términos como “presunción inocencia” y “conocimiento” conducen a enunciar la vertiente del derecho a la tutela judicial efectiva relativa al derecho de todos los ciudadanos a presumirse inocentes. Destacan términos específicos como “delito malversación”.

- **Tópico 12 – Igualdad: Laboral:** Términos como “seguridad social”, “igualdad” y “sindicato” permiten aducir la invocación del derecho a la igualdad. Destacan términos muy concretos como “diferencia trato”, “formación” o “discriminación”.

En la Figura 19 se observa la persistencia de ciertos temas recurrentes, como aquellos relacionados con la regulación mediante Real Decreto-ley, así como temas relativos a la función política. Ello es lógico y deriva de la estrecha relación que mantienen los funcionarios con la Administración, siendo más frecuentes las violaciones de derechos.

No obstante, además de la posible influencia de la reforma introducida en 2023, la evolución temporal desde los años 2005 y 2009 ha propiciado la aparición de cuestiones novedosas en los recursos de amparo.

En primer lugar, dentro del ámbito de la tutela judicial efectiva, destacan términos nuevos como “extradición”, que, aunque están presentes en el corpus de 2005, el gran volumen de sentencias no permitía apreciarlos. La extradición es el instrumento que permite a los estados recabar la presencia de las personas a las que se juzga, evitando que las fronteras territoriales sean un obstáculo a las pretensiones punitivas de un país (Pérez Manzano, 2004).

Por otro lado, en el tópico relativo a la presunción de inocencia surgen términos como la malversación o apropiación de patrimonio por parte de los funcionarios, lo cual implica que aquellos individuos acusados de malversación buscan defenderse ante la Corte Constitucional alegando el derecho a la presunción de inocencia.

Asimismo, se observa la presencia de temas de gran relevancia social y jurídica, cuya regulación ha incidido en derechos fundamentales. Por ejemplo, dentro del tema de igualdad (12), sobresalen términos como “sindicatos” y “semana”, lo que sugiere una posible relación con los permisos de maternidad y paternidad, asuntos ampliamente debatidos en el ámbito político y mediático. En cuanto al derecho a la libertad (10), aparecen términos como “mujer”

e “interrupción voluntaria”, reflejando la controversia y actualidad del debate sobre los derechos reproductivos.

Además, emerge un tema específico relacionado con la vivienda (8), que ya se adelantaba con la nube de palabras (Figura 10), y términos como “servicio” sugieren una orientación hacia la protección del Tribunal de quienes no son propietarios. Finalmente, resulta destacable la aparición de un tema exclusivo sobre los presupuestos, lo que podría estar vinculado a la inestabilidad política y la dificultad para su aprobación en ausencia de mayorías parlamentarias, una cuestión de gran impacto en la sociedad.

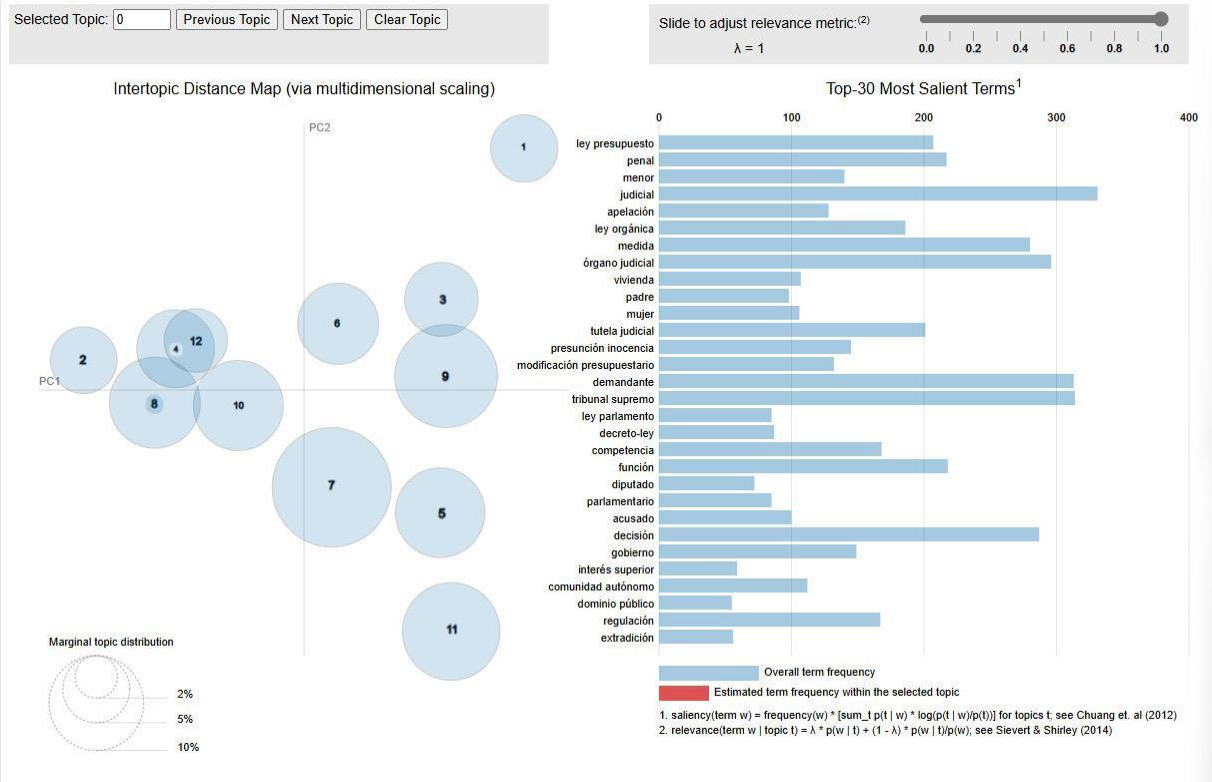

Figura 20: Mapa de distancia intertópica para el año 2024 con LDAVis Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 20 muestra una mayor diferenciación entre los temas que el año 2009, evidenciando cómo ciertos conceptos se agrupan en torno a cuestiones específicas. Destaca especialmente el tópico 7 (Presupuestos), que aparece como el más representativo en términos de tamaño, lo que

sugiere una alta frecuencia de sentencias relacionadas con la falta de regulación presupuestaria. En cambio, tópicos como el 2 (Competencias Autonómicas), el 4 (Regulación por RDL), el 10 (Libertad: Mujer) y el 12 (Igualdad: Laboral) se agrupan con mayor proximidad, lo cual podría deberse a una posible similitud semántica en los argumentos jurídicos empleados.

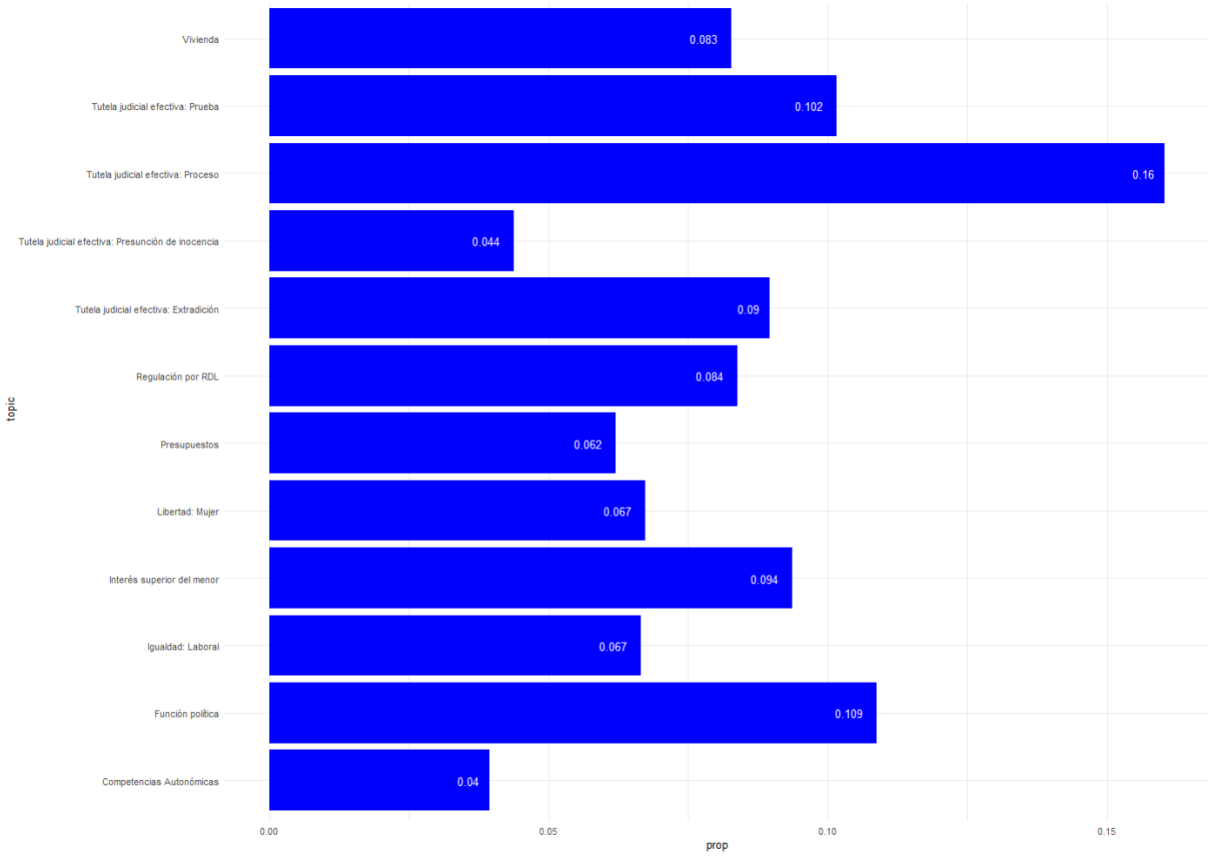

Figura 21: Peso relativo de cada tópico en el año 2024. Fuente: elaboración propia a partir de los datos del caso de estudio

La Figura 21 refleja una evolución temática en las sentencias del Tribunal Constitucional. Cada vez disminuye más la proporción de sentencias relativas a la tutela judicial efectiva pues, aunque sigue siendo predominante, cada vez se observa una mayor proporción de temas más definidos e innovadores. Ello podría indicar un cambio en la naturaleza de los recursos admitidos a trámite por el Tribunal.

Asimismo, aumenta la proporción de temas como la libertad. En contraste, emergen nuevos temas de especial interés, como la vivienda (0.083) o los presupuestos (0.062).

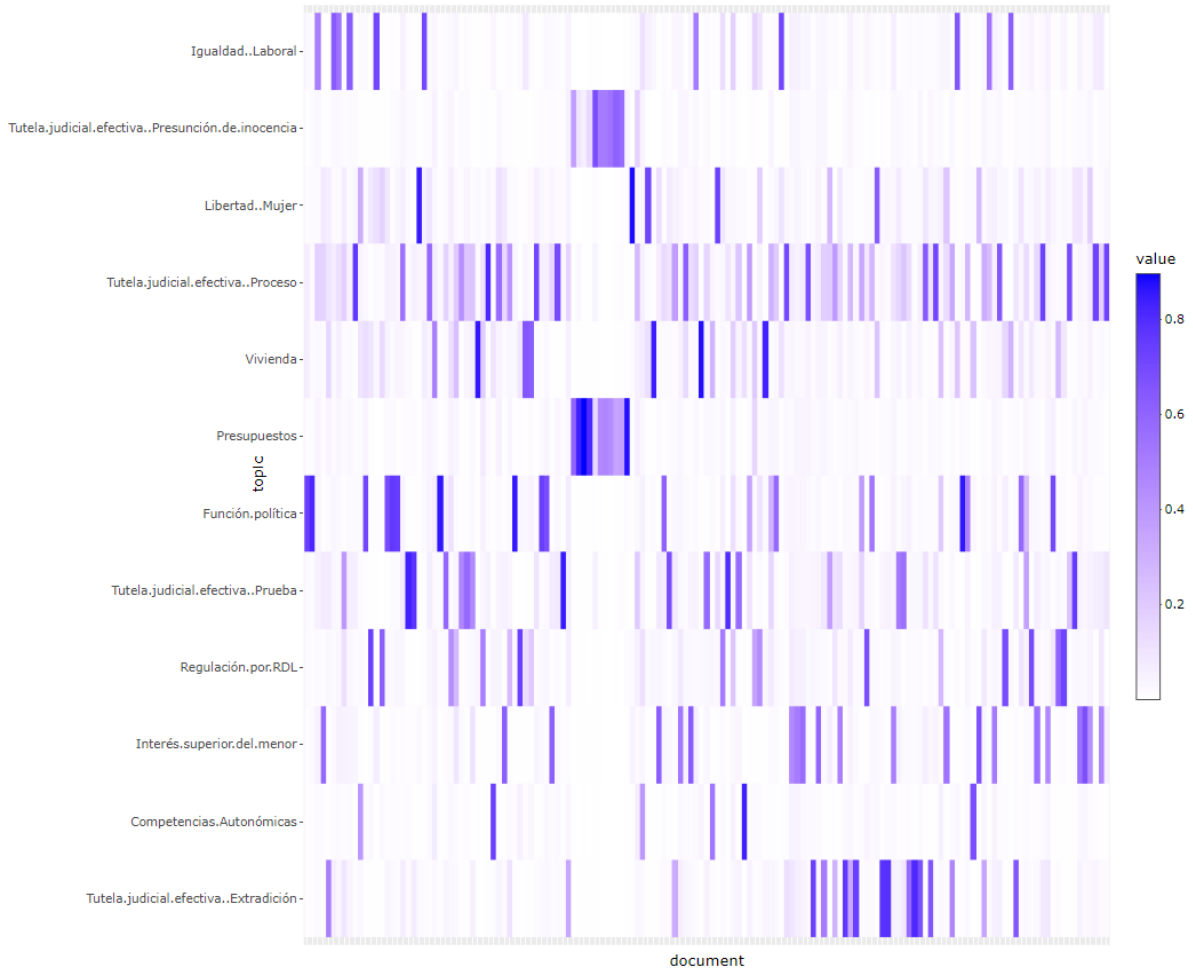

Figura 22: Mapa de calor para el año 2024. Fuente: elaboración propia a partir de los datos del caso de estudio

Finalmente, la Figura 22 muestra que, mientras los términos procesales vinculados a la tutela judicial efectiva aparecen en todas las sentencias de manera constante, ciertos temas se concentran en momentos específicos del año, reflejándose en un gran número de resoluciones simultáneamente. Un claro ejemplo de ello es el caso de los presupuestos, cuyo tratamiento se observa de forma masiva en un punto concreto del año.

6. CONCLUSIONES FINALES

El objetivo general del estudio era determinar el impacto de las reformas en los temas tratados por el Tribunal Constitucional. En este sentido, esta investigación ha permitido ofrecer una visión global sobre cómo ha evolucionado el contenido y la cantidad de sentencias dictadas tras la introducción del requisito de especial trascendencia constitucional.

En primer lugar, los datos confirman que el Tribunal no ha incrementado el número de sentencias dictadas al año en términos cuantitativos. Por el contrario, las medidas adoptadas han generado una reducción en el número total de sentencias. Sin embargo, esta disminución no responde a una reducción en la actividad, sino a un proceso de selección más riguroso. El nuevo criterio ha propiciado la admisión únicamente de aquellos recursos que abordan cuestiones jurídicas más específicas, novedosas o de mayor relevancia constitucional.

En cuanto al contenido de las sentencias, el análisis permite observar tanto la persistencia de ciertos temas como la aparición de otros nuevos. Algunos asuntos se mantienen constantes a lo largo del tiempo, sin importar el momento histórico o normativo. Tal es el caso de la regulación por Real Decreto-ley, la presunción de inocencia, la tutela judicial efectiva vinculada a la prueba y los derechos de los funcionarios públicos. Estos temas forman parte del núcleo habitual de recursos admitidos por el Tribunal y reflejan preocupaciones jurídicas transversales y permanentes.

No obstante, también se detecta la aparición de cuestiones puntuales que surgen en momentos determinados, posiblemente impulsadas por el componente innovador que el criterio de especial trascendencia constitucional favorece. Así, en 2005 destaca la presencia de sentencias relativas a extranjería y tratados internacionales; en 2009, cobran protagonismo los derechos vinculados a la libertad de expresión, el honor y la intimidad; mientras que en 2024 emergen asuntos como el derecho a la vivienda, los presupuestos y el interés superior del menor. Estos

temas reflejan no solo el contexto jurídico del momento, sino también preocupaciones sociales específicas.

Asimismo, algunos tópicos muestran una evolución significativa en su enfoque. Por ejemplo, la libertad sindical en 2005 se centraba en la defensa del derecho de asociación; en cambio, en 2024 el debate gira en torno a la mejora de condiciones laborales reclamadas desde los sindicatos, lo que evidencia una madurez del derecho reconocido. De manera similar, dentro de la tutela judicial efectiva surgen nuevas líneas jurisprudenciales, como aquellas relacionadas con la extradición o con delitos como la malversación, que en 2024 adquieren una mayor presencia.

Esta evolución temática puede atribuirse tanto al elemento discrecional que posee el Tribunal Constitucional para seleccionar qué recursos admitir a trámite como al propio devenir histórico y social. En este sentido, las sentencias analizadas no solo son reflejo del ordenamiento jurídico vigente, sino también del momento social en el que se dictan. La jurisprudencia del Tribunal podría configurarse como un reflejo de los intereses jurídicos y sociales dominantes en cada etapa.

Por otra parte, este estudio también demuestra la aplicabilidad de técnicas de procesamiento de datos para llevar a cabo un análisis de las sentencias del Tribunal Constitucional. A pesar de que no existe mucha literatura al respecto, estas técnicas permiten profundizar en el contenido de las sentencias, aportando un punto innovador a la estadística judicial.

Del mismo modo, se plantea una vía de trabajo útil para los profesionales de la abogacía, que pueden ampliar sus capacidades de análisis adoptando herramientas de inteligencia artificial. En un entorno jurídico cada vez más complejo, estas tecnologías permiten abordar fenómenos normativos y jurisprudenciales con mayor precisión.

Entre las limitaciones del estudio, cabe destacar que se ha centrado únicamente en las sentencias admitidas a trámite, lo cual deja fuera un elemento clave del proceso de selección: las demandas que no superan el filtro de admisión. Sería especialmente interesante, en futuras investigaciones, aplicar esta misma metodología a los recursos desestimados en fase de admisión para valorar si los temas tratados en ambas categorías coinciden. Lo ideal sería que existiera coherencia entre ellos, lo cual permitiría comprobar el rigor del criterio de especial trascendencia constitucional. Además, esta metodología podría aplicarse en el futuro al análisis de sentencias de otras jurisdicciones nacionales o internacionales, ampliando así su utilidad comparativa.

Durante el desarrollo del trabajo se presentaron ciertos retos técnicos, especialmente derivados del procesamiento de un corpus extenso y formado por sentencias de considerable complejidad textual. En este sentido, la elección de solo tres años de estudio —uno por cada momento clave— respondió a criterios de viabilidad técnica. Ampliar el rango temporal podría haber ofrecido una visión más matizada de la evolución jurisprudencial, pero resultaba inviable dentro de los márgenes de procesamiento asumibles para este trabajo.

En definitiva, este estudio ofrece una primera aproximación a los efectos reales del requisito de especial trascendencia constitucional sobre el funcionamiento del Tribunal Constitucional, tanto en términos cuantitativos como cualitativos. Al mismo tiempo, plantea nuevas preguntas y líneas de investigación que pueden enriquecer el debate académico y profesional en torno a la evolución del amparo constitucional en España.

7. ANEXOS

Anexo I: Código completo utilizado en RStudio relativo al año 2005

```
options(encoding="utf8") #Cambiar el encoding del ordenador

install.packages("rvest")

install.packages("quanteda")

install.packages("quanteda.textstats")

install.packages("quanteda.textplots")

install.packages("dplyr")

install.packages("stopwords")

install.packages("LDAvis")

install.packages("slam")

install.packages("wordcloud")

install.packages("RColorBrewer")

install.packages("gt")

library(rvest)

library(quanteda)

library(quanteda.textstats)

library(quanteda.textplots)

library(wordcloud2)

library(RColorBrewer)
```

```
library(dplyr)

library(udpipe)

library(text2vec)

library(ggplot2)

library(lattice)

library(reshape2)

library(tidytext)

library(tm)

library(stopwords)

library(topicmodels)

library(LDAvis)

library(slam)

library(servr)

library(plotly)

library(reshape2)

library(plotly)

library(openxlsx)

library(gt)

get_online_text <- function(url) {

  doc <- read_html(url)  # Leer la página web
```

```
  wikitext <- doc %>%
    html_nodes("p, h4") %>%  # Captura tanto párrafos <p> como encabezados <h4>
    html_text(trim = TRUE) %>%
    paste(sep = " ", collapse = " ") %>%  # Unifica todo el texto
    gsub("\"", "", .) %>%  # Elimina comillas dobles
    gsub("\\[[0-9]*\\]", "", .)  # Elimina referencias como [34]
  return(wikitext)
}
get_online_text
stc <- get_online_text("https://hj.tribunalconstitucional.es/HJ/es/Resolucion/Show/5286")
stc
# Crear una lista vacía para almacenar las sentencias
sentencias_2005 <- list()
# Bucle para recorrer las sentencias del año 2005
for (num_sentencia in 5602:5261) {
  url <- paste0("https://hj.tribunalconstitucional.es/HJ/es/Resolucion/Show/", num_sentencia)
  # Obtener el texto de la sentencia
  texto_sentencia <- get_online_text(url)
  # Almacenar la sentencia en la lista
  sentencias_2005[[as.character(num_sentencia)]] <- texto_sentencia
```

```
}

# Verificar el contenido de sentencias_2005

str(sentencias_2005)

# Crear un dataframe vacío donde almacenar las sentencias

df_sentencias <- data.frame(

  IDENTIFICADOR = character(),

  FUNDAMENTOS_JURIDICOS = character(),

  FALLO = character(),

  stringsAsFactors = FALSE

)

# Función para extraer la información de cada sentencia

extraer_info_sentencia <- function(sentencia) {

  # Extraer IDENTIFICADOR desde el ECLI

  identificador_match <- regexpr("ECLI:ES:TC:2005:[0-9]+", sentencia)

  if (identificador_match[1] != -1) {

    ecli <- regmatches(sentencia, identificador_match)

    numero_identificador <- gsub("ECLI:ES:TC:2005:", "", ecli)

    identificador <- paste0(numero_identificador, "/2005")

  } else {

    identificador <- "No encontrado"
```

```
  }

  # Extraer FUNDAMENTOS JURÍDICOS

  fundamentos_inicio <- regexpr("II\\. Fundamentos jurídicos", sentencia)

  fallo_inicio <- regexpr("\\bFallo\\b", sentencia)  # Busca la palabra "Fallo" como palabra completa

  if (fundamentos_inicio[1] != -1) {

    if (fallo_inicio[1] != -1) {

      fundamentos_juridicos <- substring(sentencia, fundamentos_inicio[1], fallo_inicio[1] - 1)

    } else {

      fundamentos_juridicos <- substring(sentencia, fundamentos_inicio[1])  # Si no se encuentra el fallo, tomar hasta el final

    }

  } else {

    fundamentos_juridicos <- "No encontrado"

  }

  # Extraer FALLO desde "Fallo" hasta el final

  if (fallo_inicio[1] != -1) {

    fallo <- substring(sentencia, fallo_inicio[1])

  } else {

    fallo <- "No encontrado"
```

```
  }

  # Devolver un dataframe con la sentencia procesada

  return(data.frame(

    IDENTIFICADOR = identificador,

    FUNDAMENTOS_JURIDICOS = fundamentos_juridicos,

    FALLO = fallo,

    stringsAsFactors = FALSE

  ))

}

# Recorrer cada sentencia en el objeto 'sentencias_2005' y añadirla al dataframe

for (num_sentencia in names(sentencias_2005)) {

  sentencia <- sentencias_2005[[num_sentencia]]

  df_sentencia <- extraer_info_sentencia(sentencia)

  df_sentencias <- rbind(df_sentencias, df_sentencia)

}

# Escribir el dataframe en un archivo Excel

write.xlsx(df_sentencias, "sentencias_2005.xlsx")

# Verificar la estructura del dataframe

str(df_sentencias)

#Tokenización
```

```
ud_model_download <- udpipe_download_model(language = "spanish")

ud_model <- udpipe_load_model(ud_model_download$file_model)

corpus<-df_sentencias$FUNDAMENTOS_JURIDICOS

corpus_tokenizado <- udpipe_annotate(ud_model, x = corpus)

corpus_tokenizado_data_frame <- as.data.frame(corpus_tokenizado)

head(corpus_tokenizado_data_frame)

corpus_tokenizado_data_frame$lemma<-tolower(corpus_tokenizado_data_frame$lemma)

unique(corpus_tokenizado_data_frame$upos)

corpus_no_PUNCT_SYM_NUM <- subset(corpus_tokenizado_data_frame, !(upos %in% c("PUNCT", "SYM", "NUM", "X", "PART")))

corpus_int<-subset(corpus_no_PUNCT_SYM_NUM,!grepl("^[&';@[:digit:]]", lemma))

corpus_int$lemma<-gsub("\\-", "", corpus_int$lemma)

corpus_int$lemma<-gsub("\\#", "", corpus_int$lemma)

#StopWords

library(stopwords)

stopwords <- stopwords("es", source = "snowball")

print(stopwords)

stopwords<-c(stopwords, "recurso","jurídico", "vulneración", "resolución", "derecho","amparo","hecho", "juzgado", "procedimiento", "instancia", "sentencia","demanda","fundamento","stc","general")
```

```
corpus_nostopwords<-subset(corpus_int, !lemma %in% c(stopwords))

corpus_nostopwords<-subset(corpus_nostopwords, !token %in% c(stopwords))

bigramas <- keywords_collocation(corpus_nostopwords, term = "lemma", group = c("doc_id",
"sentence_id"), ngram_max = 2, n_min = 10)

head(bigramas)

bigramas <- bigramas[bigramas$pmi >= 3,]

bigramas$key <- factor(bigramas$keyword, levels = rev(bigramas$keyword))

barchart(key ~ pmi, data = head(subset(bigramas, freq>3), 20), col = "blue",

        main = "Colocaciones con bigramas", xlab = "PMI")

corpus_nostopwords$term <- corpus_nostopwords$lemma

corpus_nostopwords$term<-txt_recode_ngram(corpus_nostopwords$lemma,       compound=
bigramas$keyword, ngram = bigramas$ngram, sep = " ")

corpus_nostopwords <- corpus_nostopwords[nchar(corpus_nostopwords$term)>3, ]

corpus_nostopwords <- subset(corpus_nostopwords, upos %in% c("NOUN", "ADJ",

                                        "PROPN"))

#Analisis exploratorio - Metrica TF-IDF

dtf <- document_term_frequencies(corpus_nostopwords, document = "doc_id",term = "term")

dtm <- document_term_matrix(dtf)

dfm_corpus <- as.dfm(dtm)
```

```
dtf %>%
  head(20) %>%  # Mostrar solo las primeras filas
  gt() %>%
  tab_header(title = "Frecuencia de términos por documento") %>%
  cols_label(
    doc_id = "Documento",
    term = "Término",
    freq = "Frecuencia"
  ) %>%
  fmt_number(columns = "freq", decimals = 0)
# Agrupar los términos por documento
textos_por_doc <- corpus_nostopwords %>%
  group_by(doc_id) %>%
  summarise(texto = paste(term, collapse = " "))
# Crear el corpus
corpus_q <- corpus(textos_por_doc, text_field = "texto", docid_field = "doc_id")
# Crear la Document-Feature Matrix
dfm_q <- dfm(corpus_q)
# Calcular TF-IDF
dfm_tfidf <- dfm_tfidf(dfm_q)
```

```
# Extraer los 20 términos con mayor tf-idf global
top_tfidf_global <- topfeatures(dfm_tfidf, n = 20)
# Convertir a dataframe para ggplot
df_top <- data.frame(
  term = names(top_tfidf_global),
  tfidf = as.numeric(top_tfidf_global)
)
# Crear el gráfico
ggplot(df_top, aes(x = reorder(term, tfidf), y = tfidf)) +
  geom_col(fill = "tomato") +
  coord_flip() +
  labs(title = "Top 20 términos con mayor TF-IDF global",
       x = "Término",
       y = "TF-IDF") +
  theme_minimal()
# Convertir dfm_tfidf a tidy
df_tfidf_tidy <- tidy(dfm_tfidf)
# Agrupar por término y sumar el TF-IDF en todos los documentos
df_top <- df_tfidf_tidy %>%
  group_by(term) %>%
```

```
  summarise(tfidf = sum(count)) %>%
  arrange(desc(tfidf)) %>%
  filter(tfidf > 0)  # Asegurar que todas las palabras tengan peso positivo
wordcloud2(data = df_top, size = 0.7, color = "random-dark", backgroundColor = "white")
#Coherencia - Numero de topics
range <- seq(2, 40, by = 2)
tcm <- crossprod(as.matrix(dtm))
coherence_logratio <- data.frame()
for (i in range) {
  coherenceloop_logratio <- data.frame()  # Inicialización dentro del bucle
  for (j in seq(1, 5, by = 1)) {
    topicmodeling <- LDA(dtm, method = "Gibbs", k = i, iter = 200, burnin = 100,
                         control = list(alpha = 50/i, delta = 0.1), initialize = "random")    words_topic <-
terms(topicmodeling, k = 10)
    words_topic_matrix  <-  as.matrix(words_topic)          coherenceloop_logratio[j,  1]  <-
mean(coherence(words_topic_matrix, tcm,
                                    metrics  =  c("mean_logratio"),  smooth  =  1,  n_doc_tcm  =
nrow(tcm)))
  }
  coherence_logratio[i, 1] <- i
```

```
  coherence_logratio[i, 2] <- mean(coherenceloop_logratio[, 1])  # Acceder a la primera columna correctamente
}
colnames(coherence_logratio) <- c("Número de tópicos", "Coherencia UMass")
ggplot(data = coherence_logratio, aes(x = `Número de tópicos`, y = `Coherencia UMass`)) +
  geom_line(color = 'blue', size = 1) +
  labs(x = "Número de tópicos", y = "Coherencia UMass") +
  theme_minimal()
#LDA
dfm_quanteda<-as.dfm(dtm)
dfm_trimmed<-dfm_trim(dfm_quanteda, min_docfreq = 0.005, max_docfreq = 0.99, docfreq_type="prop")
dtm=convert(dfm_trimmed, to="topicmodels")
i=12 #número de temas
modelo <- LDA(dtm, method = "Gibbs", k =i, control = list(alpha = 50/i,
                                                  delta=0.1, seed=58), initialize="random")
phi <- data.frame(as.matrix(posterior(modelo)$terms))
phi$row_name <- (rownames(phi))
topicos <- melt(phi)
colnames(topicos) <- c("topic", "term", "phi")
```

```
topicos$topic<-as.numeric(topicos$topic)

top_terms <- topicos %>%

  group_by(topic) %>%

  top_n(15,phi) %>%

  ungroup() %>%

  arrange(topic,-phi)

plot_topic <- top_terms %>%

  mutate(term = reorder_within(term, phi, topic)) %>%

  ggplot(aes(term, phi, fill = factor(topic))) +

  geom_col(show.legend = FALSE) +

  facet_wrap(~ topic, scales = "free") +

  coord_flip() +

  scale_x_reordered()

plot_topic

topicNames<-c("Indemnización", "Regulación por RDL", "Tutela judicial efectiva: Presunción de inocencia", "Competencias Autonómicas", "Tutela judicial efectiva: Indefensión", "Igualdad", "Tutela judicial efectiva: Prueba", "Penal", "Función política", "Tutela judicial efectiva: Inadmisión", "Tratados y Convenios Internacionales", "Libertad sindical" )

phi <- as.matrix(posterior(modelo)$terms)

theta <- as.matrix(posterior(modelo)$topics)
```

```
#Encontramos documentos con mucha presencia en el tópico 9

topico_9 <- theta[, 9]

top_docs <- order(topico_9, decreasing = TRUE)

top_docs[1:5]

fundamentos_251 <- df_sentencias$FUNDAMENTOS_JURIDICOS[251]

#Encontramos documentos con mucha presencia en el tópico 11

topico_11 <- theta[, 11]

top_docs <- order(topico_11, decreasing = TRUE)

top_docs[1:5]

fundamentos_27 <- df_sentencias$FUNDAMENTOS_JURIDICOS[27]

vocab <- colnames(phi)

doc.length <- as.vector(table(dtm$i))

term.frequency <- col_sums(dtm)

json_lda <- createJSON(phi = phi, theta = theta, vocab = vocab, doc.length =

                       doc.length, term.frequency = term.frequency)

serVis(json_lda)

body<-df_sentencias

exampleIds<-1:nrow(body)

exampleIds <- exampleIds[exampleIds <= nrow(theta)]

N <- length(exampleIds)
```

```
topicProportionExamples <- theta[exampleIds,]
colnames(topicProportionExamples) <- topicNames
vizDataFrame <- melt(cbind(data.frame(topicProportionExamples), document =
                           factor(1:N)), variable.name = "topic", id.vars = "document")
plot<-ggplot(vizDataFrame, aes(document, topic,
                               fill=value))+geom_tile()+scale_fill_gradient(low="white",
                                                                            high="blue")+theme(axis.ticks.x =
element_blank(), axis.text.x =
                                                                                element_blank())
ggplotly(plot)
topicproportions <- colSums(theta) / nrow(dtm)
names(topicproportions) <- topicNames
topicproportions<-sort(topicproportions, decreasing = TRUE)
proportions <- data.frame(topic = names(topicproportions), prop =
                            as.numeric(topicproportions))
ggplot(proportions, aes(x=prop, y=topic))+geom_bar(stat="identity", fill="blue")+
geom_text(aes(label = round(prop,3)), hjust = 1.5,color="white") +theme_minimal()
```

Anexo II: Código completo utilizado en RStudio relativo al año 2009

```
options(encoding="utf8") #Cambiar el encoding del ordenador

install.packages("rvest")

install.packages("quanteda")

install.packages("quanteda.textstats")

install.packages("quanteda.textplots")

install.packages("dplyr")

install.packages("stopwords")

install.packages("LDAvis")

install.packages("slam")

library(rvest)   # rvest para hacer webscraping con R

library(quanteda)  # contains all of the core natural language processing and textual data management functionss

library(quanteda.textstats)  # statistics for textual data

library(quanteda.textplots)  # plots for textual data

library(wordcloud2)

library(RColorBrewer)

library(dplyr)

library(udpipe)

library(text2vec)
```

```
library(ggplot2)

library(lattice)

library(reshape2)

library(tidytext)

library(tm)

library(stopwords)

library(topicmodels)

library(LDAvis)

library(slam)

library(servr)

library(plotly)

library(reshape2)

library(plotly)

library(openxlsx)

library(gt)

get_online_text <- function(url) {

  doc <- read_html(url)  # Leer la página web

  wikitext <- doc %>%

    html_nodes("p, h4") %>%  # Captura tanto párrafos <p> como encabezados <h4>

    html_text(trim = TRUE) %>%
```

```
    paste(sep = " ", collapse = " ") %>%  # Unifica todo el texto
    gsub("\"", "", .) %>%  # Elimina comillas dobles
    gsub("\\[[0-9]*\\]", "", .)  # Elimina referencias como [34]
  return(wikitext)
}
# Crear una lista vacía para almacenar las sentencias
sentencias_2009 <- list()
# Bucle para recorrer las sentencias del año 2009
for (num_sentencia in 6639:6420) {
  url <- paste0("https://hj.tribunalconstitucional.es/HJ/es/Resolucion/Show/", num_sentencia)
  # Obtener el texto de la sentencia
  texto_sentencia <- get_online_text(url)
  # Almacenar la sentencia en la lista
  sentencias_2009[[as.character(num_sentencia)]] <- texto_sentencia
}
# Verificar el contenido de sentencias_2009
str(sentencias_2009)
# Crear un dataframe vacío donde almacenar las sentencias
df_sentencias <- data.frame(
  IDENTIFICADOR = character(),
```

```
  FUNDAMENTOS_JURIDICOS = character(),
  FALLO = character(),
  stringsAsFactors = FALSE
)
# Función para extraer la información de cada sentencia
extraer_info_sentencia <- function(sentencia) {
  # Extraer IDENTIFICADOR desde el ECLI
  identificador_match <- regexpr("ECLI:ES:TC:2009:[0-9]+", sentencia)
  if (identificador_match[1] != -1) {
    ecli <- regmatches(sentencia, identificador_match)
    numero_identificador <- gsub("ECLI:ES:TC:2009:", "", ecli)
    identificador <- paste0(numero_identificador, "/2009")
  } else {
    identificador <- "No encontrado"
  }
  # Extraer FUNDAMENTOS JURÍDICOS
  fundamentos_inicio <- regexpr("II\\. Fundamentos jurídicos", sentencia)
  fallo_inicio <- regexpr("\\bFallo\\b", sentencia)  # Busca la palabra "Fallo" como palabra completa
  if (fundamentos_inicio[1] != -1) {
```

```
    if (fallo_inicio[1] != -1) {
      fundamentos_juridicos <- substring(sentencia, fundamentos_inicio[1], fallo_inicio[1] - 1)
    } else {
      fundamentos_juridicos <- substring(sentencia, fundamentos_inicio[1])  # Si no se encuentra el fallo, tomar hasta el final
    }
  } else {
    fundamentos_juridicos <- "No encontrado"
  }
  # Extraer FALLO desde "Fallo" hasta el final
  if (fallo_inicio[1] != -1) {
    fallo <- substring(sentencia, fallo_inicio[1])
  } else {
    fallo <- "No encontrado"
  }
  # Devolver un dataframe con la sentencia procesada
  return(data.frame(
    IDENTIFICADOR = identificador,
    FUNDAMENTOS_JURIDICOS = fundamentos_juridicos,
    FALLO = fallo,
```

```
    stringsAsFactors = FALSE
  ))
}
# Recorrer cada sentencia en el objeto 'sentencias_2009' y añadirla al dataframe
for (num_sentencia in names(sentencias_2009)) {
  sentencia <- sentencias_2009[[num_sentencia]]
  df_sentencia <- extraer_info_sentencia(sentencia)
  df_sentencias <- rbind(df_sentencias, df_sentencia)
}
# Escribir el dataframe en un archivo Excel
write.xlsx(df_sentencias, "sentencias_2009.xlsx")
# Verificar la estructura del dataframe
str(df_sentencias)
#Tokenización
ud_model_download <- udpipe_download_model(language = "spanish")
ud_model <- udpipe_load_model(ud_model_download$file_model)
corpus<-df_sentencias$FUNDAMENTOS_JURIDICOS
corpus_tokenizado <- udpipe_annotate(ud_model, x = corpus)
corpus_tokenizado_data_frame <- as.data.frame(corpus_tokenizado)
head(corpus_tokenizado_data_frame)
```

```
corpus_tokenizado_data_frame$lemma<-tolower(corpus_tokenizado_data_frame$lemma)

unique(corpus_tokenizado_data_frame$upos)

corpus_no_PUNCT_SYM_NUM <- subset(corpus_tokenizado_data_frame, !(upos %in% c("PUNCT", "SYM", "NUM", "X", "PART")))

corpus_int<-subset(corpus_no_PUNCT_SYM_NUM,!grepl("^[&';@[:digit:]]", lemma))

corpus_int$lemma<-gsub("\\-", "", corpus_int$lemma)

corpus_int$lemma<-gsub("\\#", "", corpus_int$lemma)

#StopWords

library(stopwords)

stopwords <- stopwords("es", source = "snowball")

print(stopwords)

stopwords<-c(stopwords, "recurso","jurídico", "vulneración", "resolución", "derecho","amparo","hecho", "juzgado", "procedimiento", "instancia", "sentencia","demanda","fundamento","stc","general")

corpus_nostopwords<-subset(corpus_int, !lemma %in% c(stopwords))

corpus_nostopwords<-subset(corpus_nostopwords, !token %in% c(stopwords))

bigramas <- keywords_collocation(corpus_nostopwords, term = "lemma", group = c("doc_id", "sentence_id"), ngram_max = 2, n_min = 10)

head(bigramas)

bigramas <- bigramas[bigramas$pmi >= 3,]

bigramas$key <- factor(bigramas$keyword, levels = rev(bigramas$keyword))
```

```
barchart(key ~ pmi, data = head(subset(bigramas, freq>3), 20), col = "blue",
        main = "Colocaciones con bigramas", xlab = "PMI")
corpus_nostopwords$term <- corpus_nostopwords$lemma
corpus_nostopwords$term<-txt_recode_ngram(corpus_nostopwords$lemma, compound=
bigramas$keyword, ngram = bigramas$ngram, sep = " ")
corpus_nostopwords <- corpus_nostopwords[nchar(corpus_nostopwords$term)>3, ]
corpus_nostopwords <- subset(corpus_nostopwords, upos %in% c("NOUN", "ADJ",
                                                    "PROPN"))
#Analisis exploratorio - Metrica TF-IDF
dtf <- document_term_frequencies(corpus_nostopwords, document = "doc_id",term = "term")
dtm <- document_term_matrix(dtf)
dfm_corpus <- as.dfm(dtm)
# Agrupar los términos por documento
textos_por_doc <- corpus_nostopwords %>%
  group_by(doc_id) %>%
  summarise(texto = paste(term, collapse = " "))
# Crear el corpus
corpus_q <- corpus(textos_por_doc, text_field = "texto", docid_field = "doc_id")
# Crear la Document-Feature Matrix
tokens_q <- tokens(corpus_q)   # Tokenizamos el corpus
```

```
dfm_q <- dfm(tokens_q)        # Creamos la DFM desde los tokens

# Calcular TF-IDF

dfm_tfidf <- dfm_tfidf(dfm_q)

# Extraer los 20 términos con mayor tf-idf global

top_tfidf_global <- topfeatures(dfm_tfidf, n = 20)

# Convertir a dataframe para ggplot

df_top <- data.frame(

  term = names(top_tfidf_global),

  tfidf = as.numeric(top_tfidf_global)

)

# Crear el gráfico

ggplot(df_top, aes(x = reorder(term, tfidf), y = tfidf)) +

  geom_col(fill = "tomato") +

  coord_flip() +

  labs(title = "Top 20 términos con mayor TF-IDF global",

       x = "Término",

       y = "TF-IDF") +

  theme_minimal()

# Convertir dfm_tfidf a tidy

df_tfidf_tidy <- tidy(dfm_tfidf)
```

```
# Agrupar por término y sumar el TF-IDF en todos los documentos
df_top <- df_tfidf_tidy %>%
  group_by(term) %>%
  summarise(tfidf = sum(count)) %>%
  arrange(desc(tfidf)) %>%
  filter(tfidf > 0)  # Asegurar que todas las palabras tengan peso positivo
wordcloud2(data = df_top, size = 0.7, color = "random-dark", backgroundColor = "white")
#Coherencia - Número de Topics
range <- seq(2, 40, by = 2)
tcm <- crossprod(as.matrix(dtm))
coherence_logratio <- data.frame()
for (i in range) {
  coherenceloop_logratio <- data.frame()  # Inicialización dentro del bucle
  for (j in seq(1, 5, by = 1)) {
    topicmodeling <- LDA(dtm, method = "Gibbs", k = i, iter = 200, burnin = 100,
                         control = list(alpha = 50/i, delta = 0.1), initialize = "random")
    words_topic <- terms(topicmodeling, k = 10)
    words_topic_matrix <- as.matrix(words_topic)
    coherenceloop_logratio[j, 1] <- mean(coherence(words_topic_matrix, tcm,
```

```
                                    metrics = c("mean_logratio"), smooth = 1, n_doc_tcm =
nrow(tcm)))
  }
  coherence_logratio[i, 1] <- i
  coherence_logratio[i, 2] <- mean(coherenceloop_logratio[, 1])  # Acceder a la primera
columna correctamente
}
colnames(coherence_logratio) <- c("Número de tópicos", "Coherencia UMass")
ggplot(data = coherence_logratio, aes(x = `Número de tópicos`, y = `Coherencia UMass`)) +
  geom_line(color = 'blue', size = 1) +
  labs(x = "Número de tópicos", y = "Coherencia UMass") +
  theme_minimal()
#LDA
dfm_quanteda<-as.dfm(dtm)
dfm_trimmed<-dfm_trim(dfm_quanteda, min_docfreq = 0.005, max_docfreq = 0.99,
docfreq_type="prop")
dtm=convert(dfm_trimmed, to="topicmodels")
i=12 #número de temas
modelo <- LDA(dtm, method = "Gibbs", k =i, control = list(alpha = 50/i,
                                    delta=0.1, seed=58), initialize="random")
```

```
phi <- data.frame(as.matrix(posterior(modelo)$terms))
phi$row_name <- (rownames(phi))
topicos <- melt(phi)
colnames(topicos) <- c("topic", "term", "phi")
topicos$topic<-as.numeric(topicos$topic)
top_terms <- topicos %>%
  group_by(topic) %>%
  top_n(15,phi) %>%
  ungroup() %>%
  arrange(topic,-phi)
plot_topic <- top_terms %>%
  mutate(term = reorder_within(term, phi, topic)) %>%
  ggplot(aes(term, phi, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip() +
  scale_x_reordered()
plot_topic
topicNames<-c("Recurso de casación", "Igualdad: Mujer", "Regulación por RDL", "Tutela judicial efectiva: Indefensión", "Tutela judicial efectiva: Inadmisión", "Libertad sindical",
```

"Función política", "Tutela judicial efectiva: Presunción de inocencia", "Tutela judicial efectiva: Prueba", "Libertad de expresión, honor e intimidad", "Igualdad: Función política", "Competencias del Congreso")

```
phi <- as.matrix(posterior(modelo)$terms)

theta <- as.matrix(posterior(modelo)$topics)

vocab <- colnames(phi)

doc.length <- as.vector(table(dtm$i))

term.frequency <- col_sums(dtm)

json_lda <- createJSON(phi = phi, theta = theta, vocab = vocab, doc.length =

                    doc.length, term.frequency = term.frequency)

serVis(json_lda)

body<-df_sentencias

exampleIds<-1:nrow(body)

exampleIds <- exampleIds[exampleIds <= nrow(theta)]

N <- length(exampleIds)

topicProportionExamples <- theta[exampleIds,]

colnames(topicProportionExamples) <- topicNames

vizDataFrame <- melt(cbind(data.frame(topicProportionExamples), document =

                        factor(1:N)), variable.name = "topic", id.vars = "document")

plot<-ggplot(vizDataFrame, aes(document, topic,
```

```
                    fill=value))+geom_tile()+scale_fill_gradient(low="white",
                                                    high="blue")+theme(axis.ticks.x =
element_blank(), axis.text.x =
                                                        element_blank())
ggplotly(plot)
topicproportions <- colSums(theta) / nrow(dtm)
names(topicproportions) <- topicNames
topicproportions<-sort(topicproportions, decreasing = TRUE)
proportions <- data.frame(topic = names(topicproportions), prop =
                    as.numeric(topicproportions))
ggplot(proportions, aes(x=prop, y=topic))+geom_bar(stat="identity", fill="blue")+
geom_text(aes(label = round(prop,3)), hjust = 1.5,color="white") +theme_minimal()
```

Anexo III: Código completo utilizado en RStudio relativo al año 2024

```
options(encoding="utf8")

install.packages("rvest")

install.packages("quanteda")

install.packages("quanteda.textstats")

install.packages("quanteda.textplots")

install.packages("dplyr")

install.packages("stopwords")

install.packages("LDAvis")

install.packages("slam")

install.packages("openxlsx")

library(rvest)

library(quanteda)

library(quanteda.textstats)

library(quanteda.textplots)

library(dplyr)

library(udpipe)

library(text2vec)

library(ggplot2)

library(lattice)
```

```
library(reshape2)

library(tidytext)

library(tm)

library(stopwords)

library(topicmodels)

library(LDAvis)

library(slam)

library(servr)

library(plotly)

library(reshape2)

library(plotly)

library(openxlsx)

# Función para extraer el texto de la sentencia desde la URL

get_online_text <- function(url) {

  doc <- read_html(url)  # Leer la página web

  wikitext <- doc %>%

    html_nodes("p, h4") %>%  # Captura tanto párrafos <p> como encabezados <h4>

    html_text(trim = TRUE) %>%

    paste(sep = " ", collapse = " ") %>%  # Unifica todo el texto

    gsub("\"", "", .) %>%  # Elimina comillas dobles
```

```
    gsub("\\[[0-9]*\\]", "", .)  # Elimina referencias como [34]
  return(wikitext)
}
# Lista para almacenar las sentencias válidas
sentencias_2024 <- list()
# Bucle para recorrer las sentencias del año 2024
for (num_sentencia in 31365:30062) {
  url <- paste0("https://hj.tribunalconstitucional.es/HJ/es/Resolucion/Show/", num_sentencia)
  texto_sentencia <- get_online_text(url)
  # Verificar si la sentencia tiene un ECLI con terminación ":A"
  if (grepl("^ECLI:ES:TC:2024:\\d{1,3}A", texto_sentencia)) {
    print(paste("Descartando AUTO:", num_sentencia))
  } else if (grepl("^ECLI:ES:TC:2024:\\d{1,3}", texto_sentencia)) {
    sentencias_2024[[as.character(num_sentencia)]] <- texto_sentencia
    print(paste("Sentencia válida almacenada:", num_sentencia))
  }
}
# Mostrar cuántas sentencias válidas se almacenaron
print(paste("Total de sentencias almacenadas:", length(sentencias_2024)))
# Crear un dataframe vacío donde almacenar las sentencias
```

```
df_sentencias <- data.frame(
  IDENTIFICADOR = character(),
  FUNDAMENTOS_JURIDICOS = character(),
  FALLO = character(),
  stringsAsFactors = FALSE
)
# Función para extraer la información de cada sentencia
extraer_info_sentencia <- function(sentencia) {
  # Extraer IDENTIFICADOR desde el ECLI
  identificador_match <- regexpr("ECLI:ES:TC:2024:[0-9]+", sentencia)
  if (identificador_match[1] != -1) {
    ecli <- regmatches(sentencia, identificador_match)
    numero_identificador <- gsub("ECLI:ES:TC:2024:", "", ecli)
    identificador <- paste0(numero_identificador, "/2024")
  } else {
    identificador <- "No encontrado"
  }
  # Extraer FUNDAMENTOS JURÍDICOS
  fundamentos_inicio <- regexpr("II\\. Fundamentos jurídicos", sentencia)
```

```
  fallo_inicio <- regexpr("\\bFallo\\b", sentencia)  # Busca la palabra "Fallo" como palabra completa
  if (fundamentos_inicio[1] != -1) {
    if (fallo_inicio[1] != -1) {
      fundamentos_juridicos <- substring(sentencia, fundamentos_inicio[1], fallo_inicio[1] - 1)
    } else {
      fundamentos_juridicos <- substring(sentencia, fundamentos_inicio[1])  # Si no se encuentra el fallo, tomar hasta el final
    }
  } else {
    fundamentos_juridicos <- "No encontrado"
  }
  # Extraer FALLO desde "Fallo" hasta el final
  if (fallo_inicio[1] != -1) {
    fallo <- substring(sentencia, fallo_inicio[1])
  } else {
    fallo <- "No encontrado"
  }
  # Devolver un dataframe con la sentencia procesada
  return(data.frame(
```

```
    IDENTIFICADOR = identificador,
    FUNDAMENTOS_JURIDICOS = fundamentos_juridicos,
    FALLO = fallo,
    stringsAsFactors = FALSE
  ))
}
# Recorrer cada sentencia en el objeto 'sentencias_2024' y añadirla al dataframe
for (num_sentencia in names(sentencias_2024)) {
  sentencia <- sentencias_2024[[num_sentencia]]
  df_sentencia <- extraer_info_sentencia(sentencia)
  df_sentencias <- rbind(df_sentencias, df_sentencia)
}
# Escribir el dataframe en un archivo Excel
write.xlsx(df_sentencias, "sentencias_2024.xlsx")
# Verificar la estructura del dataframe
str(df_sentencias)
#Tokenización
ud_model_download <- udpipe_download_model(language = "spanish")
ud_model <- udpipe_load_model(ud_model_download$file_model)
corpus<-df_sentencias$FUNDAMENTOS_JURIDICOS
```

```
corpus_tokenizado <- udpipe_annotate(ud_model, x = corpus)

corpus_tokenizado_data_frame <- as.data.frame(corpus_tokenizado)

head(corpus_tokenizado_data_frame)

corpus_tokenizado_data_frame$lemma<-tolower(corpus_tokenizado_data_frame$lemma)

unique(corpus_tokenizado_data_frame$upos)

corpus_no_PUNCT_SYM_NUM <- subset(corpus_tokenizado_data_frame, !(upos %in% c("PUNCT", "SYM", "NUM", "X", "PART")))

corpus_int<-subset(corpus_no_PUNCT_SYM_NUM,!grepl("^[&';:@[:digit:]]", lemma))

corpus_int$lemma<-gsub("\\-", "", corpus_int$lemma)

corpus_int$lemma<-gsub("\\#", "", corpus_int$lemma)

#StopWords

library(stopwords)

stopwords <- stopwords("es", source = "snowball")

print(stopwords)

stopwords<-c(stopwords, "recurso","jurídico", "vulneración", "resolución", "derecho","amparo","hecho", "juzgado", "procedimiento", "instancia", "sentencia","demanda","fundamento","stc","general")

corpus_nostopwords<-subset(corpus_int, !lemma %in% c(stopwords))

corpus_nostopwords<-subset(corpus_nostopwords, !token %in% c(stopwords))

bigramas <- keywords_collocation(corpus_nostopwords, term = "lemma", group = c("doc_id", "sentence_id"), ngram_max = 2, n_min = 10)
```

```
head(bigramas)

bigramas <- bigramas[bigramas$pmi >= 3,]

bigramas$key <- factor(bigramas$keyword, levels = rev(bigramas$keyword))

barchart(key ~ pmi, data = head(subset(bigramas, freq>3), 20), col = "blue",

        main = "Colocaciones con bigramas", xlab = "PMI")

corpus_nostopwords$term <- corpus_nostopwords$lemma

corpus_nostopwords$term<-txt_recode_ngram(corpus_nostopwords$lemma,       compound=

bigramas$keyword, ngram = bigramas$ngram, sep = " ")

corpus_nostopwords <- corpus_nostopwords[nchar(corpus_nostopwords$term)>3, ]

corpus_nostopwords <- subset(corpus_nostopwords, upos %in% c("NOUN", "ADJ",

                                                "PROPN"))

#Analisis exploratorio - Metrica TF-IDF

dtf <- document_term_frequencies(corpus_nostopwords, document = "doc_id",term = "term")

dtm <- document_term_matrix(dtf)

dfm_corpus <- as.dfm(dtm)

# Agrupar los términos por documento

textos_por_doc <- corpus_nostopwords %>%

  group_by(doc_id) %>%

  summarise(texto = paste(term, collapse = " "))

# Crear el corpus
```

```
corpus_q <- corpus(textos_por_doc, text_field = "texto", docid_field = "doc_id")

# Crear la Document-Feature Matrix

tokens_q <- tokens(corpus_q)   # Tokenizamos el corpus

dfm_q <- dfm(tokens_q)         # Creamos la DFM desde los tokens

# Calcular TF-IDF

dfm_tfidf <- dfm_tfidf(dfm_q)

# Extraer los 20 términos con mayor tf-idf global

top_tfidf_global <- topfeatures(dfm_tfidf, n = 20)

# Convertir a dataframe para ggplot

df_top <- data.frame(

  term = names(top_tfidf_global),

  tfidf = as.numeric(top_tfidf_global)

)

# Crear el gráfico

ggplot(df_top, aes(x = reorder(term, tfidf), y = tfidf)) +

  geom_col(fill = "tomato") +

  coord_flip() +

  labs(title = "Top 20 términos con mayor TF-IDF global",

       x = "Término",

       y = "TF-IDF") +
```

```
  theme_minimal()

# Convertir dfm_tfidf a tidy

df_tfidf_tidy <- tidy(dfm_tfidf)

# Agrupar por término y sumar el TF-IDF en todos los documentos

df_top <- df_tfidf_tidy %>%

  group_by(term) %>%

  summarise(tfidf = sum(count)) %>%

  arrange(desc(tfidf)) %>%

  filter(tfidf > 0)  # Asegurar que todas las palabras tengan peso positivo

wordcloud2(data = df_top, size = 0.7, color = "random-dark", backgroundColor = "white")

#Coherencia - Numero de topics

range <- seq(2, 40, by = 2)

tcm <- crossprod(as.matrix(dtm))

coherence_logratio <- data.frame()

for (i in range) {

  coherenceloop_logratio <- data.frame()  # Inicialización dentro del bucle

  for (j in seq(1, 5, by = 1)) {

    topicmodeling <- LDA(dtm, method = "Gibbs", k = i, iter = 200, burnin = 100,

                         control = list(alpha = 50/i, delta = 0.1), initialize = "random")

    words_topic <- terms(topicmodeling, k = 10)
```

```
    words_topic_matrix <- as.matrix(words_topic)
    coherenceloop_logratio[j, 1] <- mean(coherence(words_topic_matrix, tcm,
                                        metrics = c("mean_logratio"), smooth = 1, n_doc_tcm =
nrow(tcm)))
  }
  coherence_logratio[i, 1] <- i
  coherence_logratio[i, 2] <- mean(coherenceloop_logratio[, 1])  # Acceder a la primera
columna correctamente
}
colnames(coherence_logratio) <- c("Número de tópicos", "Coherencia UMass")
ggplot(data = coherence_logratio, aes(x = `Número de tópicos`, y = `Coherencia UMass`)) +
  geom_line(color = 'blue', size = 1) +
  labs(x = "Número de tópicos", y = "Coherencia UMass") +
  theme_minimal()
#LDA
dfm_quanteda<-as.dfm(dtm)
dfm_trimmed<-dfm_trim(dfm_quanteda, min_docfreq = 0.005, max_docfreq = 0.99,
docfreq_type="prop")
dtm=convert(dfm_trimmed, to="topicmodels")
i=12 #número de temas
```

```
modelo <- LDA(dtm, method = "Gibbs", k =i, control = list(alpha = 50/i,
                                        delta=0.1, seed=58), initialize="random")
phi <- data.frame(as.matrix(posterior(modelo)$terms))
phi$row_name <- (rownames(phi))
topicos <- melt(phi)
colnames(topicos) <- c("topic", "term", "phi")
topicos$topic<-as.numeric(topicos$topic)
top_terms <- topicos %>%
  group_by(topic) %>%
  top_n(15,phi) %>%
  ungroup() %>%
  arrange(topic,-phi)
plot_topic <- top_terms %>%
  mutate(term = reorder_within(term, phi, topic)) %>%
  ggplot(aes(term, phi, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip() +
  scale_x_reordered()
plot_topic
```

```
topicNames<-c("Tutela judicial efectiva: Extradición", "Competencias Autonómicas", "Interés superior del menor", "Regulación por RDL", "Tutela judicial efectiva: Prueba", "Función política", "Presupuestos", "Vivienda", "Tutela judicial efectiva: Proceso", "Libertad: Mujer", "Tutela judicial efectiva: Presunción de inocencia", "Igualdad: Laboral" )

phi <- as.matrix(posterior(modelo)$terms)

theta <- as.matrix(posterior(modelo)$topics)

vocab <- colnames(phi)

doc.length <- as.vector(table(dtm$i))

term.frequency <- col_sums(dtm)

json_lda <- createJSON(phi = phi, theta = theta, vocab = vocab, doc.length =

                       doc.length, term.frequency = term.frequency)

serVis(json_lda)

body<-df_sentencias

exampleIds<-1:nrow(body)

exampleIds <- exampleIds[exampleIds <= nrow(theta)]

N <- length(exampleIds)

topicProportionExamples <- theta[exampleIds,]

colnames(topicProportionExamples) <- topicNames

vizDataFrame <- melt(cbind(data.frame(topicProportionExamples), document =

                           factor(1:N)), variable.name = "topic", id.vars = "document")
```

```
plot<-ggplot(vizDataFrame, aes(document, topic,
                    fill=value))+geom_tile()+scale_fill_gradient(low="white",
                                                high="blue")+theme(axis.ticks.x =
element_blank(), axis.text.x =
                                                    element_blank())
ggplotly(plot)
topicproportions <- colSums(theta) / nrow(dtm)
names(topicproportions) <- topicNames
topicproportions<-sort(topicproportions, decreasing = TRUE)
proportions <- data.frame(topic = names(topicproportions), prop =
                    as.numeric(topicproportions))
ggplot(proportions, aes(x=prop, y=topic))+geom_bar(stat="identity", fill="blue")+
geom_text(aes(label = round(prop,3)), hjust = 1.5,color="white") +theme_minimal()
```

Anexo IV: Gráficos complementarios a las nubes de palabras

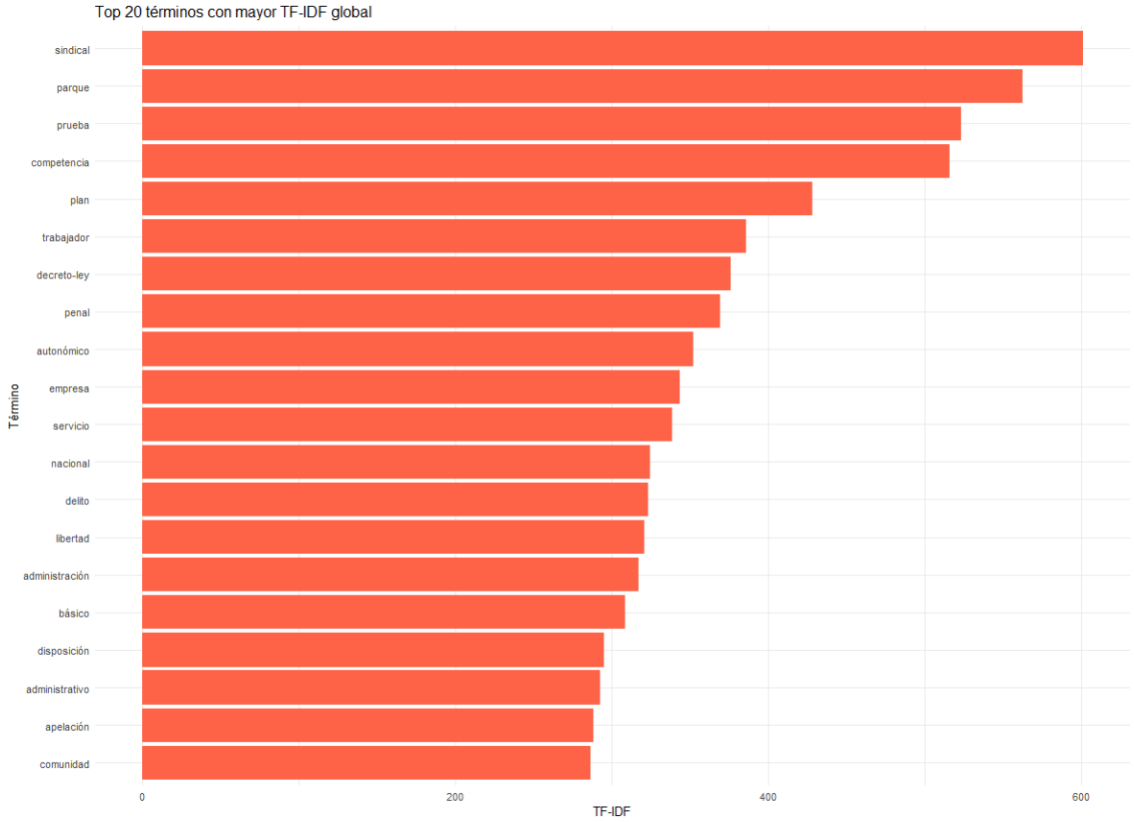

Figura 23: Términos más relevantes del 2005 con la métrica TF-IDF. Fuente: elaboración propia a partir de los datos del caso de estudio

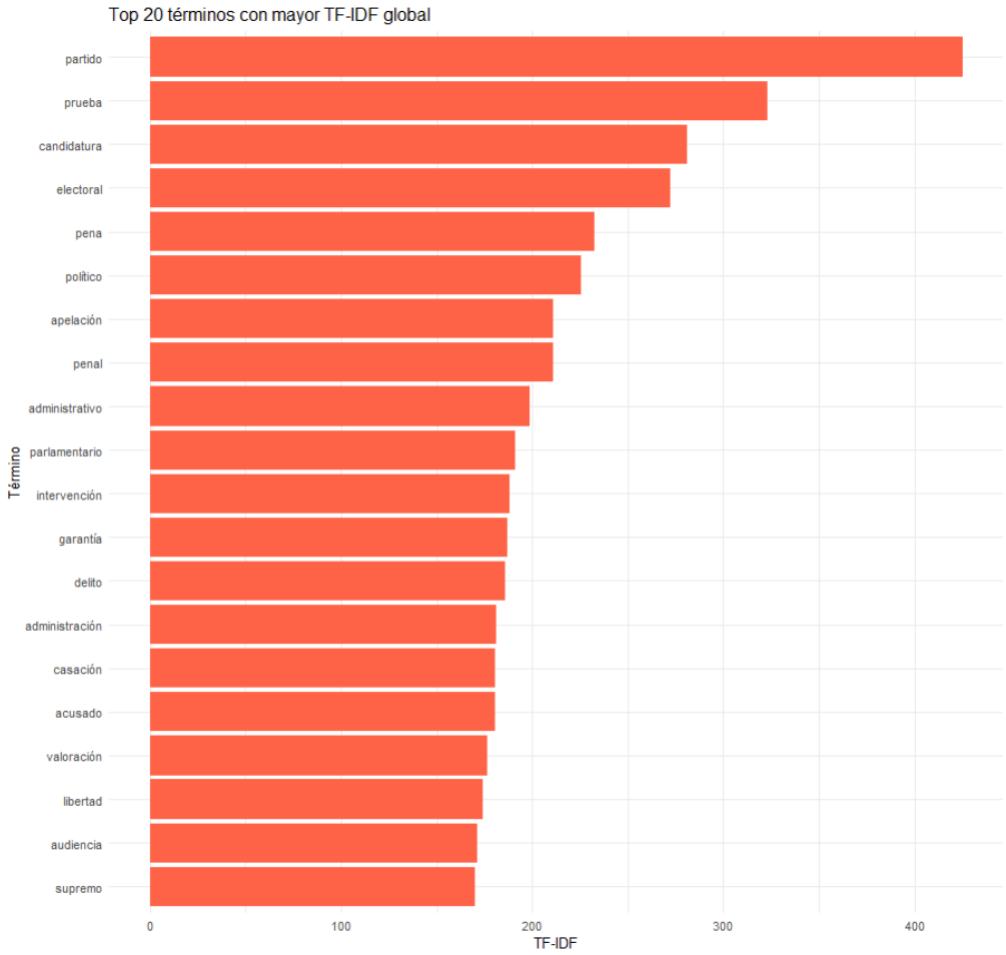

Figura 24: Términos más relevantes del 2009 con la métrica TF-IDF. Fuente: elaboración propia a partir de los datos del caso de estudio

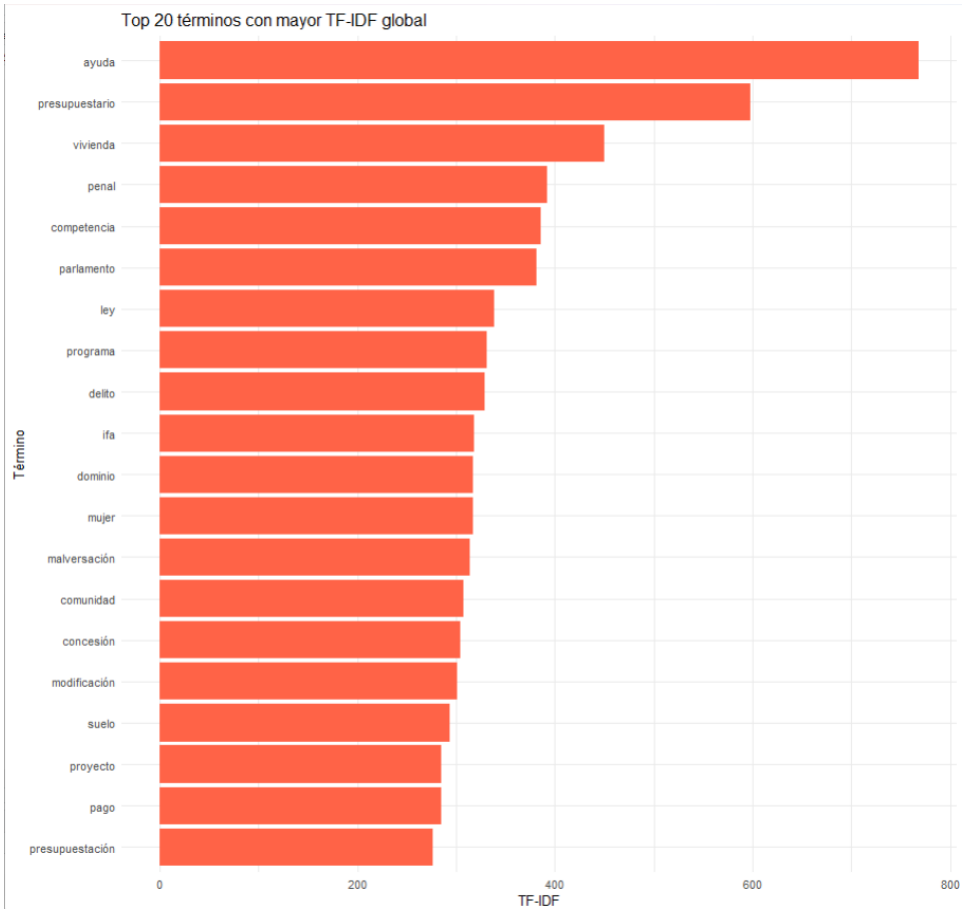

Figura 25: Términos más relevantes del 2024 con la métrica TF-IDF. Fuente: elaboración propia a partir de los datos del caso de estudio

Anexo V: Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

Por la presente, yo, Inés Vinuesa Armada, estudiante de E3 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Aplicación de técnicas de Text Mining para evaluar el impacto de las reformas legales en las sentencias del Tribunal Constitucional", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Metodólogo: Para descubrir métodos aplicables a problemas específicos de investigación.
2. Interpretador de código: Para realizar análisis de datos preliminares.
3. Estudios multidisciplinares: Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
4. Corrector de estilo literario y de lenguaje: Para mejorar la calidad lingüística y estilística del texto.
5. Sintetizador y divulgador de libros complicados: Para resumir y comprender literatura compleja.
6. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
7. Traductor: Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares).

Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 7 de abril de 2025

Firmado: Inés Vinuesa Armada

8. REFERENCIAS

Abay Mersha, M., Gemeda Yigezu, M. y Kalita J. (2024). "Semantic-Driven Topic Modeling Using Transformer-Based Embeddings and Clustering Algorithms", *Procedia Computer Science*, 244, pp. 121-132 https://doi.org/10.1016/j.procs.2024.10.185

Alaminos-Fernández, Antonio F (2023) "Introducción a la minería de texto y análisis de sentimiento con R". *Universidad de Alicante. Obets Ciencia Abierta*. Alicante: Limencop.

Auto 252/2009, de 19 de octubre, Tribunal Constitucional. https://hj.tribunalconstitucional.es/eu-ES/Resolucion/Show/22234

Blei, M.D., Ng, A.Y., Jordan, M.I. (2003). "Latent Dirichlet Allocation", Jornal of Machine Learning Research, 3. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Bues, M.M, Matthaei E. (2017). "LegalTech on the rise: technology changes legal work behaviors, but does not replace its profession" en Jacob, K. et al., "Liquid Legal. Management for Professionals", *Springer*.

Cabañas García, J.C. (2010). "El trámite de admisión del recurso de amparo y su especial trascendencia constitucional". En Montoya Melgar A., "Cuestiones actuales de la jurisdicción en España", pp. 253-273, *Dykinson*, Madrid.

Corrales, M., Fenwick, M. y Haapio, H. (2019). "Digital Tecnologies, Legal Design and the Future of the Legal Profession" en *Legal Tech, Smart Contracts and Blockchain*, Springer, p.9-10.

Dale R. (2018). "Industry Watch. Law and Word Order: NLP in Legal Tech, Natural Lenguage Engineering", *Cambridge*, 25, p. 211-217 https://doi.org/10.1017/S1351324918000475

Dieng, A.B., Ruiz, F.J.R., Blei, D.M. (2019). “Topic Modeling in Embedding Spaces”. *Cornell University. Arxiv*. https://arxiv.org/abs/1907.04907

Diez-Picazo, L.M. (1994). “Dificultades prácticas y significado constitucional del recurso de amparo”. *Revista Española de Derecho Constitucional*, 40, pp. 9-14. https://www.jstor.org/stable/44203495

Figuerola, C.G, Zazo A.F., Rodriguez Vázquez de Alana, E., Alonso Berrocal, J.L. (2004). La Recuperación de Información en español y la normalizacion de términos. *Revista Iberoamericana de Inteligencia Artificial*, 22(8), pp.135-145. http://www.redalyc.org/articulo.oa?id=92582210

Finck M. (2019). “Blockchain Regulation and Governance in Europe”, *Cambridge*, United Kingdom, pp. 6-8.

García de Enterría, E. (2014). “La posición jurídica del Tribunal Constitucional en el sistema español: posibilidades y perspectivas”. *Revista Española de Derecho Constitucional*, 100, 39-131. https://www.jstor.org/stable/24886783

García-Pelayo, M. (2014). “El “status” del Tribunal Constitucional”. *Revista Española de Derecho Constitucional*, 100, 17-37. https://www.jstor.org/stable/44202689

Gil Pascual, J.A. (2021). “Minería de texto con R. Aplicaciones y técnicas estadísticas de apoyo”. *Universidad Nacional de Educación a Distancia*. Madrid.

González Beilfuss, M. (2018). “La especial trascendencia constitucional como criterio de selección de los recursos de amparo”. *Anuario de la Facultad de Derecho de la Universidad Autónoma de Madrid*, 257-280.https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-A-2018-10025700280

Griffiths, T., & Steyvers, M. (2004). “Finding Scientific Topics.” *Proceedings of the National Academy of Sciences of the United States of America*, 101(1), 5228-5235. doi:10.1073/pnas.0307752101

Grootendorst, M. (2022). “BERTopic:Neural topic modeling with a class-based TF-IDF procedure”. *Cornell University*. *Arxiv*. https://arxiv.org/abs/2203.05794

Gudín Rodríguez-Magariños, F. (2022). “Criptoactivos: de la paralegalidad a la paulativa legalización”, *Sepín*, Madrid, p.78.

Gurcan, F., & Cagiltay, N. E. (2019). “Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling”. *IEEE access*, 7, 82541-82552.

Hernández, C.L. y Rodríguez J.E. (2008). “Preprocesamiento de datos estructurados”, *Investigación y Desarrollo*, 4 (2), pp.28-32. https://doi.org/10.14483/2322939X.4123

Hernández Ramos, M. (2016). “Incumplimiento de la buena administración de justicia del Tribunal Constitucional en la admisión del recurso de amparo. El caso Arribas Antón vs. España del TEDH”. *Revista Española de Derecho Constitucional*, 108, 307-355. http://dx.doi.org/10.18042/cepc/redc.108.10

Jain, S., Jain, S. K., y Vasal, S. (2024). “An Effective TF-IDF Model to Improve the Text Classification Performance”. *IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 1-4. doi: 10.1109/CSNT60213.2024.10545818

Kulhari, S. (2018). “The Midas touch of Blockchain: Leveraging it for Data Protection”. En “Building-Blocks of a Data Protection Revolution: The Uneasy Case for Blockchain

Technology to Secure Privacy and Identity", pp. 15–22. *Nomos Verlagsgesellschaft mbH*. http://www.jstor.org/stable/j.ctv941qz6.6

Kumar S., Kumar Kar A. y Ilavarasan P.V. (2021). "Appplications of text mining in services management: A systematic literature review". *International Journal of Information Management Data Insights*, 1(1). https://doi.org/10.1016/j.jjimei.2021.100008

López Guerra, L. (2008). "Las sentencias básicas del Tribunal Constitucional". Boletín Oficial del Estado. *Centro de Estudios Políticos y Constitucionales*. p. 117 https://books.google.es/books?hl=es&lr=&id=VKv7DwAAQBAJ&oi=fnd&pg=PA2&dq=estructura+sentencias+espa%C3%B1a+TC&ots=f7JxOOke5Z&sig=BD0oS8R0NbyNvxQQOrb5HP0cWj4#v=onepage&q&f=false

López-Nozal, C., Garcıa-Osorio, C. I., Arnaiz-González, A., y Juez-Gil, M. (2018). "Aplicaciones de la técnica de topic model en repositorios software". *IX Simposio Teoría y Aplicaciones de Minería de Datos*. https://sci2s.ugr.es/caepia18/proceedings/docs/CAEPIA2018_paper_210.pdf

Matia Portilla, F.J. (2009). "La especial trascendencia constitucional y la inadmisión del recursos de amparo". *Revista Española de Derecho Constitucional*, 86, 350-368. https://www.jstor.org/stable/24885989?read-now=1&seq=1#page_scan_tab_contents

Mielnik, D. (2022). "Análisis computacional del Derecho Argentino", *Revista Argentina de Teoría Jurídica*, 23(1). https://inteligencialegal.com.ar/_app/immutable/assets/Mielnik-ACDA.Bq-xolfe.pdf

Mielnik, D. y Altszyler, E. (2023). "Inteligencia Artificial aplicada al estudio del derecho: análisis computacional de la jurisprudencia de casación penal" en Corvalán, J.G. (2021), "Tratado de Inteligencia Artificial y Derecho", *Thomson Reuters*. https://inteligencialegal.com.ar/investigacion

Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. (2011). “Optimizing semantic coherence in topic models”. *Proceedings of the conference on empirical methods in natural language processing*, pp. 262–272, https://aclanthology.org/D11-1024.Pdf

Omar, M., On, B.-W., Lee, I., & Choi, G. S. (2015). “LDA topics: Representation and evaluation”. *Journal of Information Science*, 41(5), pp. 662-675. https://doi.org/10.1177/0165551515587839

Manzano, M. P. (2004). “La extradición: una institución ‘constitucional’”. *Revista de derecho penal y criminología*, 2, 213-242. https://revistas.uned.es/index.php/RDPC/article/download/24894/19752

Pérez Tremps, P. (2004). “La reforma del recurso de amparo”. *Tirant Lo Blanch*, Valencia, 538-544. https://dialnet.unirioja.es/descarga/articulo/1125414.pdf

Ponweiser, M. (2012). “Latent Dirichlet Allocation in R.” *WU Vienna University of Economics and Business - Institute for Statistics and Mathematics*. https://research.wu.ac.at/ws/files/18975608/main.pdf

Rawat, R., Ghildiyal, R., & Dixit, A. (2022). “Topic modeling of legal documents using NLP and bidirectional encoder representations from transformers”. Indonesian *Journal of Electrical Engineering and Computer Science*, 28(3), pp. 1749-1755. DOI: 10.11591/ijeecs.v28.i3.pp1749-1755

Reyes-Ortiz, J.A, Paniagua-Reyes, F., Sánchez L. (2017). “Minería de opiniones centrada en tópicos usando textos cortos en español”, *Research in Computing Science*, 134, pp. 155-165.

Rodriguez Puñal, E. (2023). “Nuevas reglas para presentar recursos de amparo constitucional”. *Actualidad Jurídica Aranzadi*, 996, 13-15. https://www.legaltoday.com/revista-aja/996/

Rubio Llorente, F (1988). “La jurisdicción constitucional como forma de creación de derecho”. *Revista Española de Derecho Constitucional*, 22 https://www.jstor.org/stable/44203235

Sakthi Vel, S. (2021). “Pre-Processing techniques of Text Mining using Computational Lingustics and Python Libraries”, *International Conference on Artificial Intelligence and Smart Systems (ICAIS),* pp. 879-884, doi: 10.1109/ICAIS50930.2021.9395924.

Sari, D. K., Kosasih, D., & Indarti, D. (2023). “Clustering and topic modeling of verdicts of narcotics cases using machine learning”. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 27(6), pp. 1168-1174. https://doi.org/10.20965/jaciii.2023.p1168

Santander S.A. y Broadridge Inc. (2018), “Santander y Broadridge utilizan por primera vez tecnología blockchain para votar en una junta general de accionistas”, https://www.santander.com/content/dam/santander-com/es/documentos/historico-notas-de-prensa/2018/05/NP-2018-05-17-Santander%20y%20Broadridge%20utilizan%20por%20primera%20vez%20tecnolog%C3%ADa%20blockchain%20para%20votar%20en%20una%20ju-es.pdf.

Sentencia del Tribunal Constitucional (STC) 155/2009, de 25 de junio, https://hj.tribunalconstitucional.es/es-ES/Resolucion/Show/6574

Sentencia del Tribunal Constitucional (STC) 107/1984, de 23 de noviembre, https://hj.tribunalconstitucional.es/es-ES/Resolucion/Show/360

Shamshiri A., Ryu K.R., y Park J.Y. (2024) “Text mining and natural language processing in construction”. *Automation in Construction*, 158, https://doi.org/10.1016/j.autcon.2023.105200

Shinde, P.P. y Shah S.(2018), “A Review of Machine Learning and Deep Learning Applications”. *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, India, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697857.

Taboada Villamarín, A. (2024). “Big data en ciencias sociales. Una introducción a la automatización de análisis de datos de texto mediante procesamiento de lenguaje natural y aprendizaje automático”, *Revista CENTRA De Ciencias Sociales*, 3(1), 51–75. https://doi.org/10.54790/rccs.51

Torres Díaz, M.C. (2023). “A vueltas con la “especial trascendencia constitucional” como requisito de admisibilidad en los recursos de amparo”. *AIS: Ars Iuris Slmanticensis*, 11, pp. 162-164. https://revistas.usal.es/cuatro/index.php/ais/article/view/31342/29539

Tribunal Constitucional (2025). Estadísticas jurisdiccionales. https://www.tribunalconstitucional.es/es/memorias/Paginas/Cuadros-estadisticos.aspx

Valiente Martínez, P. (2024). “A vueltas con la especial trascendencia constitucional: la problemática de la inadmisión de los recursos de amparo”. *Tutela Judicial Efectiva, Dykinson*, 135-152. https://repositorio.comillas.edu/xmlui/handle/11531/94502

Velilla Cadahía, M.R. (2023). “Guía práctica de implementación de topic modeling en R: Analizando artículos periodísticos sobre el metaverso”, *Repositorio Comillas ICADE*. http://hdl.handle.net/11531/68840

Wasiq, M., Bashar, A., Khan, I., Nyagadza B. (2024). “Unveiling customer engagement dynamics in the metaverse: A retrospective bibliometric and topic modelling investigation”. *Computers in Human Behavior Reports*, 16. https://doi.org/10.1016/j.chbr.2024.100483

Wijffels, J. (2023). Package “udpipe”. Repository CRAN R (https://cran.r-project.org/web/packages/udpipe/udpipe.pdf)

Xie, S. y Feng, Y. (2015). “A Recommendation System Combining LDA and Collaborative Filtering Method for Scenic Spot”, *2nd International Conference on Information Science and Control Engineering*, Shanghai, China, pp. 67-71. https://doi.org/10.1109/ICISCE.2015.24

Yan, F., Xu, N., & Qi, Y. (2009). “Parallel Inference for Latent Dirichlet Allocation on Graphics Processing Units”. *Advances in Neural Information Processing Systems*, 22. https://proceedings.neurips.cc/paper/2009/hash/ed265bc903a5a097f61d3ec064d96d2e-Abstract.html