UNIVERSIDAD DE MURCIA

Escuela de Doctorado

TESIS DOCTORAL

El inglés jurídico: nuevas aproximaciones, métodos y herramientas para su estudio cuantitativo

New approaches, methods, and tools for the quantitative study of Legal English

AUTOR/A Daniel Granados Meroño

DIRECTOR/ES Pascual Cantos Gómez

Ángela Almela Sánchez-Lafuente

2025

UNIVERSIDAD DE MURCIA Escuela de Doctorado

TESIS DOCTORAL

El inglés jurídico: nuevas aproximaciones, métodos y herramientas para su estudio cuantitativo

New approaches, methods and tools for the quantitative study of Legal English

AUTOR/A Daniel Granados Meroño

DIRECTOR/ES Pascual Cantos Gómez
Ángela Almela Sánchez-Lafuente

2025

UNIVERSIDAD DE MURCIA

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TITULO DE DOCTOR/A

Aprobado por la Comisión General de Doctorado el 19 de octubre de 2022.

Yo, D. Daniel Granados Meroño, habiendo cursado el Programa de Doctorado de Artes y Humanidades: Bellas Artes, Literatura, Teología, Traducción e Interpretación y Lingüística General e Inglesa de la Escuela Internacional de Doctorado de la Universidad de Murcia (EIDUM), como autor/a de la tesis presentada para la obtención del título de Doctor/a titulada:

El inglés jurídico: nuevas aproximaciones, métodos y herramientas para su estudio cuantitativo

y dirigida por:

D.: Pascual Cantos Gómez
D.: Ángela Almela Sánchez-Lafuente
D.:

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita, cuando se han utilizado sus resultados o publicaciones.

Del mismo modo, asumo ante la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada, en caso de plagio, de conformidad con el ordenamiento jurídico vigente.

Murcia, a 2 de mayo de 2025

(firma)

GRANADOS MEROÑO, DANIEL (FIRMA)

Firmado digitalmente por GRANADOS MEROÑO, DANIEL (FIRMA)
Fecha: 2025.05.02 11:19:44 +02'00'

Información básica sobre protección de sus datos personales aportados:	
Responsable	Universidad de Murcia. Avenida teniente Flomesta, 5. Edificio de la Convalecencia. 30003; Murcia. Delegado de Protección de Datos: dpd@um.es
Legitimación	La Universidad de Murcia se encuentra legitimada para el tratamiento de sus datos por ser necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento. art. 6.1.c) del Reglamento General de Protección de Datos
Finalidad	Gestionar su declaración de autoría y originalidad
Destinatarios	No se prevén comunicaciones de datos
Derechos	Los interesados pueden ejercer sus derechos de acceso, rectificación, cancelación, oposición, limitación del tratamiento, olvido y portabilidad a través del procedimiento establecido a tal efecto en el Registro Electrónico o mediante la presentación de la correspondiente solicitud en las Oficinas de Asistencia en Materia de Registro de la Universidad de Murcia

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la quinta hoja, después de la portada de la tesis presentada para la obtención del título de Doctor/a.

TABLE OF CONTENTS

LIST OF TABLES, FIGURES AND APPENDICES

FIGURES

TABLES

EXCERPTS

CODE

DATA

SUPPLEMENTARY DOCUMENTS

RESUMEN

La presente tesis doctoral aborda el estudio del inglés jurídico desde una perspectiva multidisciplinar, empírica y metodológicamente renovadora, con el propósito de contribuir a una comprensión más precisa, profunda y sistemáticamente fundamentada de los textos legales británicos, tanto en su dimensión lingüística como argumentativa. En particular, el trabajo se centra en la exploración de nuevas herramientas metodológicas y recursos analíticos aplicables al estudio de los géneros jurídicos del sistema anglosajón, integrando de forma innovadora métodos de la lingüística de corpus, el análisis multidimensional y la teoría de la argumentación. Esta integración metodológica se plantea como una respuesta a la escasa interacción entre disciplinas que, hasta ahora, han abordado el discurso jurídico desde enfoques separados, lo que ha dado lugar a análisis parciales, a menudo cualitativos, y difícilmente generalizables. La tesis se sitúa así en el cruce entre lingüística aplicada, estudios jurídicos, ciencia del lenguaje y tecnología lingüística, con el fin de establecer una base empírica sólida para el estudio del lenguaje legal en inglés.

La investigación parte de un análisis crítico de la situación actual de los estudios sobre el inglés jurídico, evidenciando que, a pesar de la gran cantidad de trabajos sobre su terminología, su traducción o su estructura sintáctica, persisten importantes lagunas metodológicas. Estas lagunas tienen que ver, fundamentalmente, con la ausencia de corpus suficientemente representativos, con la falta de validación estadística de muchas de las hipótesis formuladas en estudios anteriores y con la escasa incorporación de herramientas automáticas que permitan un análisis a gran escala. A ello se añade una desconexión patente entre los estudios lingüísticos sobre el derecho y los desarrollos más recientes en el campo de la teoría de la argumentación, especialmente en lo que respecta a la formalización de esquemas argumentativos y su aplicación en contextos legales reales. La tesis plantea, por tanto, una intervención metodológica que permita

superar estos déficits, proponiendo un enfoque integrador que combine recursos empíricos amplios, modelos teóricos sólidos y herramientas computacionales avanzadas.

En su vertiente empírica, el trabajo parte de la construcción de un corpus legal de gran envergadura, el British Statute Law Corpus (BSLC), que recoge más de diez millones de palabras provenientes de legislación reciente de las distintas jurisdicciones del Reino Unido. Este corpus se ha compilado siguiendo criterios estrictos de representatividad, equilibrio temático y cobertura temporal, e incluye textos legislativos oficiales publicados digitalmente por los parlamentos de Westminster, Escocia e Irlanda del Norte. La construcción del corpus ha supuesto un trabajo técnico minucioso, que ha incluido tareas de extracción automatizada de documentos, normalización textual, segmentación estructural y etiquetado lingüístico con herramientas especializadas como CLAWS y el Multidimensional Analysis Tagger. Este nuevo corpus se ha complementado con el British Law Report Corpus (BLRC), ya existente, que contiene textos jurisprudenciales. De este modo, se ha podido llevar a cabo un estudio comparativo entre los dos grandes géneros del derecho público británico: la legislación y la jurisprudencia, lo que constituye una de las aportaciones más relevantes de este trabajo.

A partir de los datos extraídos de ambos corpus, se ha llevado a cabo un análisis multidimensional, siguiendo la metodología desarrollada por Biber, que permite identificar dimensiones latentes del uso lingüístico en función de la coocurrencia estadística de múltiples variables gramaticales y léxicas. Este análisis se ha realizado sobre una muestra de miles de documentos, procesados y cuantificados automáticamente, lo que ha dado lugar a la extracción de seis dimensiones principales que explican la variación discursiva en el corpus jurídico. Las dimensiones identificadas permiten observar diferencias significativas entre los textos legislativos y jurisprudenciales en términos de modalidad, densidad léxica, uso de la voz pasiva, presencia de conectores discursivos, referencia pronominal y otros indicadores lingüísticos clave. Así, mientras la legislación presenta un estilo

altamente impersonal, denso, normativo y tecnificado, la jurisprudencia se caracteriza por una mayor implicación del emisor, el uso de marcadores de actitud y evidencia, y una estructura argumentativa más explícita, coherente con su función interpretativa dentro del sistema de precedentes del common law.

El valor explicativo del análisis multidimensional va más allá de la mera descripción de diferencias estilísticas. En efecto, las dimensiones extraídas permiten reformular hipótesis anteriores sobre la estructura del lenguaje jurídico, ofreciendo una base empírica para la comparación sistemática entre géneros y facilitando su incorporación en modelos de procesamiento automático del lenguaje. Además, este enfoque permite establecer una tipología funcional de los textos jurídicos, en la que el grado de formalidad, abstracción o argumentatividad se define en términos observables y cuantificables. Esta tipología resulta útil no solo para fines descriptivos, sino también para aplicaciones prácticas en enseñanza, traducción y recuperación de información jurídica, ámbitos en los que la clasificación precisa de textos es fundamental.

En paralelo al estudio estructural, la tesis desarrolla una línea de análisis centrada en la argumentación jurídica. A diferencia de otros enfoques que abordan la argumentación desde una perspectiva exclusivamente lógica o filosófica, este trabajo parte de la premisa de que los argumentos jurídicos se manifiestan en formas lingüísticas concretas, susceptibles de ser descritas y analizadas empíricamente. Para ello, se ha llevado a cabo una colaboración con el Centre for Argumentation Technology de la Universidad de Dundee, en cuyo marco se ha desarrollado un protocolo de anotación de esquemas argumentativos adaptado al contexto del discurso jurídico británico. El protocolo se basa en la teoría de los esquemas argumentativos de Walton, y ha sido aplicado manualmente a una selección de sentencias judiciales, permitiendo identificar estructuras como el argumento por precedentes, el argumento por consecuencias, el argumento por autoridad o el argumento por analogía.

El proceso de anotación manual ha sido seguido de una fase de análisis computacional, en la que se han explorado correlaciones entre los esquemas argumentativos identificados y los rasgos lingüísticos presentes en los textos. Esta correlación ha revelado patrones interesantes que sugieren la posibilidad de automatizar parcialmente la detección de esquemas argumentativos en textos jurídicos. Por ejemplo, el uso frecuente de expresiones modales, condicionales o conectores causales aparece asociado a determinados tipos de argumentos, lo que permite anticipar su presencia con un grado razonable de fiabilidad. Estos hallazgos abren una línea de investigación prometedora hacia la creación de sistemas de minería de argumentos legales, con aplicaciones tanto en el análisis jurisprudencial como en la enseñanza del razonamiento jurídico.

La tesis también discute las implicaciones pedagógicas y tecnológicas de sus resultados. Desde el punto de vista pedagógico, los conocimientos adquiridos permiten diseñar materiales de enseñanza del inglés jurídico más adaptados a las necesidades reales de los estudiantes, especialmente en contextos de formación de traductores, juristas o intérpretes judiciales. El conocimiento de las estructuras recurrentes, las funciones discursivas y los mecanismos argumentativos del discurso legal permite un enfoque más funcional y menos memorístico del aprendizaje, centrado en la comprensión de textos reales y en la adquisición de competencias discursivas. En el ámbito tecnológico, los resultados sientan las bases para el desarrollo de herramientas de apoyo a la lectura, traducción y análisis de textos jurídicos, como sistemas de extracción automática de argumentos, clasificadores de textos legales o asistentes de redacción jurídica. Estas aplicaciones resultan especialmente relevantes en un contexto globalizado, en el que el acceso eficiente a la información legal es una necesidad creciente tanto para profesionales como para ciudadanos.

La tesis reconoce, por supuesto, una serie de limitaciones que delimitan el alcance de sus contribuciones. En primer lugar, el estudio se ha centrado en textos del derecho público británico, dejando fuera géneros igualmente relevantes como los contratos, los dictámenes o la doctrina jurídica, cuya inclusión en futuras investigaciones podría enriquecer considerablemente la caracterización del inglés jurídico. En segundo lugar, la anotación de esquemas argumentativos, aunque metodológicamente sólida, sigue siendo un proceso costoso en términos de tiempo y requiere un conocimiento especializado que limita su escalabilidad. La automatización total de este proceso exigirá el entrenamiento de modelos de aprendizaje supervisado sobre grandes corpus anotados, tarea que excede los límites de la presente investigación, pero que se perfila como una continuación lógica y viable de sus resultados.

En conclusión, esta tesis ofrece una contribución significativa al estudio del inglés jurídico, no solo por sus hallazgos específicos, sino también por la propuesta metodológica que la sustenta. Al integrar técnicas de análisis cuantitativo, teoría de la argumentación y recursos computacionales, se ha logrado una aproximación empírica, replicable y útil al análisis del discurso legal. Se trata, por tanto, de una investigación que no solo describe, sino que propone, que no solo interpreta, sino que construye herramientas, que no solo teoriza, sino que aplica. Su vocación interdisciplinar, empírica y aplicada constituye su principal aportación al campo de la lingüística jurídica, y la sitúa como una referencia para futuros estudios que deseen abordar el lenguaje del derecho desde una perspectiva rigurosa, funcional y tecnológicamente informada.

su enseñanza, traducción e interpretación. De este modo, el trabajo se alinea con las tendencias actuales de la lingüística aplicada y las humanidades digitales, y contribuye a reforzar el papel del análisis del lenguaje como herramienta clave para el acceso, la transparencia y la justicia en el ámbito legal contemporáneo.

Más allá del marco metodológico y los hallazgos específicos, esta tesis también propone una reflexión sobre los fundamentos epistemológicos del estudio del lenguaje jurídico, que tradicionalmente ha oscilado entre una visión puramente normativa y una aproximación exclusivamente crítica. Frente a ambas posiciones, este trabajo defiende una concepción del análisis del discurso jurídico como una disciplina empírica, enraizada en datos observables, pero también sensible a las dimensiones ideológicas, retóricas y persuasivas del texto legal. En este sentido, el enfoque adoptado combina el rigor descriptivo de la lingüística de corpus con la profundidad analítica de la teoría de la argumentación, ofreciendo así una visión integradora que trasciende la dicotomía entre descripción y crítica. Esta visión resulta especialmente relevante en un momento en el que el lenguaje jurídico se encuentra sometido a crecientes demandas de accesibilidad, claridad y transparencia, tanto por parte de la ciudadanía como de los propios operadores jurídicos.

Una de las contribuciones más innovadoras del trabajo reside precisamente en su capacidad para tender puentes entre disciplinas que rara vez han dialogado de forma efectiva. Por un lado, se recupera el potencial explicativo de los modelos formales de la argumentación, mostrando que sus esquemas pueden ser aplicados con éxito al análisis de textos jurídicos reales. Por otro lado, se demuestra que los métodos estadísticos propios de la lingüística de corpus permiten identificar con precisión patrones de uso que se corresponden con funciones discursivas específicas. Esta convergencia metodológica no solo enriquece nuestra comprensión del discurso jurídico, sino que proporciona herramientas concretas para su tratamiento automático, lo que abre la puerta a desarrollos tecnológicos en ámbitos como la justicia digital, la inteligencia artificial aplicada al derecho o la minería de textos legales.

Además, los resultados de esta investigación tienen una clara dimensión internacional. El inglés jurídico británico, aunque enraizado en la tradición del common law, tiene un impacto global debido a su influencia en sistemas jurídicos de numerosos países, su papel como lengua de la diplomacia internacional y su uso habitual en tratados, contratos y arbitrajes internacionales. Por tanto, el conocimiento detallado de sus estructuras lingüísticas y argumentativas no solo reviste interés académico, sino que tiene implicaciones prácticas para juristas, traductores y diplomáticos de todo el mundo. En este contexto, el trabajo que aquí se presenta puede considerarse un primer paso hacia la creación de recursos multilingües que permitan comparar el discurso legal en diferentes lenguas y sistemas jurídicos, lo cual constituye una línea de investigación de gran relevancia en un mundo jurídicamente globalizado, pero lingüísticamente diverso.

Finalmente, cabe destacar que esta tesis no se limita a describir el estado actual del inglés jurídico, sino que también plantea propuestas concretas para su análisis futuro. Entre ellas, se encuentran el desarrollo de sistemas semiautomáticos de anotación argumentativa, la ampliación del corpus con textos de otras jurisdicciones anglófonas como Canadá, Australia o Estados Unidos, y la integración de técnicas de aprendizaje automático para mejorar la detección de patrones discursivos complejos. Asimismo, se propone la creación de materiales didácticos basados en corpus, orientados a la enseñanza del inglés jurídico desde una perspectiva basada en datos reales, y no en manuales prescriptivos o ejemplos artificiales. De este modo, la tesis aspira no solo a comprender mejor el lenguaje del derecho, sino también a transformarlo, contribuyendo a que sea más accesible, más transparente y, en última instancia, más justo.

INTRODUCTION

The study of legal English has been approached from the insight of many fields of research, with different interests, focuses and methods. On the one hand, authors with a background in Translation Studies, English for Specific Purposes (ESP) or even Discourse and Critical Discourse Analysis (CDA) have extensively published on the features, genres and patterns found in the discourse of the legal domain, mainly in English (Alcaraz & Hughes, 2002; Breeze, 2013), but also other languages such as Italian (Garofalo, 2009) or Spanish (Alcaraz et al., 2014). Each with a highlight in a specific aspect of this discourse, but using similar methods, mainly qualitative ones, such as text comparison, translational analysis, lexical analysis or triangulation with some natural language processing techniques.

While these authors attempted to understand the linguistic aspects of these legal texts, law and philosophy scholars had other priorities, such as the description of the logical structure of legal argumentation, or the identification of proper and not so proper (fallacies) types of argumentation. Some of them had theoretical interests, while others (especially in the Anglo-Saxon context) had much more practical concerns, so that prospective lawyers reach sufficient knowledge and expertise in legal argumentation when they become barristers. The theoretical and practical interests that drove these authors made them, similarly, not use any type of quantitative or statistical methods on a regular basis.

On the other hand, linguistics scholars have developed methods for the study of genres, discourse and register variation in (mainly) English, such as Biber's (1988) Multi-dimensional analysis. By creating these new methods, they managed to study discourse drawing on corpora containing thousands of words, that allowed them to rely on more complex statistical methods, such as factor analysis or PCA (Biber et al., 2007; Biber & Conrad, 2019; Marín & Rea Rizzo, 2012; Parodi, 2003). This new approach on the study of linguistics can be clustered as the approach of research called Corpus Linguistics (CL), which in the recent years has been

predominant in the study of language. The methodology shift eased the collaboration with computation scholars, which had been developing computation and machine learning models. This resulted in the emergence of the prominent field of Natural Language Processing (NLP), so relevant in the scope of Artificial Intelligence (AI).

In addition, computer scientists gained interest as well in the argumentation aspect of discourse, so they started to look for patterns and models in the logical structure of arguments that might enable machines to automatically identify and create arguments and fallacies (Lawrence et al., 2020; Lawrence & Reed, 2015; Mochales & Moens, 2009). They also developed tools that assisted scholars to manually annotate corpora with argumentation schemes (Janier et al., 2014; Lawrence et al., 2020).

These approaches and methodologies have not converged in a common and interdisciplinary field attempting to describe what legal discourse is (at least, not in a relevant and consistent way), even if that might come as the obvious assumption. On the contrary, they have developed and evolved as isolated fields. As such, many of the conclusions and common literature background on legal English in Translation Studies or ESP has not been consistently verified by quantitative studies based on large-scale data corpora. Similarly, argumentation theory and mining have not made used of the advances on the automatic detection of POS (Parts-of-Speech) or parsing in order to look for linguistic patterns in the use of one or another argumentation scheme.

This dissertation is structured as follows: Chapter 1 covers exhaustively the literature mentioned above, that is, the state of the art in Legal Discourse, and Argumentation Theory and Argumentation Mining regarding the study of legal English, from a comprehensive, connected and chronological view, so the reader can understand how these are interrelated, even if they have developed almost ignoring each other, and the way in which these studies framed the approach of this

dissertation; Chapter 2 establishes the purposes and justification of the dissertation; Chapter 3 explains the structure, compilation and design of the British Statute Law Corpus (BSLC), which is the corpus compiled by the author for this dissertation; Chapters 4 and 5 cover the multidimensional analysis of legal genres for the study of register variation in Legal English; and Chapter 6 regards the study developed in the Centre for Argumentation Technology. In Chapter 6's study, a series of guidelines for the manual annotation of legal argumentation schemes was developed, a small corpus with argumentation schemes was annotated by an annotation team that followed these guidelines, and, finally, Pearson's correlations were performed so as to find preliminary patterns between legal argumentation schemes, Biber's (2019) linguistic features and Hyland's (2005) metadiscourse devices.

1. LITERATURE REVIEW

1.1. Legal Discourse

Linguistics has extensively studied specialised discourse in an attempt not only to better understand its structure, cultural context and historical evolution, but also to develop more effective teaching and translation methods. Thus, scholars in genre, discourse studies, ESP or Translation Studies have conducted several studies disclosing the common patterns that they found when analysing these texts.

The notion of textual 'genre' is one of the common theoretical backgrounds these authors have drawn on to better approach these phenomena, as these help the author classify their object of study following some common variables (such as the utterer, the recipient or the communicative purpose) (Bhatia, 1993; Biber, 1988; Biber & Conrad, 2019; Swales, 1990). A further explanation of this concept is provided in Section 1.1.1. Thus, the study of language variation in different specialised areas has been framed in the study of different genres. That is, when studying scientific discourse, we will find genres such as academic papers, dissemination texts, documentaries... while if interested in journalistic discourse, we will find genres such as opinion columns, pieces of news, and editorials.

This approach to specialised discourse is especially useful in highly specialised areas such as medicine, economics or engineering, as these genres are found to be (generally) universal in (at least) western-related languages (English, French, German, Italian, etc.). This makes perfect sense if we consider the fact that these disciplines tend to homogenisation due to the historical events and advancements taking place during the 20th century. However, this is not exactly the case of Legal Discourse, as Law is a highly cultural-dependent discipline, whose system, purpose, rules and resources might be completely different depending on

the country it is developed (even when comparing two countries with a shared culture and language) (Alcaraz, 2007; Alcaraz & Hughes, 2015; Orts, 2016).

This additional constraint to the study of Legal Discourse might explain why Linguistics and Philological studies have given limited attention to this specialised discourse, whereas other disciplines with closer relationship with social and political sciences or an urgent interest in professional applications, such as Critical Discourse Analysis (CDA) and Translation Studies, have been more interested in the specificalities of this discourse.

Understanding the cultural and political context in which Legal Discourse and Legal Genres are developed is certainly a good starting point when designing a quantitative study, and that is why the Alcaraz's (2007) book on Legal English is especially useful for this dissertation: Law is primarily a cultural and ideological construct(Orts, 2015) to regulate the relationships between citizens (private law) and between citizens and the state (public law). This adapts to the sociocultural context of the state-nation this legal system belongs to, at least if we think of 'state-nations' from a European point of view. Law and language are as well correlated: the power of law is executed by using language; legal texts are given a 'legal validity,' so what it is said or written is enforceable. Legal language has unique features, and it is not improvised at all, because they are to be 'construed' by judges and lawyers in ways that may change the course of a trial.

Given law is a cultural and ideological construct, we can distinguish different legal systems, as well as types of legal discourse with different features depending on the language and the legal system: we can find on the one hand, the English Common law essentially relying on the 'precedent', that is, judicial decisions ruled by the highest jurisdictions of the country, so-called "case law". On the other hand, the Civil or Continental law (to which Spanish law belongs) relying on codified law, that is, legislation promulgated by a parliament (statutes or acts). This does not mean that statute law does not exist in common law systems, since the Parliament

does proclaim enforceable statutes and civil law systems count as well on case law, but as a secondary source of law.

The most relevant sources of the deductive and empirical common law system are the so-called 'case law' and 'equity.' The former compilation of reasoned arguments (*ratio decidendi*) in judgments ruled by senior judges of the higher courts, in charge of very relevant and challenging to solve cases, settling the precedent (binding precedent) judges in lower courts must follow in latter alike cases. This principle, the so-called stare decisis ("to stay on what has been decided"), is essential for common law and governs the functioning of this system (Alcaraz, 2007, pp. 8–9). The latter (equity) emerged as a 'law of the king,' establishing principles ruled by the king guided by "equity and fairness" regardless of what the law of the time said. These principles started to be protected systematically as 'equitable remedies' by the Court of Chancery. Nowadays, judges still apply these, such as the well-known 'injunction.' (Alcaraz, 2007, pp. 6–7).

In turn, 'statute law' is the legislation promulgated by a parliament, that is, the legislative power of the country. The UK has a series of acts that compose the 'constitutional principles' reflecting the organisation and functioning of the country, as well as its fundamental values and its citizens' rights. These acts are the equivalent to a constitution in countries such as Spain. They are the Magna Carta (1215), the Bill of Rights (1689), the Parliamentary Acts 1911 and 1949; and the Peerage Act (Alcaraz, 2007, pp. 9–13).

Bearing this in mind, we can conclude that the genres of 'judgments' representative of the primary source of law (case law) and 'acts or 'legislation', representing the second most important source of law in the Common Law system, are suitable for a comprehensive insight of what Legal English looks like.

This first section reviewing literature contains a selection of studies showing different approaches, resources and methodologies to the study of Legal Discourse from the linguistics point of view, focusing on the textual genres of judgments and

acts. In Sections 1.1.1 and 1.1.2, many studies dealing with the structural, lexical or morphological features of judgments, legislation, and other legal genres are presented. Secondly, some of the most recently compiled corpora containing legal genres are showed in Section 1.1.3. Lastly, Section 1.1.4 summarises the last trends in the use of quantitative approaches and statistical analysis in linguistics, which mainly do not deal with Legal Discourse.

1.1.1. Genre Studies and (Critical) Discourse Analysis

Álvarez (2008) uses Bhatia's (1993, p. 13) definition of 'genre' for her contrastive analysis of Spanish and English judgments, which approaches this concept from a communicative insight:

> Genres are recognisable communicative events characterized by a set of purposes identified and understood by the members of the professional or academic community in which it regularly occurs. Most often, it is highly structured and conventionalized in terms of their intent, positioning, form, and functional value. These constraints are often exploited by the expert members of the discourse community to achieve private intentions within the framework of socially recognized purposes.

This approach is beneficial since it provides the researcher with the items that must be common in different texts, so they conform to a genre. Following this approach, Swales (1990) clarifies some concepts Bhatia uses in his definition: a "speech community" is any group of speakers of the same linguistic code, while a "discourse community" is a group of experts in the same field of work, complying with the six characteristics that Swales (1990) proposes, namely common public goals, methods of communicating among members, participatory communication methods, genres that define the group, lexis, and a standard of knowledge needed for membership.

Garofalo (2009) applies this framework to determine that discourse communities producing legal genres are attorneys, solicitors and barristers, and judges (the producers of our genre, judgments), having these two groups different

public goals. He also defines four fundamental features distinguishing a genre: (1) a communicative event, in which the use of language is relevant; (2) the communicative goal shared by the discursive community members, which helps us to find the text primary purpose/intention; (3) to what extent patterns are shared by the texts within a single genre, that is, how prototypical a text is about the characteristics of a genre; (4) structure and content constraints, since texts within the same genre will share, apart from the textual focus, conventions typically appearing in this genre (Garofalo, 2009).

This notion has been used in several studies making an in-depth analysis of the discursive features of judgments, acts and other legal genres such as court orders. Furthermore, authors interested in the interpersonal and pragmatical dimension of legal discourse from the approach of CDA have also found useful departing from the notion of genre.

Regarding judgments, there is extensive literature on the structure, pragmatics and lexical bundles that are frequent and characterise them as a genre. Álvarez (2008) considers judgments fulfil the requirements that characterise a genre, when comparing English and Spanish judgments through a quantitative analysis. She draws on the four explicit communicative purposes proposed by Bhatia (1993) to define the genre of judgments: (a) genuine records of the facts relevant to the case and of the reasoned argument and the final ruling; (b) their *ratio decidendi* will be followed by lower courts in future alike cases, especially in legal systems based on case-law namely the Common Law; (c) these judgments are used by jurists as a learning resource and during trials, so they are very likely to be reread or reused in several communicative events; (d) in Common law systems, they reflect trending legal issues, so legal professionals may use them to understand better and construe legal texts.

Judgments have as well a very clear macrostructure that Bhatia (1993) names as 'moves': (I) Case identification (Heading), (II) Facts narration (Facts in

issue), (III) Judges argument, (III.a.) Case history, (III.b) Reasoned argument, (III.c.) Ratio decidendi, (IV) Final ruling.

These moves are usually present in any judgment, even if their extension may vary depending on the stage of the civil or criminal procedure on which the judgment is ruled (Magistrates' Court, Appellate Court, Supreme Court). The move III will also be more critical in Common Law judgments because of the stare decisis principle and the prevalence of case law over enacted law in this legal system. (Álvarez Álvarez, 2008). In turn, features highlighted at a microstructural level in Ruiz Moneva's qualitative analysis (2013) can be a useful starting point for predicting the results of our multidimensional analysis.

These features are the following:

(α) *Scarcity or even total absence of Latinisms and terms of Greek origin*: this may be due to the search for clarity or to the fact that anybody may be involved in legal matters, so the potential addressees may not be familiar with these legal terms.

(β) *A use of both personal and impersonal references when expressing opinions*: "Everyone in the case overlooked the fact that…", "Thus the court did not consider, but ought to have considered this dispute"; "In my view, the judgment is seriously flawed. The judges seem to be more likely to use personal pronouns when there is a controversial question and they feel necessary to admit their responsibility in the standpoint expressed: "For those reasons I would allow the appeal".

(γ) *A use of the first person singular* when the judge adopts full responsibility ("When I refer here to Seller, I am referring to its agents, who conducted all negotiations"), or when they want to admit their limitations when it comes to deciding some aspect ("I cannot resolve this dispute. Nevertheless, […] I must presently proceed on the basis that …"), while they tend to use impersonality or the third person singular when contrarily they express objectivity, i.e. when expressing

the conclusion they have reached on the evidence available ("What is evident from that is that….") or when they want to attribute the responsibility to somebody else ("The Appellant submits that the Judge was not applying the principle correctly"), what, apart from the willing to gain clarity, decrease politeness (Ruiz Moneva, 2013, pp. 83–87).

On the contrary, there are not any works providing a complete insight into the characteristics of acts, legislative provisions, laws, or legislation (not even an agreement in the term to name it), but several articles focusing on one specifical feature or dimension of the discourse from legislative acts, one specifical type of legislation (EU directives or regulations) or a mere attempt to translate them and find lexical equivalents. This fragmentary literature might be due to the fact that in Common Law the most relevant and prominent source of law is case law (and, therefore, judgments):

In an attempt to provide translation and linguistics scholars with an accessible explanation of the legal system and genres, Alcaraz (2007) defines the macrostructure of British acts, giving usual lexical bundles, examples and explaining the purpose of each of the parts in a typical British act. In the third part, (preamble) the so-called 'enacting words' are contained, which consists of a fixed and archaic phrase that provides the act with an enforceable force (*'Be enacted by the Queen / King […] as follows:*).

Bhatia was one of the most relevant authors interested in Genre Analysis, as he provided with a comprehensive definition of the concept 'genre' (see above), including legal discourse. In fact, he focused on the description of the discourse in legislative provisions and legal cases (Bhatia, 1993): the one for legal cases was used by Álvarez in her review of the genre of judgments, but the one for legal provisions did not find that much interest in Translation scholars, as this genre does not need translation that frequently, but a brief reminder of it is certainly useful for the purposes of our study:

Bhatia considers legislative provisions as a highly impersonal and decontextualised genre, whose illocutionary force (Austin, 1975; Searle, 1969) holds no matter the speaker or the reader, since the lawmaker assumes this must be understood by every citizen, law student, lawyer, judge or politician. This force is clearly directive, as provisions impose obligations and confer rights. Nonetheless, due to the impossibility to predict what will happen in the future, they guard against eventualities trying to refer to every imaginable contingency. In short, this genre has a double purpose: serve as a resource to reflect lawmaker's intentions, and simultaneously facilitate comprehension to all citizens.

Bearing in mind these communicative purposes, Bhatia (1993) attempts to describe legal provisions syntactically in a way that we can understand the possible reasons behind the selection of some features instead of others. He highlights a prominent use of unusually long nominalisations, such as '*permanent abandonment of such actions*'; a very high average sentence length (271 words), that is, the average number of words in a sentence, in comparison to regular English sentences (27.6 words); complex prepositional phrases with a PNP structure, with a purpose of reducing ambiguity, such as '*by virtue of*', *'for the purposes of*', or *'in accordance with*'; the use of binomial and multinomial expressions, again, due to the lawmaker's attempt to be as precise and all-inclusive as possible (*signed and delivered*, *wholly and exclusively*, *any sum of money or other consideration*), and, lastly, the repeated use of qualifications restricting and precising the conditions in which a law is enforceable (that is, the use of long and numerous conditional subordinate clauses), This last feature led to the presence of unwanted syntactic discontinuities.

A more recent study developed by Vass (2017) drew on the idea of judgments being useful as records of case facts, arguments and legal reasoning, as well as to inform future practitioners of law to explain the relevance of persuasive language in this genre. He explains that in the US Supreme Courts there are nine members which may agree or disagree with the others on the decision that should

be made. That is why there exist majority and dissenting opinions in almost every case, rather than just a unique common decision. He claims that the use of persuasive language, and thus, of hedging resources (Hyland, 2005), such as *may, might* or *possibly* is even more frequent in this type of decisions, as they are used in an attempt to nuance and reduce their commitment to their arguments. Some examples of hedging are the use of the structure *it is + adjective to suggest that…* or the presence of speculation by using structures such as *if… would*:

(1) **It is also quite odd to suggest that** the problem with North Carolina's law would go away if only the State provided some sort of study substantiating the idea that […]

(2) But, **if**, as the Court suggests, there are a multitude of copyright owners champing at the bit to bring lawsuits against libraries, […], might one not expect that at least a handful of lawsuits **would have been filed** over the past 30 years? […] (Vass, 2017, p. 349)

1.1.2. Legal Translation

Translation scholars have had great interest in studies related to genre and discourse studies such as the showed in Section 1.1.1, and they even collaborated in the conduction of some of these works. However, they also needed a practical application and systematic comparison dealing with the typical features found in these genres, as well as a suggestion of soul translation strategies able to transfer the properties of the genre from the source language to the target language.

Alcaraz and Hughes (2015) reviewed and summarised how Legal English and Spanish features resemble and diverge, as well as the problems when translating them. Some of them are the following: Legal English lexical source is mainly Latin, French and Normand, for example in the case of lexical bounds such as '*ratio decidendi*'. In Legal Spanish, apart from the terminology borrowed or translated from Latin ('in dubio pro reo'), we find borrowings from Ancient Greek ('amnistía') and Arabic ('albacea,' 'albarán').

Both in Legal English and Spanish, lexicon can be distinguished depending on its specialisation: (1) Technical terms, with a single meaning (univocity) and usually found in legal language, such as 'interdicto' or 'tort'; (2) Semi-technical terms, terms that acquire a new specialised meaning (equivocity). For example, 'issue', which means 'offspring' instead of 'affaire or problem' in legal language; (3) Colloquialisms, not changing in meaning, but very present in legal discourse (Alcaraz & Hughes, 2002, 2015).

Another feature to highlight is legal discourse's performative character. For instance, a judgment where a defendant is convicted does not merely constitute a facts narration, but the conviction itself: 'performative verbs' are essential for the fulfilment of legal procedure. Actions such as dismiss or uphold are done by the uttering itself, and this is why they are called 'performative'.

However, the feature most easily perceived by the recipient is its crypticness and complex style, stemming from its archaisms, complex sentence structures, and metaphors. This language, the so-called "legalese", greatly humpers the understanding of the text for readers not specialised in the field. This has led to movements such as the 'Plain Legal English Campaign' aiming to simplify this language and make it accessible to non-lawyers. This is reflected in English in register, archaic verbs ('witnessth,' 'whereof,' 'herein') and lexical redundancy ('give, devise and bequeath') or euphemisms ('act of God')

In Spanish, the creation of neologisms, as well as nominalisation, are also common discourse features. Disproportionate subordination, coordination, and juxtaposition presence are remarkable in Legal Spanish, occasionally leading to failure to follow on, an unconscious departure from the grammatical scheme with which a sentence was started, that is, syntactic discontinuity. Such cases force translators to change and simplify sentence structures either from English to Spanish or from Spanish to English (Alcaraz & Hughes, 2002, 2015).

Authors such as Piszcz and Sieroka (2020) analyse the relevance of cultural and societal contexts in the development of legal discourse and its proper translation. Differently from other types of specialised discourse, such as Medical or Technical Discourse, Legal Discourse is strongly influenced by the cultural context of the societies is developed, since law is eminently a cultural product. This consequently means that the writing style, argumentation, legal reasoning or reasoning skills shown by the judges may differ depending on their country and the language used. Gozdz-Roszkowski's study (2020) exemplifies this by explaining how common law and civil law traditions lead to a different legal reasoning development. In the former, judgments show puns, humorous and metaphorical expressions, as well as persuasive devices so as to make the discourse more appealing and easier to understand to the parties involved. On the contrary, judges belonging to the latter tradition, which is rooted in codified laws, tend to use a more impersonal and stylised language.

Piszcz and Sieroka's article (2020) also shows examples of the effect of culture on the challenges legal translators face, such as the translation of the terms *court* and *tribunal* from Korean. On this matter, Wojtasik-Dizekan (2020) analyses the difficulties this job poses, as Korean legal terminology is mainly imported from Chinese, due to their adoption of Chinese writing (*hanzi* ideograms) until the development of their own writing system.

Comparative Law has been another of the interests shown by Translation Studies scholars, with the aim of improving the (dis)similarities of the legal systems in which their working languages are framed. Engberg (2020) proposes a multidimensional approach (Knowledge Communication Approach), which consists of a system of rules and steps that a translator may follow so as to take advantage of their knowledge on the legal systems involved in their translation process systematically. He considers that the legal translator chooses strategically relevant parts of the complex conceptual knowledge they understand from the source text, and they attempt to transfer the relevant aspects of it into the target language, so

the reader receives a similar conceptualization of this knowledge. He divides this process in three steps: (1) the translator delimits which aspects are central, (2) they select those aspects that are also relevant in the target text, (3) they attempt to reformulate the concept so they can transfer those aspects irrelevant in the target legal system but crucial to understand the meaning conveyed in the source text.

For Engberg (2020), the appreciation of the multi-faceted character of law is crucial. Law is not only (1) a set of rules, but also represents an (2) organization (legal system of the country / supranational institution, (3) a symbol (containing symbolic aspects of the national identity), (4) and performance (enforceability of the rules).

The importance of Comparative Law as a useful strategy for the preparation of the translation of legal documents is noticed by other scholars, when attempting to perform translational analyses. Granados-Meroño and Orts (2021) exhaustively examined the terminology regarding corruption and their codification as offences in the Spanish and British legal system, as well as the regulation recommendation of international organizations specialized in corruption offences, such as GRECO and UNCAC. This work allowed them to propose relevant and well-documented translation strategies for corruption-related terms in a court order from Spanish into English.

Genre, Discourse, and Translation Studies scholars produced a considerable amount of research dealing with legal discourse from many different angles, interests and approaches. Nonetheless, many of them had something in common: the reflective, qualitative and great detailed character of their studies. This is not essentially bad. Contrarily, they provided with a strong and wide theoretical basis future researchers interested in the field of legal discourse will surely find enlightening. However, to obtain more reliable, replicable and refutable results, these conclusions must be at some point verify by quantitative studies, which, even if increasing in number, are still lacking.

1.1.3. Corpus Linguistics

The expansion of corpora and computer tools extraordinarily boosted the studies of applied linguistics, which now could draw on tons of data, that is, millions of words from language naturally produced (Goźdź-Roszkowski, 2021). This "has offered insights into the language that have shaken the underlying assumptions behind many well-established theoretical positions in the field" (Bonelli, 2010), that is, the ones obtained from the many studies reviewed in previous sections. This is mainly due to the fact that qualitative and reflective studies having conducted by genre, discourse and translation scholars until the implementation of corpora as a regular source of data for linguistics research was based on intuitions, deductions and categories produced by the knowledge these authors obtained from their expertise of the field, but not from actual representative data. This is extremely useful to start designing a hypothesis or to interpret results from quantitative data, but their conclusions are very likely to be refuted if not based on actual data (R. Z. Xiao, 2008).

Thus, the adoption of the CL approach, which essentially means (1) the use of corpus as the main source of data and (2) the use of statistical analysis, marked a turning point for the study of legal discourse. Some aspects of legal discourse were especially responsible to CL, such as its formulaic language, when faced from a comparative perspective. This provided new insights into legal discourse, such as the presence of much more fixed phrases than what used to be thought (Goźdź-Roszkowski, 2021). Nonetheless, this does not mean that the only discipline taking advantage of computerised tools was CL, since other areas such as Argumentation Theory have extensively used them developing argument mining (Section 1.2).

As mentioned above, the main feature of CL is the use of corpora as its main source of information. The definition of what a corpus exactly is has been controversial and discussed for long, but all the definitions usually agree on them being "*a collection of texts according to certain criteria"* that are aimed at fulfilling the representativeness of the corpus of the particular variation / register /

specialised discourse that is the object of study. These criteria might regard sampling, finite size, representativeness or the machine-readability of the texts obtained (Goźdź-Roszkowski, 2021, p. 4; Llisterri & Torruella Casañas, 1999; Marín & Rea Rizzo, 2012). Moreover, one of the aspects that makes the use of corpora for research even more useful and interesting is the fact that they can be added information by means of corpus annotation, that is, the addition of information about metadata, such as author, year of publication or presumed receptors, and about the syntactic (parsing), morphological (POS tagging) structure and semantic content (Goźdź-Roszkowski, 2021). This task has been in many cases automatised thanks to the use of NLP tools such as POS taggers or parsers, more and more extended and accurate with the development of machine learning and deep learning techniques (Benoit et al., 2021; Benoit & Matsuo, 2020; UCREL, 1987).

The most two relevant approaches of CL to use the information obtained from corpora are two: (1) corpus-based approach, used in studies aimed at verifying or refuting results from previous studies and (2) corpus-driven approach, whose studies use the corpus as the sole source of information for their conclusions, not considering other qualitative or reflective studies. The latter is more common in exploratory studies that are concerned with languages or language variations that have been scarcely studied (Goźdź-Roszkowski, 2021).

As far as legal discourse is concerned, several corpora have been recently compiled, mainly containing public law genres, such as legislation, judicial decisions, law reports, but also some private law genres such as wills or contracts. The latter are much more limited, due to the obvious privacy concerns that surround them. Moreover, some oral legal corpora have also been compiled, containing genres such as witness examinations (Goźdź-Roszkowski, 2021). The most common type of legal corpora are possibly the one containing legislation and judicial decisions, and the ones from European Union (EU) and United Nations (UN) bodies are prominent, due to their accessibility and availability of multilingual versions.

For instance, the UN Parallel Corpus (Ziemski et al., 2016), which is a compilation of manually translated UN documents from 1990 to 2014 for the six official UN languages, Arabic, Chinese, English, French, Russian, and Spanish; the Digital Corpus of the European Parliament (DCEP) (Hajlaoui et al., 2014), containing around 1.37 billion words from several types of documents in the 23 official EU languages between 2001 and 2012; or tools such as EUR-Lex,[1] where every legislation promulgated by the European Parliament in all the EU official languages. These corpora and other similar can be found in the webpage developed by the CLARIN Project.[2]

Many other corpora contain national judicial decisions, mainly from superior courts such as the national Supreme Courts: the HOLJ corpus contains 188 judgments of the House of Lords from 2001 to 2003 (Grover et al., 2004), while the British Law Report Corpus (BLaRC) contains around 6 million words from judgments by different Higher Courts (Scotland, England and Wales, Northern Ireland and other Commonwealth countries) between 2008 and 2010 (Marín & Rea Rizzo, 2012). In addition, it is worth to mention the multilingual English-Italian Bononia Legal Corpus[3] (Rossini Favretti et al., 2007), one of the most comprehensive legal corpus existing because of its selection of varied genres. It contains documents of legislative, judicial and administrative nature, in an attempt to be a representation of English and Italian legal systems. One of the most recent corpus compiled regarding legal discourse is the COCELD (Corpus of Contemporary English Legal Decisions) (Rodríguez-Puente & Hernández-Coalla, 2023). This diachronic corpus contains legal decisions from 1950 to 2021, making it suitable for analysis interested in exploring the changes legal English might have undergone. It is divided in one subcorpus for the Privy Council decisions and another for the House of Lords and Supreme Court decisions. The main purpose of this compilation was

[1] EUR-Lex, online access to EU law: https://eur-lex.europa.eu/homepage.html.
[2] CLARIN The research infrastructure for language as social and cultural data: https://www.clarin.eu/.
[3] Bononia Legal Corpus: https://corpora.ficlit.unibo.it/bolc_eng.html.

the exploration of the effects that the Plain Language Movement in the 1970s might have had on the development of legal English.

One of the main interests that the authors conducting studies on legal discourse using CL have had during the recent years has been, on the one hand, the exploration of legal phraseology, on a comparative basis. Phraseology is one of the most relevant aspects of legal discourse, due to their use of formulaic language and fixed expressions in the different types of legal documents. The so-called doublets and triples mentioned in previous sections had been studied by discourse and translation scholars on a qualitative approaches, and researchers such as Ruth Breeze aimed to confirm the conclusions made by them. On her study, Breeze (2013) analysed the differences in terms of lexical bundles among four legal genres: academic legal texts, case law, legislation and legal documents (such as contracts). She used a 2 million-word corpus divided into four 500 thousand-word subcorpora (one per genre) for this purpose. She found some interesting patterns, such as the presence of more lexical bundles in legislation and documents than case law and academic legal texts, which, in turn, used a more complex, unconventional language (thus, with a higher type-token ratio).

Other studies explored these variations in lexical bundles regarding translated. For example, Biel (2017) used an European legal corpus and its translated counterpart (the Polish and English Eurolect Corpora) and a Polish Domestic Law Corpus to do this. He was unable to confirm the hypothesis stating translated texts show fewer lexical bundles than the source texts. Instead, he found similar numbers between both, and a very low coincidence between the lexical bundles used in the Polish Eurolect Corpus and the Polish Domestic Law Corpus.

In turn, Giampieri (2024) recently conducted a study comparing N-grams between two English corpora: one containing EU legislation and another containing UK National acts. She confirmed the existence of the "Eurolect", that is,

terminology and jargon that ensue from EU law, not having a correspondence or use in the national legal texts of English speaking countries. For example, there is a preference for the word *withdrawal* with the meaning of *cancel* in EU legislative texts, or a more frequent use of *shall* rather than *must*.

Language variation from the approach of register and genre analysis has been, on the other hand, of great interest for legal discourse authors. Now that they had the data and methods allowing them to do so, they started to explore the features salient in legal discourse through the exploration of its genres: case law, legislation, contracts, wills, witness statements, and so on. They approached this exploration internally, that is, comparing legal genres between themselves, and externally, that is, comparing legal genres with other specialised and general genres. Moreover, this *cross-genre* approach was also added *cross-language* and *diachronic* comparison, improving the insight into the differences not only between genres, but also between languages and different periods of time (Goźdź-Roszkowski, 2021). To achieve this, legal discourse authors needed a powerful, multi-layered methodology that was able to compare genres not only in one aspect of language, such as terminology, or phraseology, but holistically, that is, bearing in mind many variables simultaneously. Fortunately, Douglas Biber took advantage of the benefits of factor analysis to develop in 1988 the so-called *Multi-Dimensional Analysis.*

1.1.4. Multi-Dimensional Analysis (MDA)

Multidimensional analysis (MDA) was designed by Biber (1988) to identify the underlying linguistic dimensions of variation in language from a quantitative approach and to compare spoken and written language registers in the linguistic space defined by those dimensions. Before designing its methodological process, he followed and structured some theoretical concepts on speech theory, namely *speech situation* and *linguistic function* as aspects marked by linguistics features of a genre.

Biber (1988, pp. 29–33) distinguishes eight components of a speech situation:

i. *Participants roles and characteristics*: the communicative roles of participants and the individual characteristics of each participant
ii. *Relations among participants*: participants' social role, their personal relationship, shared cultural knowledge.
iii. *Setting:* aspects of physical and temporal context (when and where the utterance is produced, what activity it is framed in, etc.)
iv. *Topic*: what the message is about
v. *Purpose*: outcomes the participants hope for, aims they have, the objective they want to achieve with the utterance.
vi. *Social evaluati*on: participants' attitudes (and of the culture at large) to a communicative event
vii. *Relations of participants to the text*: ability to interact with the text
viii. *Channel*: the medium used for the message utterance.

Regarding the speech situation, Biber (1988, pp. 33–36) also delimits the main linguistics functions that can be found in any discourse:

I. *Ideational function*: conveyance of propositional or referential content
II. *Textual* function: information about the structure and prominence (topic-commentary, or coherence) and cohesion (such as ellipsis, substitution, repetition, demonstratives, deixis)
III. *Personal* function: group of membership or a personal style
IV. *Interpersonal function*: attitudes towards the communicative event or the message, depending on the relationships between the participants and their shared knowledge
V. *Contextual function*: setting, purposes and perception of the event
VI. *Processing function*: production and comprehension demands of the communicative event
VII. *Aesthetic function*: personal and cultural attitudes about the forms of language.

As Biber has claimed in many occasions (1988, 1995; 2019), the sole quantitative MD analysis is far from being sufficient to obtain a proper view of how

different genres or registers are; a proper theoretical background that helps the researcher to interpret the factors indicated by the factor analysis to the speech situation and the functions of the genres analysed is essential as well. That is why he provided a robust theoretical framework based on genre and translational analyses (Alcaraz & Hughes, 2015; Bhatia, 1993; Cao, 2016; Swales, 1990).

The steps to undertake MDA will be later described and detailed on the materials and methods section. Still, an overview of the steps that must be followed are provided below (Biber & Conrad, 2019, p. 225):

1. Design and compilation of the appropriate corpus drawing on previous research and analysis. Documentation of the situational characteristics of registers involved.
2. Conduction of research for the identification of the set of linguistic features to be included in the analysis.
3. Development or selection of computer programs for automated grammatical analysis; analysis of the entire corpus of texts to compute the frequency counts of each linguistic feature in each text.
4. Analysis of the co-occurrence patterns among linguistic features by means of factor analysis of the frequency counts.
5. Computation of the factor scores for each text; the mean factor scores for each register are then compared to analyse the linguistic similarities and differences among registers.
6. Factors interpretations as underlying dimensions of variation.

The application of Biber's framework to our study is further explained in Chapter 3. MDA's major upside is the use of factor analysis so that a large number of variables are reduced to small underlying variables, the so-called *textual dimensions*, showing the relationship between the variables belonging to each dimension.

Since its first implementation in 1988, the number of studies applying this methodology in cross-genre, cross-linguistic or diachronic analysis is countless.

They have been developed for the study of academic discourse (Biber, 2006), a recreation of Biber's 1988 analysis for Spanish language (Parodi, 2003), the search of universals in literate languages (Biber, 1995), word Englishes (R. Xiao, 2009) or, more recently, English in Twitter (Clarke, 2022), Southern Asian digital Englishes (Shakir, 2024), English textbooks (Le Foll, 2024), and even comparing ChatGPT and Human produced texts. It is such the relevance of MDA for the development of CL, that a monograph on this topic was published celebrating the 25th anniversary of Biber's 1988 publication (Berber Sardinha & Veirano Pinto, 2014).

As far as the study of legal discourse is concerned, there have been some studies that have implemented MDA for the study of language variation. *Cross-genre* variation has been covered by Goźdź-Roszkowski (2011). For this study, he used the *American Law Corpus (ALC)*, containing around 5,5 million words and 7 different legal genres (academic journals, briefs, contracts, legislation, opinions, professional articles and textbooks. After the corpus processing and annotation, he performed MDA externally, that is, comparing these legal genres with other non-legal genres, (broadcasts, scientific papers, etc.) (Matulewska, 2014). More recently, Huang and Sang (2024) developed a MDA comparing oral discourse produced by legal professionals using the *CABank English SCOTUS Oral Arguments Corpus*. The most relevant differences were found in Dimension 1 – Instructive Argumentation vs. Information production: justices' discourse was strongly skewed towards the *instructive argumentation* side of the dimension, while factor score loadings of discourse produced by prosecutors and defence attorney were leaning towards the *information production* side (Figure 1).

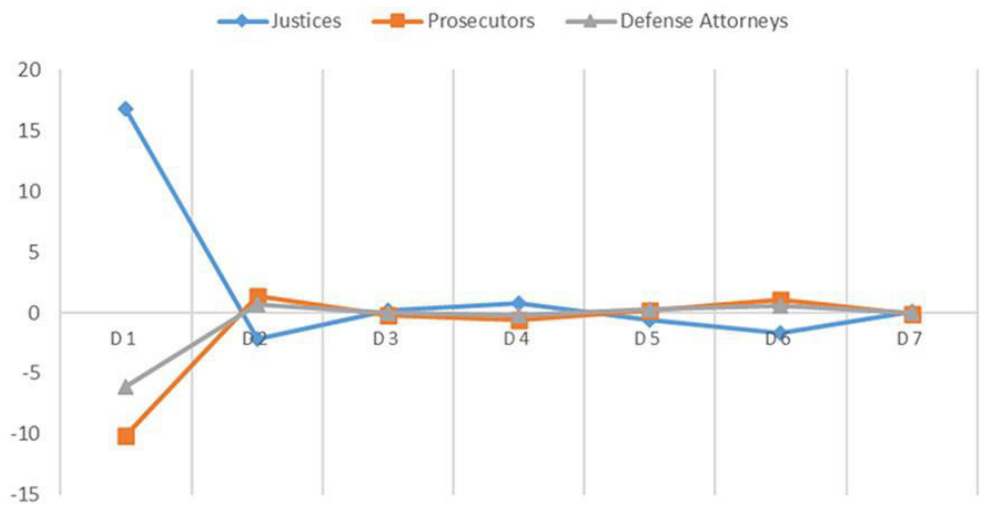


Figure 1: Mean dimension scores of legal professionals' corpora. (Huang & Sang, 2024, p. 8)

Cross-linguistically, Sun and Cheng (2017) explored the linguistic variation of Chinese legislation, using corpora that contained 1202 texts (6 million words) for Chinese legislation, the same number of texts for the translated version (with around 4 million words) and the ALC for the American legislation (54 texts with a thousand words). Their findings show legislative texts share certain features both in Chinese and American legal discourse, namely a non-narrative, explicit, highly informational and decontextualised style. The authors assumed this is due to the constant use of conditionals in legislation. In turn, Chinese legislation appears to be much more abstract and informational than American, whereas American shows a more persuasive discourse than Chinese. Regarding Chinese to English translated texts, they found a higher presence of discourse markers than in source texts, as a compensation to the lack of them in Chinese that would lead to a poorly structured discourse in English.

In turn, Granados-Meroño (2023) designed a comparative MDA between judgments of the Supreme Courts of the UK and Spain, in an attempt to verify the conclusions made on the English-Spanish judgments differences by translation scholars in previous qualitative studies. The conclusions were limited, due to the size of the corpus used (a 44-thousand-word corpus containing 10 judgments for the English corpus, and a 73-thousand-word corpus containing 10 judgments for

the Spanish one), but he found interesting patterns such as the presence of a different dimension in each language: one dimension was interpreted as *Persuasion vs. Power Distance* in English, while in Spanish as *Intertextuality.* This was understood as a consequence of the difference in legal systems between these countries, since Spain belongs to the civil law family, with a prominence of codified law, while Anglo-Saxon countries belong to the common law system, with a preference for the *stare decisis* system of legal precedents forcing the judges to draw their decisions on legal reasonings and previous judgments (Alcaraz et al., 2014).

1.2. Argumentation Technology and Legal Argumentation

Linguistics studies are not alone in the exploration of the language produced by legal professionals. They do not refer to it as *legal* discourse, though, rather *legal* language, *legal* English or *legal Argumentation*. Philosophy and computer scientists have had extensive interest in the understanding of Argumentation. Philosophers have struggled to develop a common theoretical framework that defines what argumentation exactly means, and which the parts these are compounded of are (Eemeren et al., 2014; Perelman & Olbrechts-Tyteca, 1969; Toulmin, 1958; Walton, 2006).

In turn, computer scientists gained interest in argumentation to develop guidelines, tools and models for the (automated) annotation, detection and (ultimately) generation of arguments and fallacies (Lawrence & Reed, 2015), creating fields of study: *Argument Technology and Argument Mining.* These scholars, nonetheless, seem to have not fully taken advantage of the advanced in NLP and CL for the refinement of these technologies, possibility due to the lack of interdisciplinary communication, which is one of the gaps this dissertation tries to fill.

This does not mean there has been no interaction between linguistics and argumentation theory, but that this has been limited. The more relevant is the

development of argumentation theories from the perspective of pragmatics approaches, the so-called *pragma-dialectics* (Eemeren et al., 2007, 2014; Feteris, 2012, 2017), which will be covered at the last part of the following section.

1.2.1. Argumentation Theory and Legal Argumentation Schemes

The notion of argument is described from a wide range of insights, but the one used by Walton (2006, p. 1) might be one of the most comprehensible and straightforward provided in literature, but still accurate and specific enough for using it before a specialised audience: "*the giving of reasons to support or criticise a claim that is questionable, or open to doubt*". This simple definition of the concept can be divided into two fundamental parts of what constitutes an argument, following van Eemeren's pragma-dialectical approach (Eemeren et al., 2014).

On the one hand, the nucleus of any argument is a *standpoint at issue*. In other words, any claim or idea, which might be descriptive, prescriptive, or evaluative, trying to be defended or criticised by the arguer (utterer, writer) and that is arguable by the recipient. These standpoints in issue, also called *conclusions*, (C) will be surrounded, explicitly or not, by some *premises* (P) that lead to the conclusion the arguer is trying to defend. On the other hand, these *premises* (P) might be logical relations, common knowledge between the utterer and the recipient, opinions, statements, or facts that will be used by the arguer to defend the *conclusion* (C) (Figure 2).

In a nutshell, Walton's definition, combined with Eemeren's elementary explanations of what standpoints are, would provide us with a more accurate definition of arguments. According to their definition an argument essential consists of *the giving of reasons (P, premises) to support or criticise a claim that is questionable (C, conclusion).*

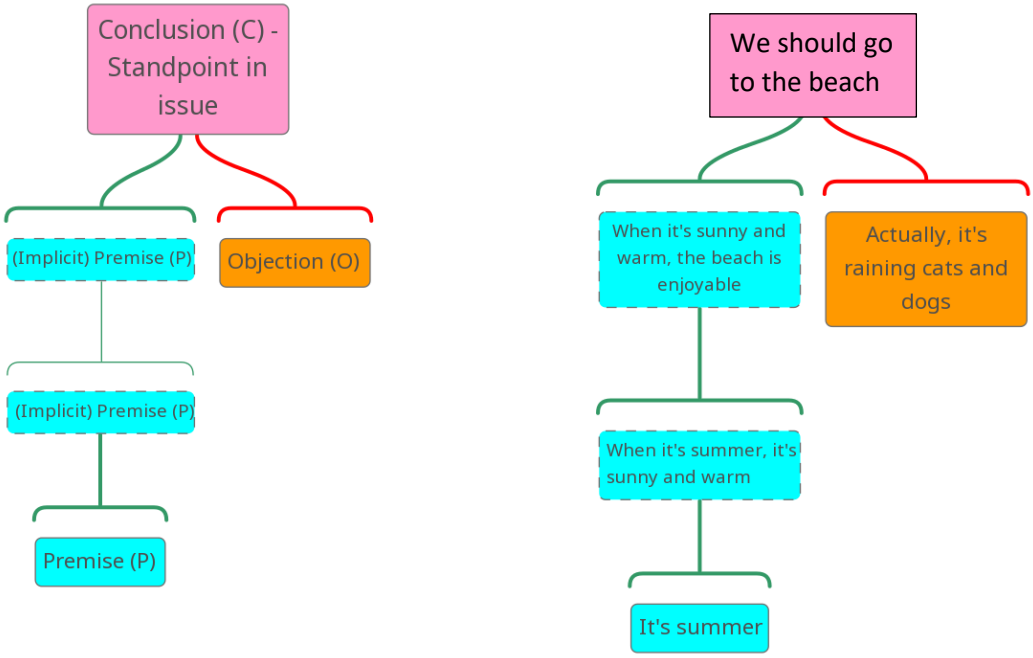


Figure 2: Basic Argument structure diagram (left) and argument example (right)

Argument schemes are types of inference (different types of forms that present arguments in discourse) that represent structures of common types of arguments in either common day language or specialised contexts such as legal argumentation, being named up to 60 different types of argument schemes (Walton et al., 2008). This set was developed to help authors identify and classify arguments, as well as better understand how argumentation works in the real world. These schemes have been broadly used in recent years to develop both manual and automatic argument annotation guidelines. There are other approaches to the detection and classification of arguments, such as Wagemann's Periodic Table of Arguments (Hinton & Wagemans, 2022), but due to the existent Walton's (2010) selection and explanation of prominent argument schemes used by legal professionals, his model is more adapted to legal contexts.

Nine are the argument schemes considered by Walton (2010) the most prominent in legal contexts, extracted from his broader set of schemes: Arguments (1) from Analogy, (2) from an Established Rule, (3) from Sign and Abductive Argument, (4) from Position to Know, (5) from Verbal Classification, (6) from

Commitment, (7) Practical Reasoning, (8) Ad Hominem and (9) Slippery Slope Argument, not excluding that other kind of less common argument schemes might be encountered as well.

In order to make a more comprehensible and useful list of argument schemes for our study, the classification of schemes Walton and Macagno later developed (2015) is also considered in this dissertation, in an attempt to create a more hierarchical set of schemes, clustering them into three big groups: source-independent argument schemes, source-dependent schemes (both of them being defeasible schemes) and a third one including practical reasoning argumentation schemes. Having a quick look at the schemes included in each of the big categories in this new classification, most schemes found in legal argumentation are defeasible ones. That is not surprising at all, as legal argumentation will be in most cases based on precedents, the law, the statement of a witness, evidence and other type of empirical sources (Walton, 2010).

1.2.2. Argument mining on legal texts

As well-accepted frameworks, despite their problems and improvements still to be made, models such as Walton's set of argument schemes became more prominent and used in the field to identify and distinguish the types of arguments found in texts. Simultaneously, the interest in creating systematic guidelines, assistant tools, and automatic detection algorithms to perform that task increased during the last years.

Mochales and Moens (2008) were interested in better understanding the argumentative structure underlying legal texts so as to develop an automatic detection tool of legal arguments. With that purpose in mind, they analysed a corpus based on judgments and decisions from the European Court of Human Rights (ECHR). They found a common macrostructure: Introduction, the Facts, and Proceedings before the Commission, complaints and the Law. They also noticed the

argumentative structure of the final decision and many rhetorical markers such as *however*, *although* or *in particular*. (Mochales & Moens, 2008).

They used this corpus and the *Araucaria Corpus* in many attempts to develop a more accurate tool able to detect and classify arguments in legal texts, by using maximum entropy models and a naïve Bayes classifier in combination with rules based on a generative grammar. This classifier considered features such as unigrams, bigrams or trigrams, keywords, adverbs, or sentence length to perform that task, obtaining an accuracy around 60-70% (Mochales & Moens, 2009).

Lawrence and Reed (2015) revisited the three of the most relevant argument approaches. The first approach is using discourse indicators as a tool for finding argumentative connections between adjacent propositions in a piece of text. These indicators are "explicitly stated linguistic expressions of the relationship between the statements" (Webber, 2011), which might clearly indicate its argumentative structure. The results of their study showed that, when present in text, these discourse indicators certainly manifest the connection between propositions (precision of 0.89), being their low frequency of presence in the text a downside when considering them as a reliable tool to find most connections (recall of 0.04).

1.2.3. Legal Discourse and Legal Argumentation

As seen in the previous sections, the theoretical study of argumentation, as well as the development of computational tools dealing with them, have been focused on the logical structure behind them, but the linguistic features shaping them are far from being completely explored. Some authors such as Feteris and Eemeren have attempted to establish that connection between the logical and linguistic sides of argumentation, but their focus has been on the pragmatic aspects of them, due to the dialectical character any argumentative text shows:

The judge is the utterer of any type of judgment. To completely understand the purpose of the genre of judgments, it is fundamental to understand who and what is the role of the individuals producing these genres. From the approach of

pragma-dialectics, Feteris analysed what is the role of the judges (2012) and how they performed their purposes to fulfil that role by using certain argumentative structures in legal decisions (2017). She explains that the institutional goal of any legal proceeding is that the discussion process regarding any legal claim is organised in a way that finishes with an impartial decision in accordance with the Rule of Law. The role of the judge is to guarantee that the right process is followed and to come up with a decision that was based on the correct application of the law.

This 'ideal' discussion proposed by the institution to solve conflicts was implemented with a series of stages that facilitate the correct succession of events that lead to the result intended: a *confrontation stage*, establishing the scope and content of the dispute; an *opening stage*, establishing common legal starting points in codes of law and common factual starting points; *argumentative stage*, establishing the acceptability of the argumentation in defence of different legal claims based on common testing methods; and *a concluding stage*, establishing the result of the discussion. In each of the stages, the judge will play a crucial role to ensure the process to be properly followed (Feteris, 2012). This analysis of the role of judges is rather complementary and insightful when compared to the one made by Bhatia (1993) mentioned in Section 1.1.1.

The most important work of the judge is the establishment of a final impartial decision finishing the discussion, but always in accordance with the Rule of Law. To do that, the judge surrounds their final decisions with a record of the facts recognised by the parties, the succession of arguments proposed by each of the parties and, most importantly, a bunch of legal precedents (case law) and articles of the law in which they ground their decision, justified by a very defined argumentative structure. For instance, in clear cases, the justification of the decision implies that the court must specify the factual and legal grounds of the decision, with an argumentative pattern consisting of the following parts (Feteris, 2017):

(1) A standpoint specifying the decision that legal consequence Y must or must not follow.
(2) An argument specifying the legal qualification of the facts of the case in terms of the conditions for applying the legal rule R
(3) An argument specifying the applicable legal rule R

Pragma-dialectics is useful as it adds the layer of pragmatics and a communicative-linguistic insight to the study of the logical structure of argumentation, analysing their actors, purposes and social contexts. Nevertheless, it does not explore the linguistic features that might be involved in the 'discourse indicators' explored by the authors mention above (Mochales & Moens, 2008, 2011). They do have been further explored, studied, used in the area of linguistics with the purpose of better understanding genre and register variation, as well as understanding and teaching more efficiently how to write proper argumentative texts.

For example, Biber developed MDA (see Section 1.1.4) applied to university language, in which he interpreted a factor as the 'Oral vs. literate discourse' dimension. Some features belonging to this dimension such as the use of pronouns, contractions or the semantic content of the verbs used represented the oral discourse side of the dimension, opposed to features such as the use of nominalisations, the use of abstract nouns, passives or prepositional phrases, which represented the literate discourse side of the dimensions, as the two sides are negatively correlated between each other (in other words, the two clusters repel each other). Some of these linguistic features have been associated to a persuasive and argumentative dimension of the discourse, such as the use of the first and the second person rather than the third one, the use of active voice rather than passive, or the use of conditionals or possibility modal verbs so as to accept counterarguments to a standpoint in issue (Biber, 1988; Biber & Conrad, 2019).

The potential of these linguistic features has not been explored yet in the field of argument mining, since only a few of them have been considered as

discourse indicators to enhance the detection of arguments and the different argumentation schemes that one might encounter in a text.

The same happens with Hyland's metadiscourse (2005), closely related to the pragma-dialectical discourse indicators explored in the field of argumentation theory. He described metadiscourse as a cover term "for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community" (Hyland, 2005, p. 37).

Metadiscourse shows three key principles that are common to all the expressions under this term, and that particularly support the idea that argumentation not only is framed by logical structures but also linguistic patterns:

Principle 1: *Metadiscourse does not only refer to propositional content*, or 'communicative content, that is, what makes a text coherent, intelligible, and persuasive, but also to material that conveys writer's beliefs and attitudes towards it. Propositional and metadiscoursal elements co-occur in texts, that is, therefore might be conveying at the same time a reference to a cause expressed in a text, but also to beliefs, intentions, argumentations the writer is trying to convey by creating a cause-effect link in a sentence.

Principle 2: *Metadiscourse expresses the interaction between the utterer and the recipient of the discourse*. It takes account of the recipient's knowledge, their textual experiences and their processing needs. In other words, Hyland (2005) considers that the 'textual devices' the utterer may use, such as conjuncts (so, because, and) or adverbials (first, therefore), are always interpersonal, but the way they perform this task differs mainly in two ways:

On the one hand, *interactive metadiscourse* accounts for ways the utterer signals the arrangement of their discourse, this includes strategies such as rephrasing, or devices such as deictics or conjunctions. Even if, apparently, these could be interpreted as an ideational purpose of conveying the structure of an

organised and well-uttered discourse, they are doing certainly this task but as a means of helping the recipient better understand, engage in and easily follow the utterance. In short, interactive metadiscourse discreetly performs an interpersonal purpose in the discourse.

On the other hand, *interactional metadiscourse* overtly expresses the intention of the utterer to engage the recipient to the utterance/text/discourse. They persuade the recipient, so they share the ideas expressed in the text by using devices such as hedges, reducing the certainty of their statements, or attitude markers, expressing the utterer's feelings towards what it is being stated.

Principle 3: *Metadiscourse distinguishes between external and internal reference*: internal reference accounts for references made to other parts of the discourse, while external reference refers to references to the external world.

2. RESEARCH PURPOSE AND JUSTIFICATION

This dissertation is aimed at developing new approaches, methods and tools that combine the abovementioned fields, so an easier, faster and more comprehensive research can be conducted by future authors interested in the multi-faceted field of Legal English. To accomplish these broad purposes, a series of more limited research questions / objectives were to be applied, namely:

(1) An exploration of patterns found in linguistic features of British Public Law genres, as a complement to the previous MD analysis performed on American legal genres (Goźdź-Roszkowski, 2011) and a quantitative contraposition to qualitative approaches on British judgments (Álvarez Álvarez, 2008).

(2) A general insight into the factor scores of British judgments and legislation for the original Biber's MD analysis textual dimensions in comparison to other genres.

(3) Potential applications of Biber's approach to register analysis to other areas relevant for Legal English (in our case, Legal Argumentation).

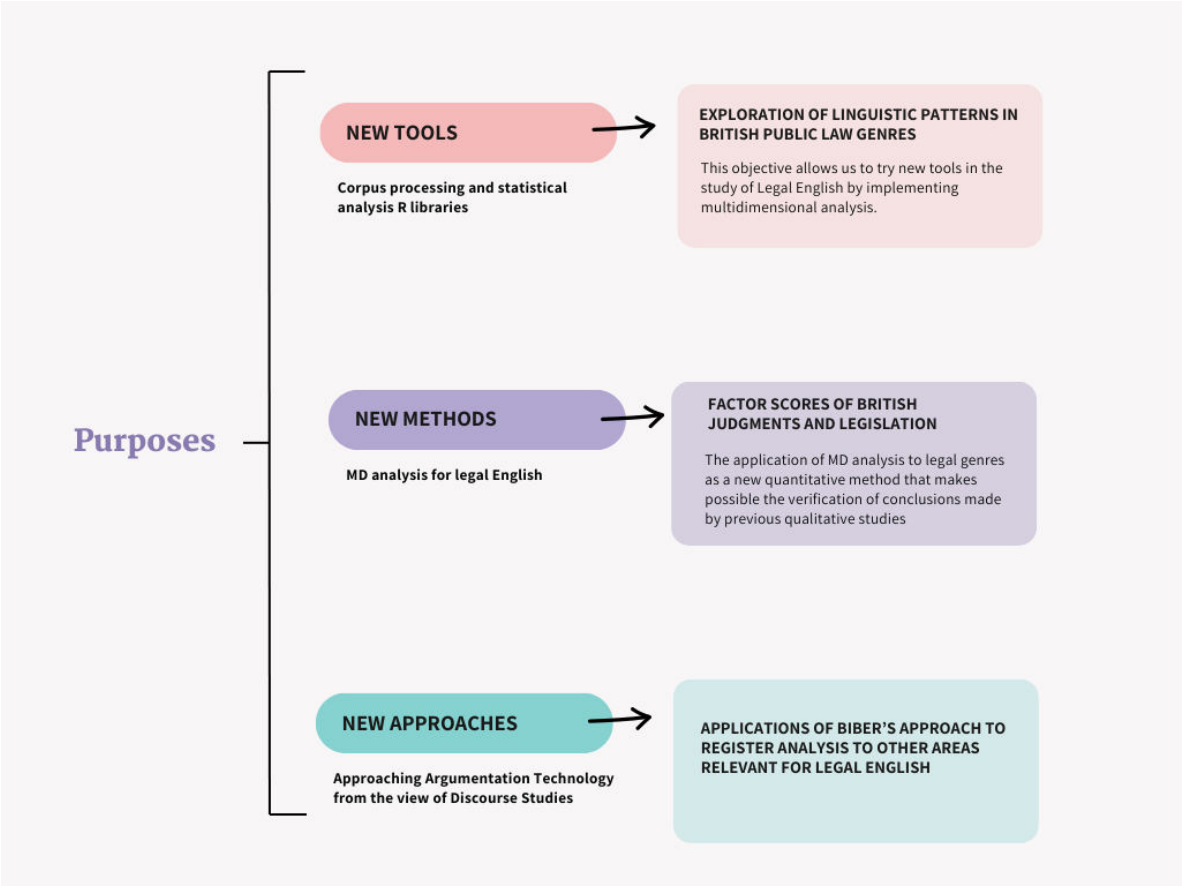


Figure 3: Research Purposes outline

The BLRC accomplished to become a representation of British case law during the 2010s. Thus, a representative corpus of British legislation would be needed to achieve a complete image of public law genres. For this dissertation, the BSLC (British Statute Law Corpus) was compiled, containing more than 10 million words written in statute acts from all the British Parliaments. The corpus compilation process made use of *readtext* and *pdftools* R libraries (Benoit et al., 2021; Ooms, 2023) to dramatically enhance and speed up the text preprocessing process.

Secondly, using the data from the BLRC and the BSLC, a multidimensional analysis was conducted to understand how the legal discourse's linguistic features are clustered and correlated so they are interpreted as textual dimensions. This analysis gives us a complete insight on how legal discourse behaves and whether conclusions drawn by previous qualitative studies agree with it. Results will also

allow researchers to perform comparative analyses with other MD analyses performed so far regarding legal texts.

Thirdly, a second MD analysis is performed, but this time using Biber's original MD analysis dimensions, so a comparison between these legal genres and genres from different fields and specialisations (broadcasts, literature, oral conversations, etc.) under a common framework. This was conducted by using the MAT tool (Nini, 2019).

Finally, as a result of a three-month stay in the Centre for Argumentation Technology (University of Dundee, Scotland), the author conducted a study with the purpose of finding patterns in the linguistic features appearing in different argumentation schemes, by performing a sequence of Pearson's correlations between these linguistic features and argumentation schemes.

To do so, a series of guidelines for the manual annotation of legal argumentation schemes was previously designed so as to obtain a sample of legal argumentation schemes annotated from a legal corpus. Once the guidelines were designed and applied for the manual annotation of a legal corpus of 4 judgments, the linguistic features were automatically annotated using NLTK. Finally, Pearson's correlations were conducted between the argumentation schemes and the linguistic features.

This new approach to the understanding of argumentation, argumentation schemes and their identification arises as a first attempt to blend the MD analysis framework with argumentation schemes research.

3. THE BRITISH STATUTE LAW CORPUS (BSLC): STRUCTURE, DESIGN AND COMPILATION

3.1. Relevance of the BSLC

In Section 1.1.3 a selection of the most relevant legal corpora recently compiled is provided. That selection reveals that the legal corpora compiled so far include national case law, such as the BLRC, and legislation belonging to European and international institutions, such as the DCEP, EUR-Lex or the ones in the CLARIN Project. There are only a few exceptions, namely the subcorpus of American legislation included in the American Law Corpus, used by Goźdź-Roszkowski for his MD analysis (2011). Thus, the attention to the provisions coming from the European Union and other international institutions, affecting the national British legal system, may have made researchers forget about legislation promulgated directly from national parliaments, as the focus has been always on case law, which is the primary source of law in common law systems.

A compilation of a corpus containing a comprehensive collection of acts promulgated by British national parliaments, namely the House of Commons, and other parliaments such as the Scottish parliament, is therefore needed. This corpus would provide researchers with reliable data on the discourse produced by the British legislator in their provisions, in contrast to the legal texts produced by European and international institutions. Studies comparing British and European / international legal English could be conducted by making use of it. The author of this dissertation, in turn, compiles this corpus with the aim of comparing its discourse with the discourse contained in British case law, so as to have a comprehensive and updated insight on British public law texts (case law and legislation). Finally, this corpus is also needed regarding chronological issues, as the majority of legal corpora published included texts of periods ending around 2010 and 2015.

3.2. Corpus design

The British Statute Law Corpus (BSLC) aims to cover the gap the BLRC left. This comprehensive corpus contained case law written by a wide range of British courts, from 2000 to 2010. Marín and Rea Rizzo (2012) used several criteria to obtain a corpus that fulfilled their needs and was useful for the area:

In terms of geographic criteria, as this corpus aimed at representing 'British case law', the delimitation of 'Britain' was crucial. As the British judicial system is not homogeneous and shows differences depending on whether we are located in England and Wales or other parts of the UK, they decided to divide the corpus in five branches depending on the jurisdictions of the judicial systems, namely the Commonwealth countries, the UK, England and Wales, Northern Ireland and Scotland. For the Commonwealth countries, judgments by the Judicial Committee of the Privy Council were included, since it is the highest court of appeal for many current and former Commonwealth countries; regarding the UK as a whole, the House of Lords (later, the Supreme Court), and the net of administrative courts; finally, the different courts pertaining to England and Wales as a single section, and Northern Ireland and Scotland independently.

As far as chronological criteria are concerned, they used judgments delivered in the 10 years prior to the date of compilation of the corpus, that is, 2000 to 2010, following Pearson's guidelines (1998). When dealing with the distribution of the representation of each year, that is, how many judgments per year were to be included, they did not distribute them evenly, finding great variation depending on the court or tribunal the texts were obtained from. The reasons explaining these are several, such as tribunals starting their operations in different years or disappearing in one of the years of the time scope selected.

For the design of the BSLC, the same criteria used in the BLRC were followed, since the type of corpus aimed is very similar to the one compiled by Marín and Rea Rizzo. Therefore, the time scope selected for this corpus was 2010 to 2020, ten

years prior to the compilation of it. Regarding geographic criteria, we covered all the UK national parliaments (that is, Northern Irish Parliament, the Scottish Parliament and the Welsh Parliament or Senedd) and the House of Commons, which promulgates acts enforceable in the whole territory of the UK. England does not have any parliament different from the one in Westminster for its own territory. The distribution objectives were aimed at 10 acts per year and parliament, but that was not possible for every year and parliament, due to the lack of such a number of acts available in the official repository.

Similarly to the BRLC, the BSLC is ultimately designed as a monolingual specialised corpus aiming to represent the whole legislative production in the UK between 2010 and 2020.

3.3. Compilation process

For the corpus compilation process (see Figure 4), the R libraries 'readtext', 'pdftools' and 'reticulate' (Benoit et al., 2021; Kalinowski et al., 2023; Ooms, 2023) were used. These allowed for the text processing workflow into readable txt files to be automated. The legislative texts used for the corpus were obtained from the official repository provided by the UK government[4], where these were classified by body of promulgation, year and topics. Texts were downloaded and stored in PDF format and later converted into txt files by using the libraries abovementioned (R code available in Code 1). This automated workflow allowed the researcher to obtain a 13-million word corpus in a matter of weeks.

[4] UK Government Official Legislation Repository: https://www.legislation.gov.uk/

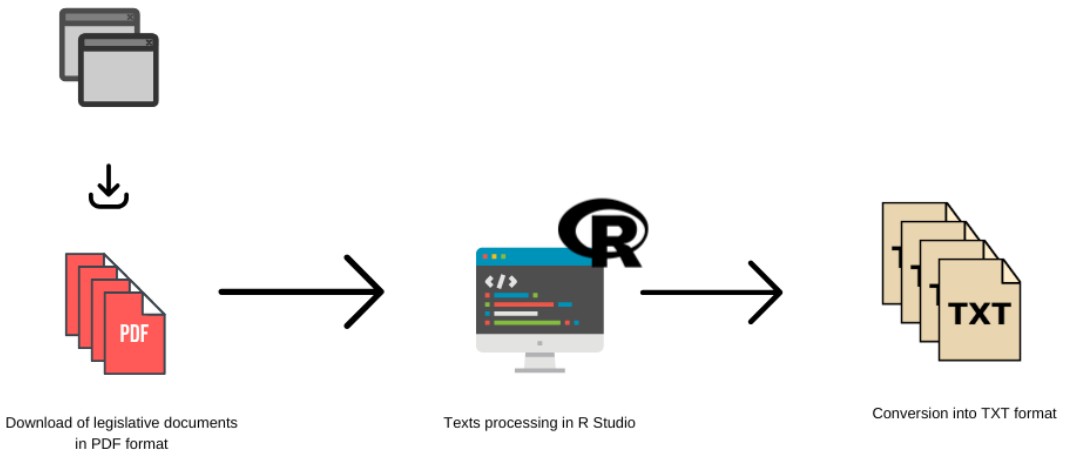


Figure 4: Compilation process workflow

Although not necessary for the purposes of the dissertation, the corpus is prepared for the removing of stopwords and punctuation, as well as basic text data mining (n-grams, wordclouds, etc.) by using the 'quanteda' package. A proposal of code for this matter is provided in Code 2.

3.4. Corpus structure and results

As a result of the compilation process, we obtained a corpus with a total of 714 legislative texts and 13 million words. Nonetheless, due to the great presence of conventions, fixed structures and common vocabulary in this genre, the number of types is 782 580. The distribution of documents and words is mostly even among the documents from national parliaments (that is, Scottish, Northern Irish and Welsh Parliament), but their number is significantly lower than the documents from the House of Commons (see Table 1 and Table 2). This is unsurprising, as the legislation enforceable in the whole UK territory is promulgated in this parliament.

The BSLC means therefore a new crucial source of linguistic data for those in need of reliable sources of information regarding the recent legislation promulgated by British parliaments, as it provides a representative sample of the legislative production during the last 10 years in the UK.

Table 1: Distribution of the documents in the BSLC

Parliament	**Number of statutes**
House of Commons (UK Public General Acts)	370
Scottish Parliament	179
Northern Ireland Assembly	116
National Assembly for Wales (Senedd)	49
Total	714

Table 2: Types and Tokens in the BSLC

Parliament	**Types**	**Tokens**
House of Commons (UK Public General Acts)	473 933	9 663 939
Scottish Parliament	162 541	2 113 747
Northern Ireland Assembly	91 946	1 078 113
National Assembly for Wales (Senedd)	54 160	845 655
Total	782 580	13 701 454

4. A MULTIDIMENSIONAL ANALYSIS OF BRITISH LEGAL GENRES: STATUTE LAW VS. CASE LAW

As explained in Section 2, with the aim of contrasting the conclusions made by literature on the discursive features of British judgments and legislation, which mainly relied on qualitative methodologies, such as case studies and translational analyses (Section 1.1), a multidimensional analysis based on the data of the BLRC and the *ad hoc* BSLC corpora has been conducted. In this section the materials and methods, and results of the MD analysis are detailed. At the end of the section, there will be a discussion of the results, and the limitations encountered during the development of the study.

4.1. Materials and methods

In order to conduct a MD analysis suitable for making a comparison between British statutes and case law genres, the BLaRC (Marín & Rea Rizzo, 2012) and the corpus compiled *ad hoc* for this dissertation, the BSLC (Section 3), were used as our primarily source of data. These corpora were processed by the *Multidimensional Analysis Tagger* (Nini, 2019), a software developed to automatically POS tag the corpus provided and extract an updated version of the variables used by Biber's first MD analysis (Biber, 1988).

This software also automatically calculates the factor loadings in the corpus provided for the dimensions interpreted by the abovementioned analysis, but in this study, *ad hoc* factors were extracted to have a better insight on how legal British discourse behaves and the differences between the two genres object of study. Nevertheless, this last feature provided by the MAT is applied in Chapter 5 to obtain an insight on where these two legal genres are positioned in the landscape of different specialised genres analysed by Biber's MD analysis.

MAT extracted 68 variables from 1941 observations (documents in the corpus), 1229 from the BLaRC and 714 from the BSLC. The description of these

variables can be found in Data 1. Once the frequency of these variables in the corpora was obtained from the data processing performed by the MAT software, these data were imported into R to compute the FA using the libraries 'readtext', 'dplyr', 'psy', 'psych' and 'nortest'. Prior to the FA, several tests and analyses are needed to obtain essential information about the nature of the data for the selection of the proper adjusting parameters for the FA.

When conducting a FA using the function provided by these R libraries, the selection of the number of factors to be extracted and the rotation method for the analysis are required. The rotation method is a mathematical technique easing the factor structure depending on the needs of the researcher by highlighting some of the most relevant components (variables) of each factors when there is a high amount of cross-loadings (oblique methods), that is, high correlation between the variables in the factors, or, on the contrary, by emphasising less relevant variables when each factor is very independent (orthogonal methods) (Yong & Pearce, 2013). A correlation matrix is helpful to obtain that information, as it shows whether our dataset's variables are highly correlated to each other and therefore whether type or another of rotation method is better, by performing simultaneous Pearson's correlation tests with any combination of our variables.

Furthermore, in order to decide the number of factors to be extracted, a parallel analysis needs to be computed. This is a statistical method used in factor analysis to determine the optimal number of factors to retain. It works by comparing the eigenvalues from real data, which represent the amount of variance in the data that a particular factor or component accounts for, with those from randomly generated data. If a real eigenvalue is higher than the corresponding random eigenvalue (typically at the average or 95th percentile), the factor is considered meaningful; otherwise, it is likely noise. This method is more reliable than traditional approaches like the *eigenvalue > 1* rule or the scree plot, reducing the risk of over- or under-extracting factors (Joaristi & Lizasoain, 2008; Yong & Pearce, 2013).

The 'psych' library provides with the function *fa.parallel()* for this analysis, with two factoring methods available: the Minimal Residual Method (minres), for skewed and data with high variability, and the Maximum Likelihood Method (lm), for data with a normal and homogeneous structure. Kolmogorov-Smirnov with Lilliefors correction and a Fligner-Killen tests are previously conducted to determine the extent to what the data set variances are normal and homogeneous. Due to the significance of these tests, the dataset for this corpus requires a parallel analysis using the Minimal Residual Method.

Once the correlation matrix, and the normality and homogeneity tests together with the parallel analysis are conducted, the information regarding the nature of the dataset to determine the rotation method and the number of factors for the FA are available.

Usually, the parallel analysis allows the researcher to reduce the range of possible number of factors between 2 or 3 options. For narrowing this range until the best possible option, several FAs can be performed and compare their Bayesian Information Criterion (BIC). By selecting the FA with a lower absolute BIC value, the best fit (most information with the less factors possible) is obtained.

The results of the FA will provide a report with the factors extracted as well as their variables and corresponding loadings in the factor, accounting for their relevance in the composition of the factor. When dealing with a high number of variables, the lower loadings might not be relevant enough to consider, so literature using FA tends to use a cut-off between |0.30| and |0.35|. In this study, we used the |0.35| cut-off. (Biber & Conrad, 2019; Joaristi & Lizasoain, 2008; Yong & Pearce, 2013)

The purpose of extracting these factors by computing a FA is reducing the high number of possible variables that influence genre variation, disclosing a latent structure in which these variables are intertwined between each other. Thus, the resulting factors of the FA are object of interpretation by the researcher drawing on

existing literature, in an attempt to understand the logic behind the (positive and negative) correlation between the variables, in our case, linguistic features. This has been done with the results of our corpus drawing on previous research on legal discourse, translational analysis, MD analysis on specialised genres and MD analysis on English variation. The interpretation of textual dimensions will also consider the factor loadings of each of the factors in the corpus, that is, how present and in which side (positive or negative) this factor appears in the corpora analysed. Figure 5 provides with a visual summary of the methods and steps followed for the study in this chapter. The R code used for this section can be found in Code 3.

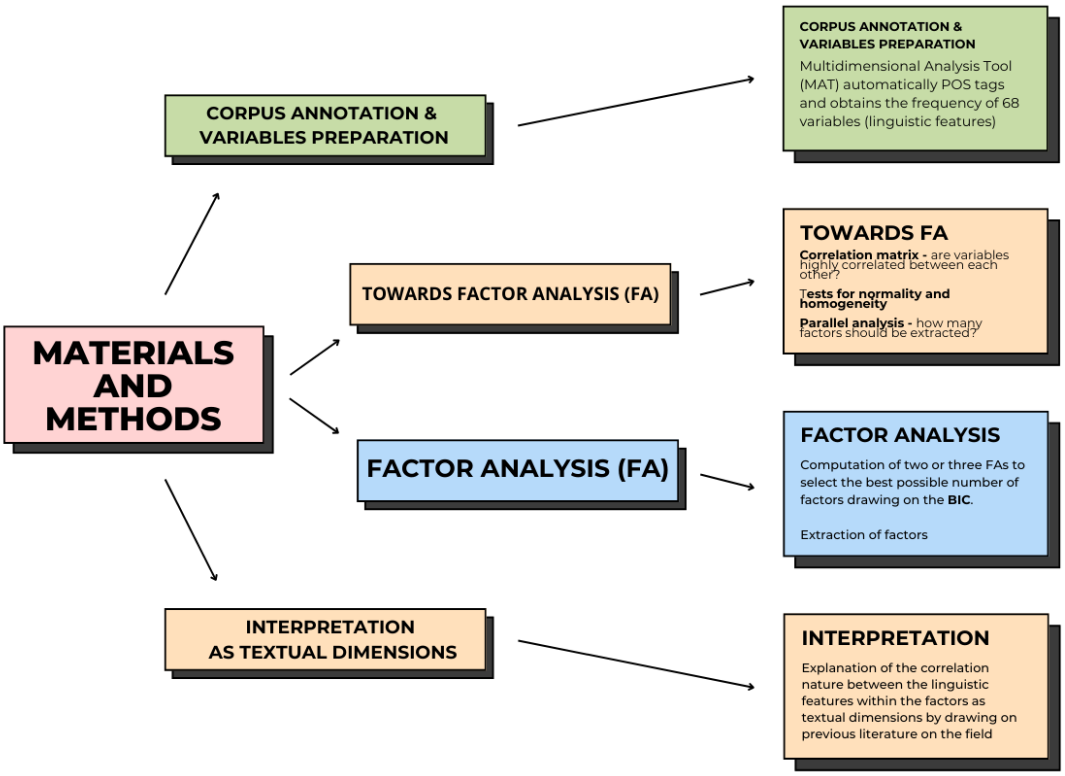


Figure 5: Materials and methods outline

4.2. Results

The descriptive statistical values of the 68 linguistic features extracted from the BSLC and the BLaRC by the MAT can be consulted in Data 2 and Data 3. The correlation matrix results show a high correlation between several combinations of linguistic features. Consequently, the promax rotation will be used for the FA, so it highlights the most relevant components of each factor (Revisit Section 4.1 for further explanation).

The factoring method selected was the Minimal Residual Method, due to the results from the tests of normality and homogeneity (Table 3). Both tests show a high test statistic variable, indicating the difference between the dataset concerned and a normal and homogeneous dataset, while the extremely low p-value indicates there is virtually no options that this result is due to random (in other words, the high significance of the results).

Table 3: Normality and homogeneity tests results

Test	**Test Statistic**	**p-value**
Lilliefors	0.43023	$< 2.2^{-16}$
Fligner-Killeen	88557	$< 2.2^{-16}$

The parallel analysis results showed that the variance representing the eigenvalues from the first 14 factors in the real data are higher than the ones from the simulated data (Table 4). Nevertheless, as Figure 6 helps us notice, around the sixth factor the difference of variance between the real and the simulated data might not be appreciable. Thus, three FA analyses will be computed, comparing the BIC resulted from extracting 5, 6 and 7 factors.

Table 4: Eigenvalues in real FA vs Random Simulations (Parallel Analysis)

Factor Number	Eigenvalue (Real Factor Analysis)	Eigenvalue (Random Simulations)
1	19.54	0.39
2	3.34	0.36
3	1.22	0.34
4	1.16	0.32
5	0.90	0.30
6	0.66	0.29
7	0.55	0.27
8	0.49	0.26
9	0.39	0.24
10	0.36	0.23
11	0.32	0.22
12	0.28	0.21
13	0.23	0.19
14	0.21	0.18
15	0.17	0.17

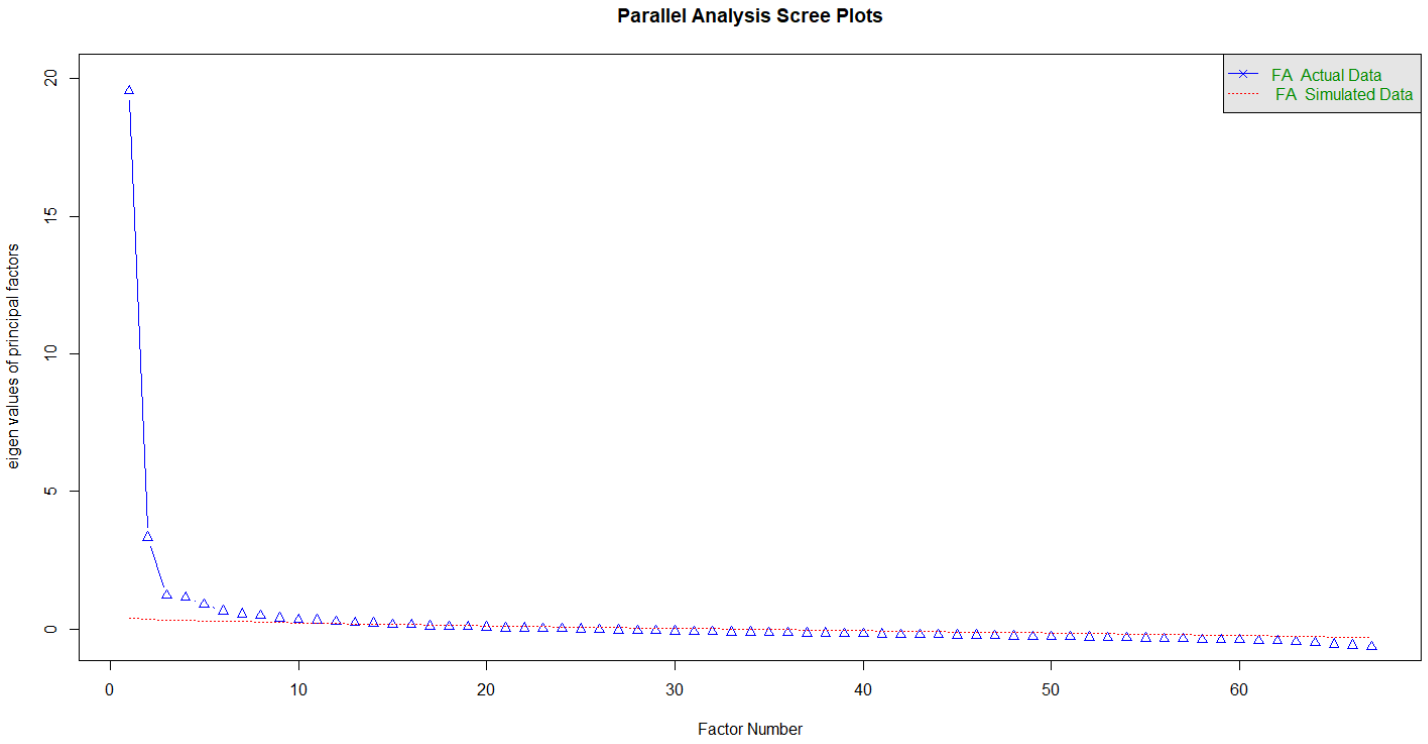


Figure 6: Scree plot (Parallel Analysis)

The Bayesian Information Criterion (BIC) results () from the three FA analyses computed showed that the FA extracting 6 factors is the most efficient as it has the lowest BIC absolute number. Extracting 5 factors would ignore too much information, while extracting 7 would create a model too complex for the information obtained in return.

Table 5: Bayesian Information Criterion results

BIC	**Number of factors extracted**
1288	5
121.19	6
-682	7

As a result, the FA extracting 6 factors is selected. After the application of a |0.35| for the reasons explained in Section 4.1, the resulting factors obtained were the ones displayed in Figure 7. Factor 1 is the most relevant factor for our corpora as well as the most complex one, accounting for almost a 20 % of the variance.

When interpreting the factors as dimensions, Factor 1 will be considered the most important when differentiating one genre from another, while the rest of the factor's eigenvalue only account for 3 % or less of the variance, so they will be much less relevant or useful for that purposes. Still, both the parallel analysis and the BIC indicates us that, even if only as complementary, the remaining 5 factors might provide with useful information about the differences between these two legal genres.

Factor 1	
Variable	Weight
Predicative adjectives	0.98
THAT verb complements	0.94
BE a main verb	0.9
Past tense	0.78
That relative clauses on object position	0.74
Private verbs	0.71
Analytic negation	0.68
Public verbs	0.67
THAT adjective complements	0.64
Synthetic negation	0.62
Perfect aspect	0.58
Pronoun IT	0.52
Third person pronouns	0.48
WH-clauses	0.46
Conjuncts	0.46
Subordinator THAT deletion	0.42
Existential THERE	0.42
Suasive verbs	0.41
Other advervial subordinators	0.39
Seem / appear	0.38
Causative adverbial subordinators	0.38
Demonstrative pronouns	0.38
Total adverbs	0.37
Total prepositional phrases	-0.4
Nominalisations	-0.45
Independent clause coordination	-0.63

Factor 2	
Variable	Weight
Amplifiers	0.72
Adverbs	0.69
Demonstrative pronouns	0.58
First person pronouns	0.57
Emphatics	0.55
Split auxilaries	0.54
Time adverbials	0.5
Third person pronuns	0.46
Concessive advervial sub.	0.46
Seem / appear	0.45
Downtoners	0.4
Predictive modals	0.38
Discourse particles	0.38
Causative adverbial sub.	0.38
Type-token ratio	0.35
Contractions	0.35
Total other nouns	-0.4

Factor 3	
Variable	Weight
Conditional adverbial sub.	0.62
Present tense	0.52
Possibility modals	0.49
Pied-piping relative clauses	0.45
Nominalisations	0.44
Necessity modals	0.37
Average Word Length	-0.51
Phrasal coordination	-0.59
Nouns	-0.99

Factor 4	
Variable	Weight
Average Word Length	0.77
Attributive adjectives	0.63
Phrasal coordination	0.56
Present participial WHIZ deletion relatives	0.45
Demonstratives	-0.37

Factor 5	
Variable	Weight
Past tense	0.54
Third person pronuns	0.44
Present tense	-0.53
Nouns	-0.56

Factor 6	
Variable	Weight
BE as main verb	0.63
Predicative adjective	0.57
Present tense	0.5

Figure 7: Factor analysis results

The descriptive statistical values for the factor loadings of each factor in the corpora are displayed in Data 4 and Data 5. A density plot comparing the distribution of the six factors in the BSLC (accounting for statute law or legislation) and the BLaRC (accounting for law reports or case law) is displayed in Figure 8. This plot shows that the distribution of the factors in the two genres have a very similar shape, even if it shows relevant differences in the values being more frequent in each genre. This is an indication of the relevance of the textual dimensions to be interpreted behind these factors are inherent for the configuration of these two legal genres, therefore the relevance of its extractions to better understand the discursive behaviour of legal English in the context of the UK.

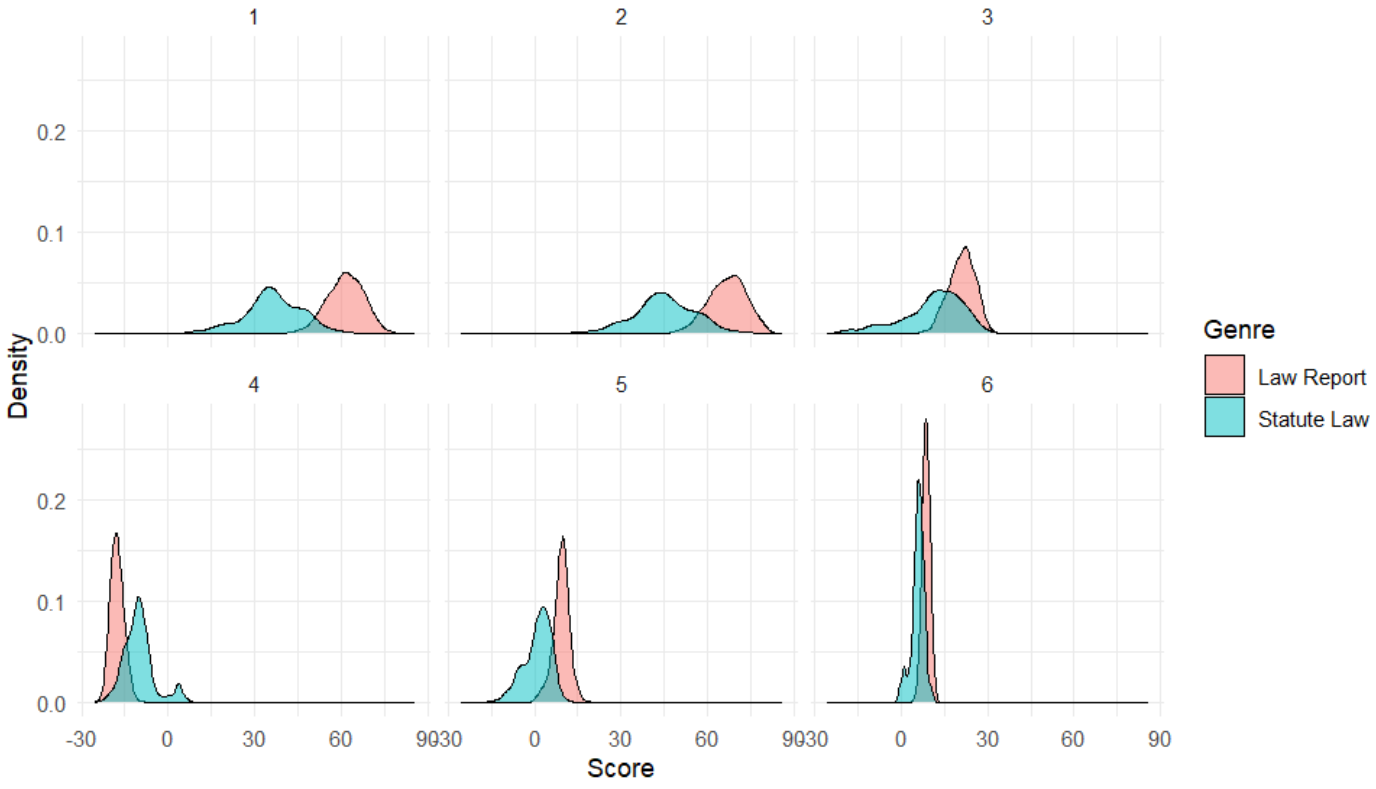


Figure 8: Comparison of the factor score distributions

4.3. Discussion

For the discussion of the results, firstly the factors extracted have been interpreted as textual dimensions, and, secondly, there is a discussion on the relevance of each textual dimension in the two genres object of comparison, law reports (BLaRC) and statute law (BSLC), relying on the factor scores and t students test statistics computed to determine whether the difference of values between the two corpora of each is significant, and what the possible reasons behind these differences are. This discussion will give a better insight on the latent discursive structure existent behind legal genres and how different this structure is depending on the genre.

4.3.1. Interpretation of factors as textual dimensions

In Factor 1, several linguistic features related to unstructured, unformal, intimate or persuasive discourse have positive weights. For instance, *predicative adjectives* and *BE as a main verb* are usually employed for making clear descriptions of states and conditions; *THAT verb complements*, *analytic negation* and synthetic *negation* are usually associated with instructive texts and oral legal texts, due to their ability to provide with explicit relations between ideas and reduce ambiguity; *relative clauses on object position* and *THAT adjective complements* add additional information about the main subjects of the text, increasing accuracy; *public verbs*, *past tense* and *present tense* are usually related to reasoning, argumentative and narration of facts, since they provide with temporal information as well as guiding through explicit statements; finally, *private verbs* and *the pronoun* IT are present in personal or persuasive utterances in an attempt to increase involvement and clear reference to abstract ideas (bringing abstract concepts to real daily situations). There are other variables such as *wh-clauses* or *subordinator THAT deletion*, but they all can be associated with similar functions, purposes or communicative situations. (Biber, 1988; Ehret & Taboada, 2021; Huang & Sang, 2024)

On the contrary, only three linguistic features have negative weights for this dimensions, and these are *prepositional phrases, independent clause coordination*

and *nouns*: both are consistently related to highly specialised or technical contexts, and the production of more complexed, with high lexical density and less-reader friendly discourse (Álvarez Álvarez, 2008; Biber, 2006; Granados-Meroño, 2023).

Regarding the context of legal discourse and the previous analysis of the uses, purposes and effects of the different linguistic features belonging to Factor 1, this has been interpreted as the textual dimension *'Guided vs. Unguided Recipient*'; every linguistic feature with positive weights contributes by means of clarity, involvement and descriptions to the guidance of the recipient through the arguments, narrations and opinions in the text, while the negative weighted features achieve precisely the oppositive, prioritising accuracy and complexity. Table 6 provides with an overview of the components of Dimension 1, their classification in the two-sided dimension and a brief description of their role.

This factor explained almost the 20 % of the variance in our corpora, being there for the most relevant for understanding the structure of correlations that the variables have showed. In other words, the resources and strategies of guidance (or lack of it) that the authors, in this case, judges and lawmakers, employ in order to make the legal texts more readable, convincing and familiar for the readers.

Table 6: Dimension 1 - Guided vs. Unguided Recipient

Variable	Weight	Classification	Description
Predicative adjectives	0.98	Guided	Adjectives used to describe actions (e.g., "The decision is important").
THAT verb complements	0.94	Guided	Subordinate clauses with "that" introducing verb complements (e.g., "I think that...").
BE as a main verb	0.9	Guided	Use of "be" as a main verb, often in passive or descriptive constructions.
Past tense	0.78	Guided	Verbs in the past tense, indicating narrative or description of past events.
That relative clauses (object)	0.74	Guided	Relative clauses with "that" in object position, adding specificity and detail.
Private verbs	0.71	Guided	Verbs expressing thought, perception, or feeling (e.g., "think", "feel").
Analytic negation	0.68	Guided	Negation using "not" or other auxiliaries (e.g., "do not"), providing clarity.
Public verbs	0.67	Guided	Verbs indicating communication (e.g., "say", "argue").

Variable	Weight	Classification	Description
THAT adjective complements	0.64	Guided	Subordinate clauses with "that" introducing adjective complements (e.g., "It is clear that...").
Synthetic negation	0.62	Guided	Negation using prefixes or suffixes (e.g., "unhappy", "impossible").
Perfect aspect	0.58	Guided	Use of perfect tenses to indicate relevance or connection to the present.
Pronoun IT	0.52	Guided	Use of the pronoun "it" for impersonal or referential purposes.
Third person pronouns	0.48	Guided	Pronouns like "he", "she", "they", referring to third parties.
WH-clauses	0.46	Guided	Clauses introduced by WH-words (e.g., "what", "where"), adding detail or explanation.
Conjuncts	0.46	Guided	Words or phrases that connect ideas (e.g., "however", "therefore").
Subordinator THAT deletion	0.42	Guided	Omission of "that" in subordinate clauses, making the text more conversational.
Existential THERE	0.42	Guided	Constructions like "There is/are", used to introduce new information.
Suasive verbs	0.41	Guided	Verbs expressing persuasion or recommendation (e.g., "suggest", "propose").
Other adverbial subordinators	0.39	Guided	Subordinating conjunctions introducing adverbial clauses (e.g., "although", "because").
Seem | appear	0.38	Guided	Verbs indicating appearance or perception (e.g., "It seems that...").
Causative adverbial subordinators	0.38	Guided	Subordinators expressing cause or reason (e.g., "because", "since").
Demonstrative pronouns	0.38	Guided	Pronouns like "this", "that", "these", "those", used to refer to specific elements.
Total adverbs	0.37	Guided	General use of adverbs to modify verbs, adjectives, or other adverbs.
Total prepositional phrases	-0.4	Unguided	Phrases starting with prepositions (e.g., "in the house"), often adding complexity.
Nominalisations	-0.45	Unguided	Turning verbs or adjectives into nouns (e.g., "decision" from "decide").
Independent clause coordination	-0.63	Unguided	Joining independent clauses with conjunctions (e.g., "and", "but"), making text less structured.

In Factor 2, there are several variables enhancing clarity, persuasion, and engagement on positive weights, while the negative weighted variables express more abstract and formal language. *Amplifiers*, *emphatics*, *split auxiliaries* and *downtoners* are some of the most common tools used in common oral language to add rhetorical force, clarity or emphasis, increasing or decreasing their commitment towards the utterance (Hyland, 2005). These are used in legal written discourse to mimic oral discourse (Alamri, 2023; Huang & Sang, 2024). *First person*

pronouns, *discourse particles* and *contractions* create a sense of dialogue, personal involvement or interactivity, and add informality, even in the written form (Biber, 2006; Huang & Sang, 2024; Shakir, 2024). Other features such as *adverbs*, *demonstrative pronouns* or *time adverbials* are context dependent tools that enhance clarity and a more readable reasoning. Finally, other positively weighted features are focused on expressing nuances, acknowledging counterarguments or conceding more than one possibility, such as *concessive adverbial subordination*, *seem* | *appear*, and *predictive modals* (Alamri, 2023; Biber, 1988; Ehret & Taboada, 2021; Huang & Sang, 2024).

The single negatively weighted variable, *nouns*, generates, in turn, a rigid, synthetic and dense discourse, usually relevant in written discourse and not followed or preceded by the orality concerned features previously mentioned.

These combination of linguistic features in the context of legal genres leads to the interpretation of Factor 2 as the textual dimension '*Elaborated Oral Discourse vs. Written Discourse'*, due to the common focus that the positively weighted features show on the reproduction of aspects related to oral language in the context of legal discourse (Alamri, 2023; Ehret & Taboada, 2021; Huang & Sang, 2024). The summary of the dimension is provided in Table 7.

Table 7: Dimension 2 - Elaborated Oral Discourse vs. Written Discourse

Variable	Weight	Classification	Description
Amplifiers	0.72	Elaborated Oral Discourse	Words that intensify meaning (e.g., "very", "completely"), adding emphasis.
Adverbs	0.69	Elaborated Oral Discourse	Words that modify verbs, adjectives, or other adverbs, adding precision and detail.
Demonstrative pronouns	0.58	Elaborated Oral Discourse	Pronouns like "this", "that", referring to specific points, creating focus.
First-person pronouns	0.57	Elaborated Oral Discourse	Pronouns like "I", "we", reflecting a personal or collective voice.
Emphatics	0.55	Elaborated Oral Discourse	Words that add emphasis (e.g., "indeed", "certainly"), strengthening arguments.
Split auxiliaries	0.54	Elaborated Oral Discourse	Constructions like "will **never** accept", adding rhythm and emphasis.

Variable	Weight	Classification	Description
Time adverbials	0.5	Elaborated Oral Discourse	Words like "now", "previously", situating events in time.
Third-person pronouns	0.46	Elaborated Oral Discourse	Pronouns like "he", "she", referring to parties or actors.
Concessive adverbial subordinators	0.46	Elaborated Oral Discourse	Words like "although", "even though", acknowledging counterarguments.
Seem | appear	0.45	Elaborated Oral Discourse	Verbs expressing appearance or perception (e.g., "It seems that...").
Downtoners	0.4	Elaborated Oral Discourse	Words that reduce intensity (e.g., "somewhat", "slightly"), adding nuance.
Predictive modals	0.38	Elaborated Oral Discourse	Modal verbs like "will", "might", expressing prediction or possibility.
Discourse particles	0.38	Elaborated Oral Discourse	Words like "well", "however", managing conversation flow.
Causative adverbial subordinators	0.38	Elaborated Oral Discourse	Words like "because", "since", explaining reasons or causes.
Type-token ratio	0.35	Elaborated Oral Discourse	A measure of lexical diversity, reflecting varied vocabulary.
Contractions	0.35	Elaborated Oral Discourse	Shortened forms like "can't", "won't", adding informality.
Total other nouns	-0.4	Written Discourse	General use of nouns, reflecting a formal, abstract style.

Factor 3 has 6 positively weighted linguistic features while 3 negatively weighted. On the positive side, *conditional adverbial subordination* and *possibility modals* express hypothetical statements, increase uncertainty and concede more than one solution or outcome for one situation (Biber & Conrad, 2019; Granados-Meroño, 2023). Present tense, pied-piping relative clauses and nominalisations are likely to be used in statements aimed at legal interpreting, concept construction or abstraction. Finally, *necessity modals*, though usually associated with formal, power distanced discourse, in legal contexts, these modals verbs precisely express opinions, views or decisions adopted by judges, rather than categorical or objectives realities being created by the law (Álvarez Álvarez, 2008; Granados-Meroño, 2023).

In turn, negatively weighted features, *average word length*, *phrasal coordination*, and *nouns* construe together analytic, direct and clear statements, leaving almost no option for counter argumentation. Therefore, the most suitable interpretation for this factor was the textual dimension *'Subjectivity vs. Objectivity'*. A summary of the dimension is included in Table 8.

Table 8: Dimension 3 - Subjectivity vs. Objectivity

Variable	Weight	Classification	Description
Conditional adverbial subordinators	0.62	Subjectivity	Subordinators expressing conditions (e.g., "if", "unless"), reflecting hypotheses.
Present tense	0.52	Subjectivity	Verbs in present tense, used for general principles or interpretations.
Possibility modals	0.49	Subjectivity	Modal verbs expressing possibility (e.g., "may", "might"), indicating uncertainty.
Pied-piping relative clauses	0.45	Subjectivity	Relative clauses with prepositions (e.g., "in which"), adding complexity.
Nominalisations	0.44	Subjectivity	Nouns derived from verbs or adjectives (e.g., "decision"), reflecting abstraction.
Necessity modals	0.37	Subjectivity	Modal verbs expressing necessity (e.g., "must", "should"), with flexibility.
Average Word Length	-0.51	Objectivity	Average length of words, associated with direct and clear language.
Phrasal coordination	-0.59	Objectivity	Coordination of phrases (e.g., "X and Y"), reflecting structure and clarity.
Nouns	-0.99	Objectivity	General use of nouns, associated with formal and fact-based language.

Factor 4 the positive weighted features are *Average Word Length (AWL)*, *attributive adjectives*, *phrasal coordination*, and *present participial WHIZ deletion relatives*. These features are related to the construction and enhancement of descriptive statements: *AWL* indicates the use of long words; *attribute adjectives* are associated with the modification of nouns to describe them more precisely, while *phrasal coordination* and *present participial WHIZ deletion relatives* contribute to the concise, clear and less dense addition of information (Biber, 1988; Ehret & Taboada, 2021).

In contrast, *demonstratives*, the single negatively weighted feature, appears as a negatively correlated feature to the rest of variables in the factor. This features is common in argumentative discourse, with the purpose of linking ideas appearing in different sentences of a text construing an argument, that is, connecting premises and conclusions (for further explanation of argumentation structures, consult Section 1.2.1).

This factor has consequently been interpreted as the textual dimension *'Descriptive vs. Argumentative Focus'* (Table 9).

Table 9: Dimension 4 - Descriptive vs. Argumentative Focus

Variable	Weight	Classification	Brief Description
Average Word Length	0.77	Descriptive Focus	Longer words, associated with technical or detailed descriptions.
Attributive adjectives	0.63	Descriptive Focus	Adjectives modifying nouns directly, adding specificity and detail.
Phrasal coordination	0.56	Descriptive Focus	Coordination of phrases (e.g., "X and Y"), reflecting structured and detailed language.
Present participial WHIZ deletion relatives	0.45	Descriptive Focus	Relative clauses with present participles, adding concise descriptive information.
Demonstratives	-0.37	Argumentative Focus	Pronouns or determiners like "this", "that", used to refer to specific points in arguments.

Dimension 2 explained a 3 % of the variance in the dataset, while dimensions 3 and 4 only around a 1 % of the variance. Still, even if the first dimension has by far the most important on the explanation of the variance, these three dimensions provide addition information that might be useful or complementary to the insights given by the first dimension. However, the last two dimensions, 5 and 6, account for less than 1 % of the variance. Still, they were included in the FA and interpreted as the rest of dimensions due to explaining more variance than in the random simulated data from the Parallel Analysis (Section 4.2). This hinders the interpretation of these last two dimensions and the possible insights given by it must be taken with caution.

Factor 5 has two positively sided variables, *past tense* and *third person pronouns*; and two negatively sided variables, *present tense* and *nouns*. These are features used in a wide range of contexts, the interpretation is therefore less accurate and more speculative. Still, the context given by the previous dimensions and the knowledge on legal discourse allows a better understanding of the possible reasons behind this combination of linguistic features. *Past tense* and *third person pronouns* opposed to *present tense* and nouns possibly indicates, on the one hand, a focus on the narration of (past) facts where there was an involvement of different (third) actors, and, on other hand, a focus on the argumentation of complex legal concepts, relationships or cause-effect / correlation linkages (Alcaraz, 2007; Álvarez Álvarez, 2008; Ehret & Taboada, 2021; Feteris, 2012; Granados-Meroño, 2023).

This would explain the correlation between past tense verbs and third person pronouns, while narrating past events in which different people are involved, and the correlation between present tense and nouns, in the creation of connections between different abstract legal concepts or reasonings. This makes reasonable, given the legal specialisation of the corpora, the interpretation of Factor 5 as the textual dimension *'Facts-narration vs. Legal Reasoning'*.

Table 10: Dimension 5 - Facts-narration vs. Legal Reasoning

Variable	Weight	Classification	Brief Description
Past tense	0.54	Facts-Narration	Verbs in past tense, describing events or actions that occurred.
Third person pronouns	0.44	Facts-Narration	Pronouns like "he", "she", "they", referring to parties or actors.
Present tense	0.53	Legal Reasoning	Verbs in present tense, expressing principles or applications of the law.
Nouns	0.56	Legal Reasoning	General use of nouns, reflecting abstract and concept-based language.

Finally, Factor 6 is the only factor with only positive-weighted variables: *BE as a main verb*, *predictive adjectives* and *present tense*. These all are features commonly associated with evaluative stance, that is, expressing judgments,

assessments or opinions, which would make sense in the legal context of judges conveying a judgment on precedents present in previous law reports, depending on their adequacy or relevance for the legal case in issue (Alamri, 2023; Granados-Meroño, 2023; Matulewska, 2014). Thus, this factor has been interpreted as the textual dimension *'Evaluative Stance Focus'* (Table 11)

Table 11: Dimension 6 - Evaluative Stance Focus

Variable	Weight	Classification	Brief Description
BE as main verb	0.63	Evaluative Stance Focus	Use of "to be" as the main verb, expressing states or qualities.
Predicative adjectives	0.57	Evaluative Stance Focus	Adjectives functioning as predicates, expressing judgments or evaluations.
Present tense	0.5	Evaluative Stance Focus	Verbs in present tense, expressing general principles or interpretations.

As a result of this interpretation, six dimensions reflecting different aspects of the discursive structure defining the nature of legal genres is obtained (Table 12), being the first of the dimension the one explaining more variance in the dataset, and thus, the most relevant to distinguish from genre from another and defining the genres as they are. The rest of dimensions might be nonetheless helpful to explain aspects or nuances differentiating both genres that the first dimension does not consider.

Table 12: Factors as textual dimensions

Dimension	Interpretation	Description
1	Guided vs. Unguided Recipient	Distinguishes between a style that leaves the recipient to infer meaning (unguided) and one that actively guides the recipient through the information (guided).
2	Elaborated Oral Discourse vs. Written Discourse	Reflects the distinction between language that mimics oral discourse (emphasis, interactivity) and more formal, written language.
3	Subjectivity vs. Objectivity	Captures the difference between interpretive and flexible language (subjective) and categorical, fact-based language (objective).
4	Descriptive Focus vs. Argumentative Focus	Distinguishes between a style that describes facts or characteristics (descriptive) and one that presents reasoning or justifications (argumentative).
5	Facts-Narration vs. Legal Reasoning	Separates the narration of facts (past tense, third person) from legal reasoning (present tense, abstract language).
6	Evaluative Stance Focus	Reflects a focus on the expression of judgments, evaluations, or interpretations by the author

4.3.2. Textual dimensions in British Legal Genres

The interpretation of the factors extracted as textual dimensions will provide with a much deeper and rich insight to understand the differences in factor loadings between the two genres (already showed in Data 4, Data 5, and Figure 8), that is, between law reports (from the BLaRC corpus) and statute law / legislation (from the BSLC).

In this section, results of a t student comparing the mean values of the factor scores loaded in each genre are provided, along with a possible explanation behind the differences in the factor loadings between the genres, as well as excerpts from the corpora showing how these factors act in the text. These excerpts were obtained using SketchEngine tools in https://www.sketchengine.eu/ (Kilgarriff et al., 2004, 2014).

Factor 1 scores' mean values are significantly different in law reports and legislation, as the t test results show in Table 13.

Table 13: t test results for Dimension 1 Factor Scores

Statistic	Value
t-value	58.21
Degrees of Freedom (df)	1043.4
p-value	$< 2.2^{-16}$
Confidence Interval (95%)	[24.10, 25.78]
Mean (Law Report Group)	61.39
Mean (Statute Law Group)	36.44

The t test results show the mean values are sufficiently different to consider that this variation between the Factor 1 scores in one genre and another are not due to random. Moreover, returning to Data 4 and Data 5 a SD of around 6 and 10 is observed, with a trimmed mean value very near to the standard mean value. Thus, the insights from the boxplots in Figure 9 are valuable for the analysis. In this figure (and from the descriptive values) a conclusion arises: both genres skew towards the positive side of Dimension 1, that is, *Guided Recipient*, meaning that the authors

(judges and lawmakers) in both genres use resources and strategies to guide to certain extent the reader along the text. This is expectable, as legal texts are produced by highly specialised speakers, with a deep knowledge of the legal discipline. Nonetheless, according to these values, the *statute law* genre shows a lower degree of guidance for the recipient (around 25 points lower than *law reports*). This might be explained by the fact that judges tend to use a considerable amount of discourse markers, expressions and metadiscourse resources in general (Hyland, 2005) to make the judgments more readable to the parts involved in a trial.

Even if the solicitors working for the private parts involved in the case are the ones who will read the most these texts, non-expert citizens are also interested to understand, at least, the general reasoning the judge uses for their ruling. Moreover, when facing a trial of the interest of the public opinion, these judgments are also object of interpretation by journalists or other citizens that are not familiar with legal terminology or even the factual context of the case.

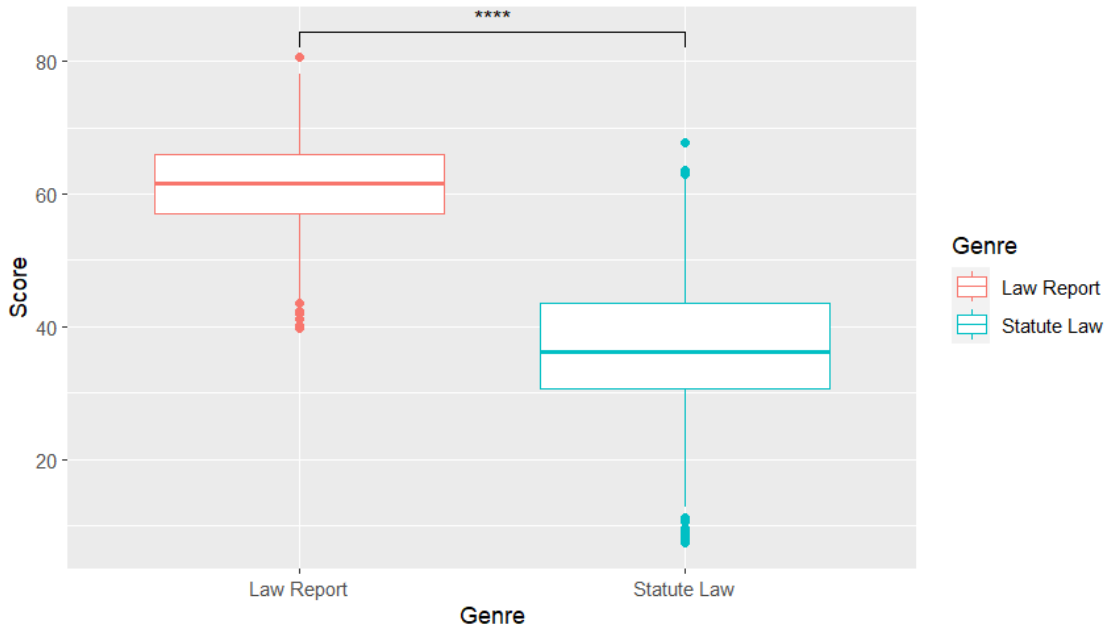


Figure 9: Factor scores in Dimension 1 with t test p-value results

On the contrary, lawmakers certainly prefer to prioritise accuracy and technical correction when developing new legal provisions, rather than making

them accessible to the citizens, since the reader and interpreter of these provisions will be highly specialised recipients, that is, the judges that will apply them, or the law professionals consulting them to assist their clients. In fact, ironically, the judges will be the ones explaining the law to the public through their legal reasoning expressed in judgments. In Excerpt 1 the combination of the positively-weighted features in Dimension 1 produces an organised, easy to follow, though formal and technical, discourse. Some examples are *public verbs* (say, report, give for), *THAT verb complements, predicative adjectives* (crucial, proactive), *perfect aspect* (had acquiesced, had been included, has already tried), or nominalisations (exercise, misrepresentation).

Excerpt 1: Examples of Guided recipient discourse in law reports

A. *He went on to **say that** whilst some progress was **reported** in assessment work, little had fundamentally changed within the couple dynamics. In oral evidence the ISW reiterated his disappointment that the parents **had not been** more **proactive**. He **pointed out** that the parents needed to work openly and honestly with professionals and that this is **crucial** as the foundation of success if those professionals are to have the evidence to gauge how the parents will behave in the future. He was referred to the fact that Mother had allowed her own daughters overnight contact with their step-grandfather at a time when Mother was saying that he **had** sexually **abused** her as a child of similar age. Father **had acquiesced** in this, just as he appeared to **have acquiesced** in the proposal that mother own father should be put forward as a carer for the children.*

B. *But that is **not** the question in this case. The question is whether Mr Koshy **has lost** his right to make an application to the court to set aside the order of Harman J dated 20 March 1998. Mr Koshy **has already tried** a number of ways of achieving his end. Importantly, he obtained permission to appeal the order of Harman J, **not** on the grounds **that** the **exercise** of the discretion to order costs was erroneous but on the grounds **that** the judge **had been induced** to make the order by **misrepresentation** by DEG. The **misrepresentation** alleged was as to the date **when** it discovered **that** (on its case) Mr Koshy and Lasco **had deceived** it as to the true cost of their investment in GVDC. But **that** appeal **failed**. I **will need to examine** in detail below some of the exchanges with counsel in the course of argument and the reasons **which** this court **gave** for its decision.*

That is the case as well in the statute law texts from our corpus, since the mean has a positive value, but significantly lower than the one in law reports. In Excerpt 2, negatively-weighted features of Dimension 1 are showed in the BSLC. It

is easily perceived how the consistent use of *complex prepositional phrases* (such as, given effect to), *nominalisations* (amendment) and *independent clause coordination* (services provided or to be provided; he services referred to in subparagraph (7) have been provided, and (b) if applicable, the reasons why the Commission has not provided any of the services referred to in sub-paragraph (7) in both official languages).

Excerpt 2: Examples of Unguided recipient discourse in statute law

*The Scheme must identify those **services provided or to be provided** in the official languages and explain **how those services are to be provided** in accordance with paragraph 8(5)**.** (8) The **Assembly Commission** must, in respect of each financial year**,** lay before the Assembly a report setting out **how the Commission has, during the year in question, given effect to** the Scheme. (9) The report prepared by the **Assembly Commission** under subparagraph (8) must include– (a) whether and to what degree the **services** referred to in subparagraph (7) **have been provided, and** (b) if applicable, the **reasons why the Commission has not provided** any of the **services** referred to in sub-paragraph (7) in both official languages. (10) The **Assembly Commission**– (a) must review the Scheme as soon as is reasonably practicable after each ordinary general election, **or** after an extraordinary general election to which section 5(5) applies, and (b) may, at any time, adopt a new Scheme or an **amendment** to the existing Scheme.*

Factor 2 scores' mean values are also significantly different in law reports and legislation. The results of the t tests are shown in Table 14. Mean values are sufficiently different to consider that the variation between the scores between genres are not due to random. The SD values are similar to the ones in Dimension 1.

Table 14: t test results for Dimension 2 Factor Scores

Statistic	Value
t-value	47.452
Degrees of Freedom (df)	1021.6
p-value	$< 2.2^{-16}$
95% Confidence Interval	20.74242 - 22.53194
Mean (Law Report)	66.7366
Mean (Statute Law)	45.09942

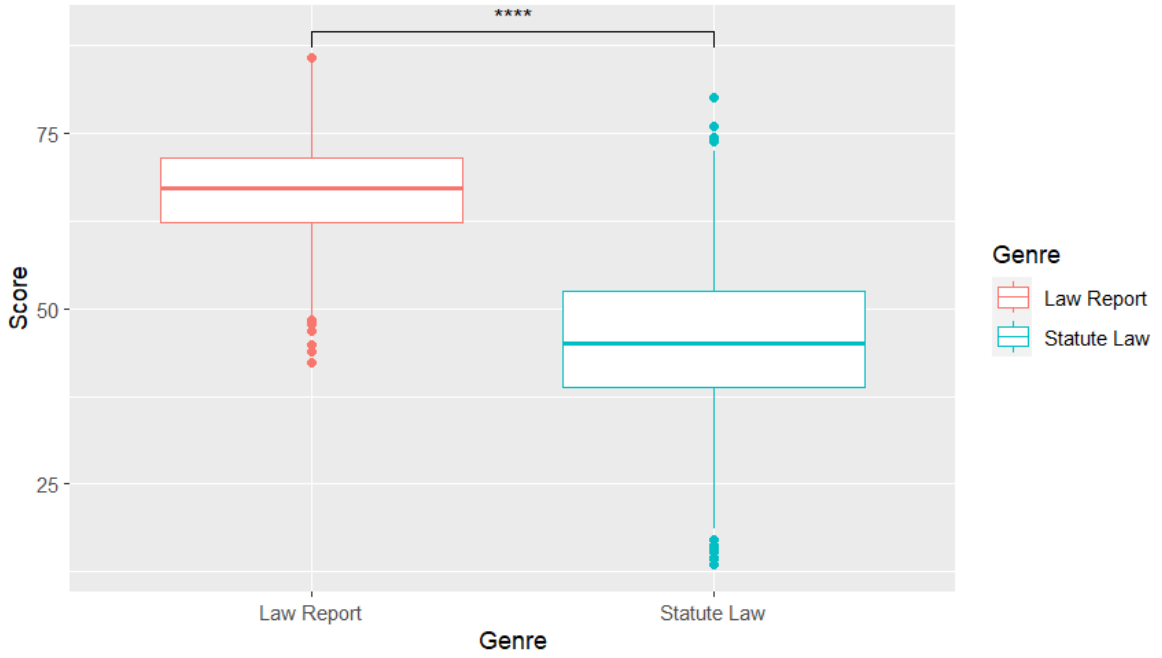


Figure 10: Factor scores in Dimension 2 with p-value results

Figure 10 shows how both genres are skewed towards the *Elaborated Oral Discourse* side of the dimension, but the statute law genre does it with a lower score. The higher score in law reports in this dimensions is certainly explained by the 'legal orality' phenomenon encountered in legal documents written by judges (such as judgments, court orders or providences), as they function consistently as a response to documents written by other judges. Thus, even if formal archaic or technical, there exists a conversation between the judges that is reflected on some of the expressions used by the authors of these type of documents. On the contrary, lawmakers do not communicate with other professionals when elaborating provisions.

In Excerpt 3 the use of adverbs, the first person pronouns, concession or time adverbials give a sense of orality as if the judge were talking to someone, despite the highly formal and sometimes archaic language.

Excerpt 3: Examples of Elaborated Oral Discourse in law reports

A. ***I*** *take into account the importance of ensuring that parties* ***actively*** *pursue their cases and keep in contact with their representatives.* ***I*** *adopt the findings of the employment judge that the Claimant failed to keep in contact with her solicitor or respond particularly to the email of 2 April. As a result, she placed herself out of contact with the Tribunal when the unless order was made. Having* ***finally*** *learnt of that unless order, she* ***promptly*** *applied for relief on 1 August.* ***Whilst I*** *am not persuaded that she intentionally failed to comply with the unless order, she has no good explanation for her failure to do so.*

B. ***We*** *do not accept that submission. Mr McCartan claimed to be unable to remember his telephone number when he was interviewed about this; he accepted that the 0 0 8 telephone could have been his but said that he had sold his telephone* ***a few weeks previously****.* ***But*** *the record of 'Ricky' against the number of this telephone in Mr McKinley's mobile phone and on the workplace name tag of Isobel Laing was ample evidence that Mr McCartan did* ***indeed*** *own that telephone, in our judgment.* ***Again****, his failure to give evidence on this crucial issue fully warranted the drawing of an adverse inference against him.*

In turn, legal provisions (statute law) usually are conceived as much more aseptic, concise texts, with no interpretations, opinions or evaluations on any matter. As Excerpt 4, there is high density in the use of highly formal or technical nouns (even one after another), used in third person singular and present tense, whilst there is a rarer presence of the positively weighted variables.

Excerpt 4: Example of Written Discourse in statute law

(5) The ***Scheme*** *must include (amongst other things)* ***provision*** *about– (a) simultaneous* ***interpretation*** *from one official* ***language*** *into the other– (i) in all* ***Assembly proceedings****, (ii) in public* ***meetings*** *conducted on behalf of the* ***Assembly Commission****, and (iii) in such other meetings connected with the functions of the Assembly or the Assembly Commission as may be provided for in the Scheme, (b)* ***publication of documents*** *in both official languages, subject to any exceptions identified in the Scheme, (c) public* ***engagement*** *with– (i) Assembly proceedings, and (ii) other functions of the Assembly and of the Assembly Commission, through the medium of either of the official languages [..] (f) the* ***allocation of responsibilities*** *for implementing the Scheme, (g) objective means of measuring progress in implementing the Scheme, and National Assembly for Wales (Official Languages) Act 2012 (anaw 1) 3 (h) a strategy for ensuring that the staff of the Assembly have, collectively, the language skills necessary to enable the Scheme to be implemented (6) The Scheme must include provision relating to the receipt,* ***investigation*** *and* ***consideration of complaints of failures*** *to give effect to provisions of the Scheme*

Factor 3 scores' mean values also show significant differences between the two genres. These results are in Table 15 and Figure 11. In this case, even if significantly different, a more similar value between the two genres is observed. Dimension 3 opposes subjectivity vs. objectivity, therefore the conclusion that law reports show more nuances of subjectivity than statute law is reasonable, and in accordance with the results in the previous dimensions. Again, law reports are related with a more guided, oral and persuasive discourse than statute laws, where objectivity is typically associated with written specialised texts.

Table 15: t test results for Dimension 3 Factor Scores

Statistic	**Value**
t-value	23.717
Degrees of Freedom (df)	871.73
p-value	$< 2.2^{-16}$
95% Confidence Interval	9.250481 - 10.919667
Mean (Law Report)	21.05301
Mean (Statute Law)	10.96793

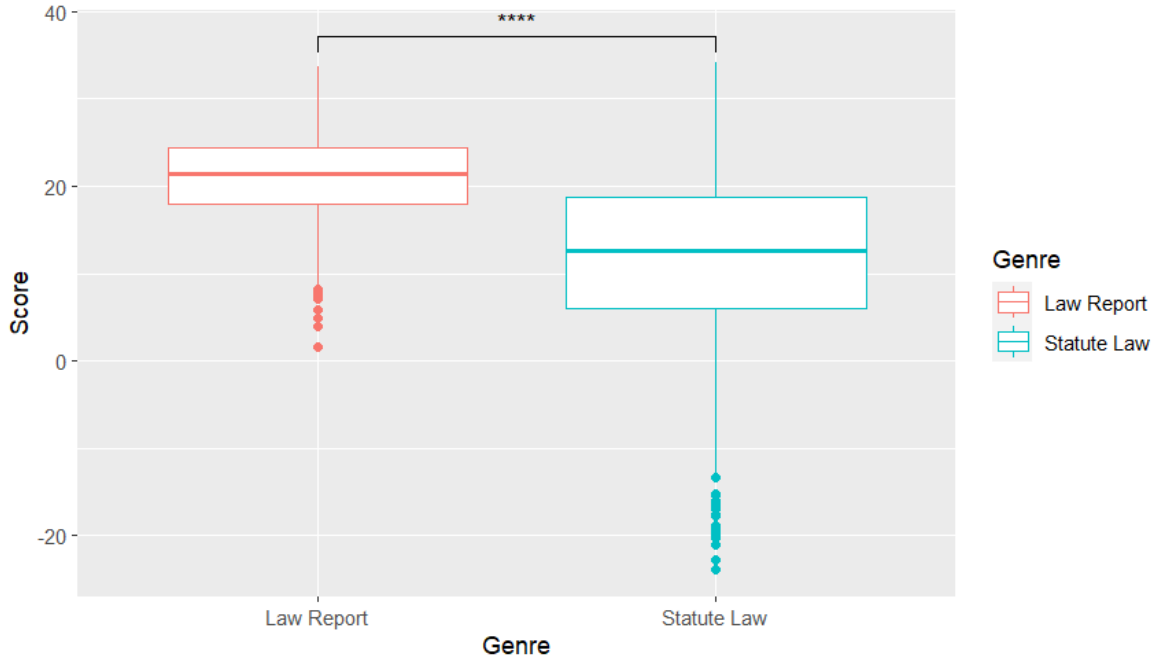


Figure 11: Factor scores in Dimension 3 with p value results

Excerpt 5 shows the presence of *possibility modals* reducing certainty (should, might, would), *conditional adverbial subordination* and pied-piping relative clauses (the submission was that Mr Koshy…), some of the variables increasing the subjectivity expressed in the statements expressed by the judge (either their own or other involved people's). In turn, Excerpt 6 shows the high presence of long and specialised *nouns* (several being nominalisations), and *phrasal coordination.* This combination conveys a much more distant, technical and objective message than the one portrayed in law reports.

Excerpt 5: Examples of Subjectivity in law reports

A. ***The submission was that what Mr Koshy could and should have done*** *was to make an application to me after the delivery of my judgment on 26 October 2001 for an order discharging Harman J's original ex parte freezing order dated 8 November 1996. Any success on that* ***would*** *not, by itself, have resulted in a reversal of the inter partes Harman Order, but* ***if*** *I had made a finding that the ex parte order ought to have been discharged for deliberate non-disclosure or misleading,* ***that would have provided*** *a proper factual basis for an appeal against the Harman Order.*

B. *By this point in the argument****, it is apparent that*** *the court was disenchanted by the prospect of hearing an appeal against the Harman Order at which further evidence which would be subject to cross-examination - was to be adduced. Mummery LJ suggested that the way forward* ***might*** *be to regard Mr Page as having identified sufficient material: " to enable us to direct an issue to the trial [sic: to be tried?], not by us [but?] by people who try issues? That is the issue of non-disclosure to impact on the correctness of the order for costs. The last thing I am going to allow is this court to be turned into conducting a trial by admitting evidence and then having cross-examination, having discovery.*

Excerpt 6: Examples of Objectivity in statute law

Anything which is in the ***process*** *of being done by the* ***Alcohol Education and Research Council*** *under an* ***enactment*** *immediately before* ***abolition*** *may be continued by the* ***Secretary of State****. (2) Anything which the* ***Council*** *is required to do under an* ***enactment*** *before* ***abolition*** *may, in so far as it has not been done by the Council, be done by the Secretary of State after abolition. (3) The Secretary of State must prepare a* ***report*** *on the activities of the* ***Council*** *during the* ***period*** *that begins with the 1 April before* ***abolition*** *and ends with* ***abolition****. </s><s> (4) In this* ***paragraph****– "abolition" means the* ***commencement*** *of section 278(1); "****enactment****" includes an enactment contained in subordinate* ***legislation*** *(within the meaning of the* ***Interpretation*** *Act 1978)*

Factor 4 scores' mean values show significant differences between the two genres. Results are in Table 16: t test results for Dimension 4and Figure 12. Both genres show negative values in their mean, so they tend to show features associated with the Argumentative Focus. This goes in accordance with the fact that among both legal genres' purposes are the explanation, justification and defence of a change produced by the author. In the case of judges, when arguing the reasons and legal precedents backing their ruling, while lawmakers, when explaining the reasons that make necessary the promulgation of that piece of legislation.

Nevertheless, the argumentation is the main focus of the judgments, as it resides in the nature of this genre that the judges properly justify their ruling, while the main focus of lawmakers, even if they need to give reasons to promulgate the new law, the accurate description of the intricacies of the new provisions is equally important.

Table 16: t test results for Dimension 4 Factor Scores

Statistic	**Value**
t-value	-35.87
Degrees of Freedom (df)	879.58
p-value	$< 2.2^{-16}$
95% Confidence Interval	-0.82371
Mean (Law Report)	-17.70895
Mean (Statute Law)	-10.18188

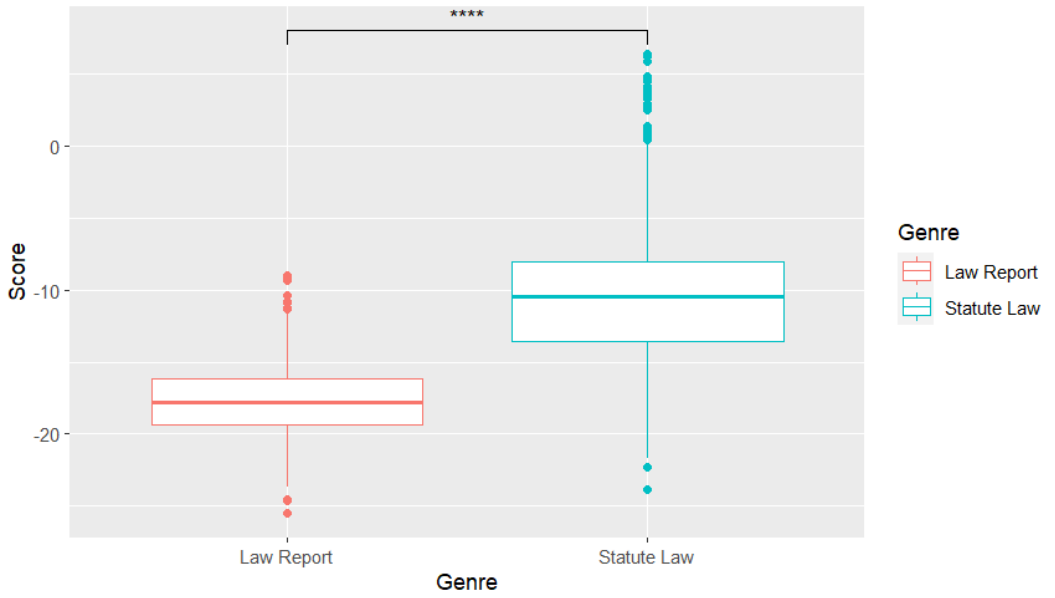


Figure 12: Factor scores in Dimension 4 with p-value results

A slightly more focus on description can be perceived in Excerpt 7, where there is an important presence of attributive adjectives, sometimes being nouns modified by two simultaneous adjectives, and WHIZ deletion relatives, for a concisely addition of information.

Excerpt 7: Example of Descriptive Focus in statute law

*Following the **consultation**, the authority must consider the responses and decide **whether making a byelaw is the most appropriate way of addressing the issue**. Local Government Byelaws (Wales) Act 2012 (anaw 2) 3 (4) The authority must then publish on its website a **second written statement** which contains – (a) **the initial written statement**; (b) a summary of the consultation and the responses; (c) its decision; (d) the reasons for that decision. At least six weeks before the byelaw is made, notice of the intention to make the (5) A byelaw must be published – (a) in one or more **local newspapers circulating in the area to which the byelaw is to apply**; (b) on the authority's website. (6) For at least six weeks before making the byelaw, the authority must ensure that – (a) a draft of the byelaw is published on the authority's website; (b) a copy of the draft is deposited at a place in the authority's area; (c) a copy is open to **public inspection** at all **reasonable hours** without payment; (d) where applicable, a copy is sent to all community councils whose areas the authority thinks are likely to be affected by the byelaw*

In turn, the high density of demonstratives shows the strong argumentative focus, in an attempt to properly and clearly connect different sentences as parts of an argument, is showed in the examples from Excerpt 8.

Excerpt 8: Examples of Argumentative Focus in law reports

*In the proceedings before Harman J, DEG alleged that Mr Koshy had made two fraudulent misrepresentations, first as to the cost of the funds which Lasco had invested in GVDC, and, secondly, as to the ownership of Lasco. DEG lost **those** proceedings at trial. 3. In brief, on the issue of the original proceedings, DEG made an application to Harman J for worldwide freezing orders on an interim basis. Mr. Koshy and Lasco made an application to discharge **those** orders which were dismissed, and they were ordered to pay the costs of **that** application in any event. It is **that** order for costs, which is at the heart of **these** proceedings. The costs were very substantial. They have not been assessed, but DEG has served a bill of costs in the sum of £ 359,415. The order made by Rimer J at trial meant that the freezing orders were then discharged, but that did not affect the order for costs.*

Table 17 shows the results from the t test comparing the mean values from the two genres regarding Factor 5 scores. These indicate that the difference between these values is significantly difference, and, together with Figure 13, that law reports skew slightly towards the *Facts-narration* side of the dimension, while statute law remains rather neutral, barely surpassing the score 1.

Table 17: t test results for Dimension 5 Factor Scores

Statistic	Value
t-value	42.747
Degrees of Freedom (df)	1023.3
p-value	$< 2.2^{-16}$
95% Confidence Interval	7.829803 - 8.583244
Mean (Law Report)	9.244483
Mean (Statute Law)	1.03796

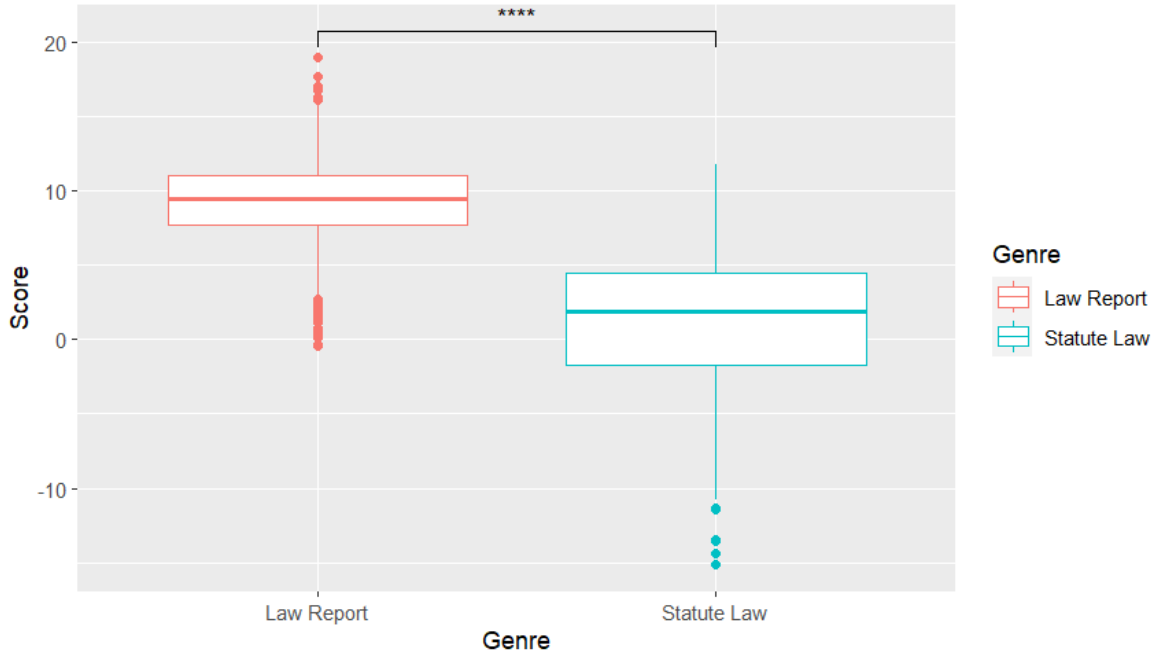


Figure 13: Factor scores in Dimension 5 with p value results

The slight predominance of the narration of facts over legal reasoning in the discourse produced by judges in their judgments makes perfect sense when considering the macrostructure of the genre, which has a whole section (usually of a considerable length) exclusively dedicated to the narration of events prior to and during the trial, that is, the so-called *facts in issue* section (Álvarez Álvarez, 2008; Goźdź-Roszkowski, 2020; Granados-Meroño, 2023). Excerpt 9 clearly shows how the use of *past tense* in *third singular person* is part of the nature structuring the narration of past events, and a relevant section in any judgment.

Excerpt 9: Example of Facts-narration in law reports

*The conflict between the parties **arose** from an agreement on mutual investment by Mr Koshy and DEG in GVDC. Mr Koshy's associated company, Lasco, **made** a loan to GVDC as **agreed** between the parties. The amount of the loan **was** 56.4 m Zambian kwacha ("K"). At the time, the dollar equivalent of this amount **was** approximately US $ 5.8 million. But Lasco **had acquired** the K 56.4 m for US $ 1.4 million using a system known as "pipeline dismantling ", which **was** available in Zambia. It is described in detail by Rimer J in [14] to [18] of his judgment following the trial of the original proceedings (reported as DEG-Deutsche Investititions- und Entwicklungesselschaft mbH v Koshy [2002] 1 BCLC 478). This **was** an official method of obtaining domestic currency in return for foreign currency. As a result of using the system, Lasco **increased** the potential profit on its investment. GVDC's farming project, however, **collapsed** and GVDC **went** into receivership.*

With regards to statute law, Excerpt 2, Excerpt 4, Excerpt 6 share the consistent use of a dense, highly specialised variety of *nouns* within *present tense* sentences that convey what is allowed, supported, enhanced, prohibited or discourage and through which processes by the enactment of the law.

Factor 6 t test results in Table 18 and show that the difference between the mean values is significant, although very scarce (around 3 points). This means that law reports present a higher degree of *Evaluative Stance* than statute law, but the difference might be hard to appreciate.

Table 18: t test results for Dimension 6 Factor Scores

Statistic	Value
t-value	31.585
Degrees of Freedom (df)	1042
p-value	$< 2.2^{-16}$
95% Confidence Interval	2.689762 - 3.046107
Mean (Law Report)	8.743716
Mean (Statute Law)	5.875781

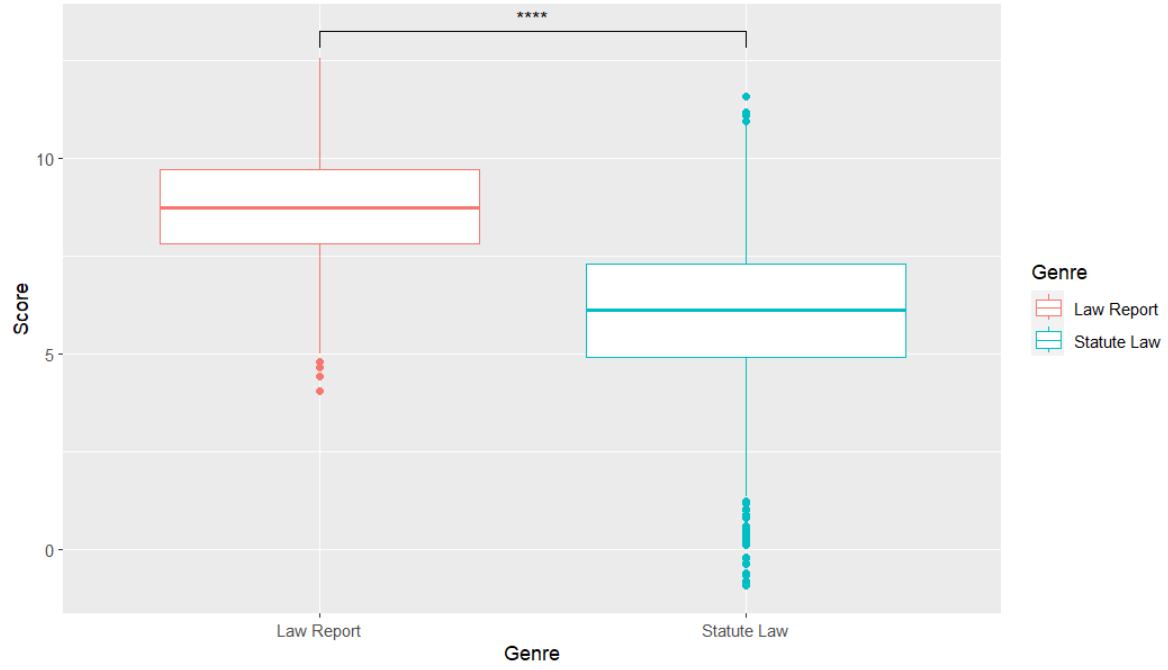


Figure 14: Factor scores in Dimension 6 with p value results

This higher degree of evaluation in law reports can be explained by the fact that judges tend to evaluate, assess, or give opinions on rulings made in previous trials that might support or argue against their own ruling. Judges will describe the legal reasoning on which previous rulings drew on, and then evaluate their accuracy or adequacy, giving their own opinion on those rulings. When they agree on that reasoning, they will use it as a precedent to support the ruling they are creating in the judgment in issue. They can also deny their effects or their relation to the current case or simply argue against that ruling, giving the proper reasons for it.

In fact, the structure created by the variables in this dimension, *BE as main verb*, *predicative adjectives*, and *present tense* is precisely the most representative one to express evaluation or assessment (x is x). Excerpt 10 shows how judges do make evaluations on other judges' decisions or hypothetical rulings different from the one is adopting.

Excerpt 10: Example of Evaluative Stance in law reports

When Rimer J made his findings of fact the issue of non-disclosure by DEG to Harman J was not before him. He was not addressing that issue. ***I do not think that it is satisfactory*** *simply to lift findings of material fact out of his judgment and use them without more to set aside the Harman Order****. In my judgment, it would be wrong and potentially unfair*** *to DEG in these circumstances for the court to set aside the costs order made by Harman J. If the exercise of discretion is to be reviewed in circumstances of alleged material non-disclosure with a view to making a different order for costs,* ***it can only be fairly and satisfactorily done in this case by an application at first instance*** *[…]*

On the contrary, this evaluative structure is only found as a mere condition by which one part of the provision is of application, letting the evaluative task to third parties, as can be observed in the examples of Excerpt 11.

Excerpt 11: Examples of Evaluative Stance in statute law

(2) But the Welsh Ministers may only implement a recommendation with modification ***if****– (a) in a case involving recommendations for change to electoral arrangements for a principal area,* ***they have considered the matters described in section 30 and are satisfied that it is appropriate*** *to make the modification, Local Government (Democracy) (Wales) Act 2013 (anaw 4) 20 (b) in a case involving recommendations for change to electoral arrangements for a community,* ***they have considered the matters described in section 33 and***

> ***are satisfied that it is appropriate*** *to make the modification, and (c) in any case,* ***they are satisfied that the modification is in the interests of effective and convenient local government****.*

4.4. Conclusions

MD analysis applied to the most important genres of British Public Law, law reports and statute law, emerges as a powerful tool that provides with a comprehensive insight on the configuration of the discursive structure by which judges and lawmakers fulfill their communicative and performative purposes in the writing of judgments and legislation. This analysis helps researchers in linguistics, translation studies, law and NLP better understand the combination of linguistic features that are of essence in the structure of these genres, and which are the ones that make them distinguishable from one to the other.

As a final summary of the results and the insights provided for them in the discussion section, a summary table is provided in Table 19.

Table 19: Textual Dimensions for British Legal Public Genres

Dimension	Legal Genre	Prevalent Focus
1. Guided vs. Unguided recipient	Law Reports	Guided recipient
	Statute Law	(Less) Guided recipient
2. Elaborated Oral Discourse vs. Written Discourse	Law Reports	Elaborated Oral
	Statute Law	(Less) Elaborated Oral
3. Subjectivity vs. Objectivity	Law Reports	Subjectivity
	Statute Law	Both
4. Descriptive vs Argumentative Focus	Law Reports	Argumentative
	Statute Law	(Less) Argumentative
5. Facts narration vs. Legal reasoning	Law Reports	Facts-narration
	Statute Law	Legal Reasoning
6. Evaluative Stance Focus	Law Reports	Evaluative Stance Focus
	Statute Law	Non-Evaluative

5. REGISTER VARIATION ACROSS ENGLISH GENRES: AN ELABORATION ON PUBLIC LEGAL GENRES

The MD analysis in Chapter 4 has provided with an exhaustive insight on the discursive nature of British public legal genres (law reports and statute law / legislation). This analysis has concluded that British public legal genres are defined by six different textual dimensions (Table 19), each one containing a cluster of different linguistic features that combined allow the utterers to achieve certain communicative or performative purposes, in this case, all of them belonging to the usual found in literature on Legal English discourse (Álvarez Álvarez, 2008; Goźdź-Roszkowski, 2011; Granados-Meroño, 2023).

However, although being revealing in terms of what public legal genres really look like, and what features are the one making them indistinguishable from each other, this MD analysis' scope of study is highly restricted, as only considers the universe of legal discourse, and makes distinctions within it, but it does not describe these genres as opposition to genres belonging to other expertise or communicative contexts, such as science, journalism, social media or oral conversations.

To have a complete understanding of the nature of the discourse produced in legal genres, the automated MD analysis computed by the Multidimensional Analysis Tagger or MAT (Nini, 2019) is considered and discussed in this chapter.

5.1. Materials and methods

This study has used the same structure and design in terms of corpora from the study in the previous chapter, explained in Section 4.1. In turn, this study, instead of preparing and computing a new FA has used the automated FA computed by the MAT, which uses a set of predefined textual dimensions that applies to the corpus provided. These textual dimensions are an updated version of the ones used by Biber on his original MD analysis (1988), and they provided a common framework

for the comparison of discursive structures among considerably different genres in terms of mode, purposes, register or specialisation, from TV series, radio broadcasts, to scientific papers. The set of predefined textual dimensions is provided on Table 20.

Table 20: Nini's (2019) adaptation for Biber's original textual dimensions (1988)

Dimension	Description	Linguistic features
1. Involved versus informational production	High scores: affective/ interactional text Low scores: informationally dense discourse.	Involved production features: that-deletions, contractions, present tenses, second person pronouns, do as pro-verb, analytic negation, demonstrative pronouns, emphatics, first person pronouns, pronoun it, be as main verb, causative subordinators, discourse particles, indefinite pronouns, hedges, amplifiers, sentence relatives, wh-questions, possibility modals, non-phrasal coordinations, wh-clauses, stranded prepositions. Informational production features: nouns, average word length, prepositions, type/token ratio, attributive adjectives.
2. Narrative versus non-narrative concerns	High scores: narrative text	Narrative concern features: past tenses, third person pronouns, perfect aspects, public verbs, synthetic negations, present participial clauses.
3. Explicit versus situation dependent reference	High scores: context-independent	Situational-dependent reference features: time adverbials, place adverbials, general adverbs. Explicit reference features: wh-relative clauses on object position, pied-piping relatives, wh-relative clauses on subject position, phrasal coordinations, nominalizations.
4. Overt expression of persuasion	High scores: explicitly marks the author's point of view as well as their assessment of likelihood and/or certainty	Overt expression of persuasion features: infinitives, prediction modals, suasive verbs, conditional subordinations, necessity modals, split auxiliaries.
5. Abstract versus non-abstract information	High scores: provides information in a technical, abstract and formal way.	Abstract informational features: conjuncts, agentless passives, past participial clauses, WHIZ deletion relatives, other adverbial subordinators.
6. Online informational elaboration	High scores: informational in nature but produces under certain time constraints	Online informational elaboration features: that clauses as verb complements, demonstratives, that relative clauses on object position, that clauses as adjective complement.

After the completion of the computation of the MD analysis, the MAT provides with a set of Excel spreadsheets containing the descriptive data of the extracted linguistic features and the factor score of each of the predefined textual dimensions on the corpus provided. This computation was applied to the BLaRC, representing the genre of law reports, and to the BSLC, representing statute law.

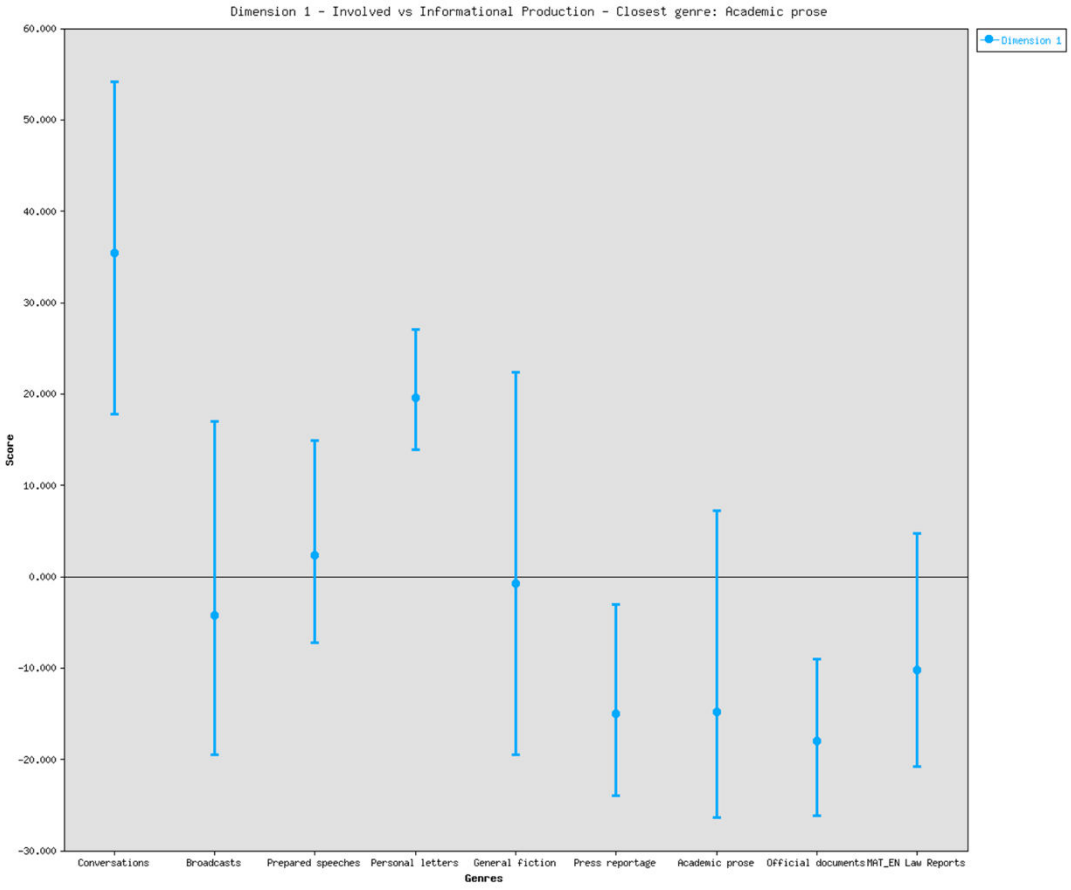
The MAT provides as well an Excel spreadsheet with the results in scaled as z-scores, and a set of plots to compare the scores of the corpus provided with the ones in the genres studied in the original MD analysis (Biber, 1988). These plots also show the most similar genre among the original MD analysis to the one provided to the software in terms of factor scores per each dimension.

5.2. Results and discussion

The automated MD analysis computed by the MAT shows that, when compared to genres from different specialised fields or communicative contexts, law reports and statute law, though different enough, they are similar in the general structures and share more characteristics between each other than they do with other genres not belonging to legal discourse, as expected. Data 6 and Data 7 provide with the descriptive statistical values of the linguistic features in our corpora.

In regard to textual dimension 1 – *'Involved vs. Informational Focus'*, the descriptive values and the plots in Figure 15 show the two genres how the two genres, though having significantly different values in the dimension, are very close to each other. Both genres skew negatively in the dimension, thus having the Informational Focus a predominance in them. They are among the genre with a lowest score in dimension, together with *academic prose*, *official documents*, and *press reportage*, very distanced from genres such as *conversations* or *personal letters.*

Regarding the differences between the two legal genres, law reports show a mean value of -10, while statute law of -16. This reinforces the idea that judges usually a wider range of persuasive, guiding and involving linguistic features than lawmakers (Section 4.3.2). In line with this reasoning, the MAT has considered that the closest genre to law reports (not considering its legal genre peer) is *academic prose*, while the closest one to statute law is *press reportage.* Similarly to law reports, *academic prose* is a genre characterised by its strong objective and highly density in information but conveyed in a persuasive and guided language that eases the reading to certain extent, while *press reportage* documents and statute law tend to provide with analytical and completely objective or categorical data.



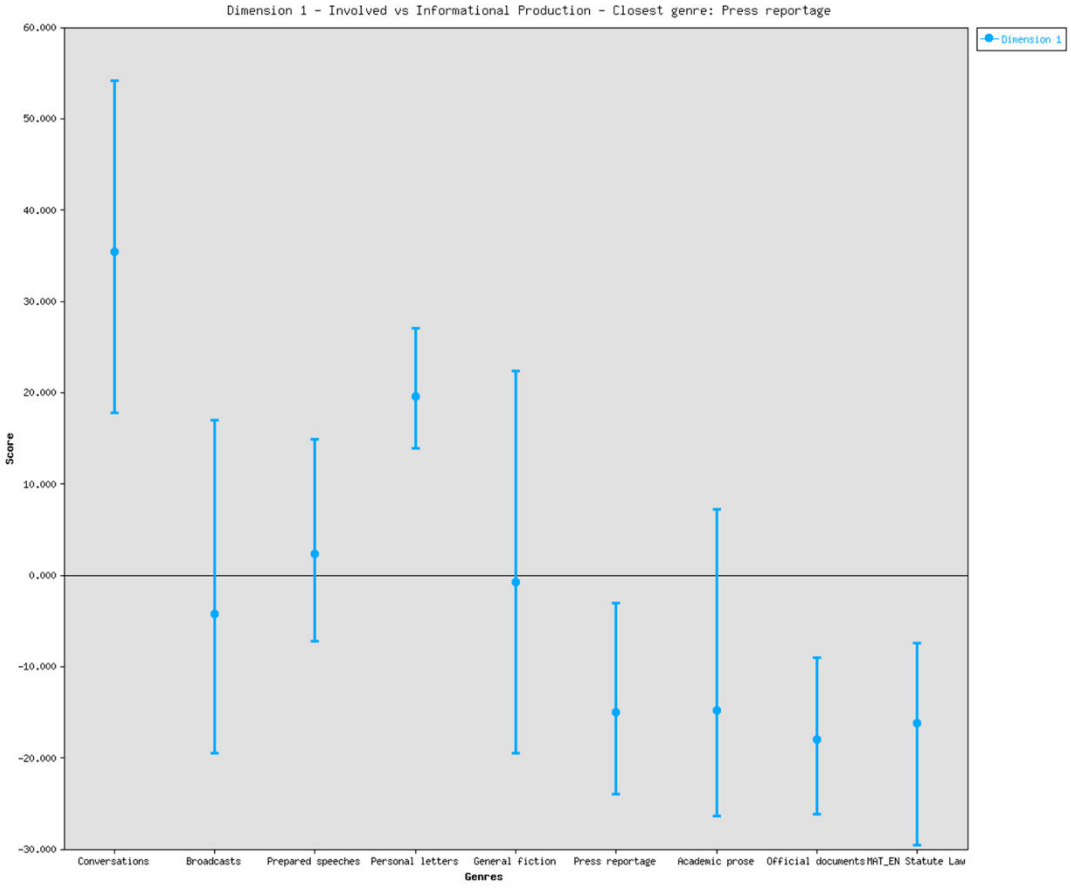


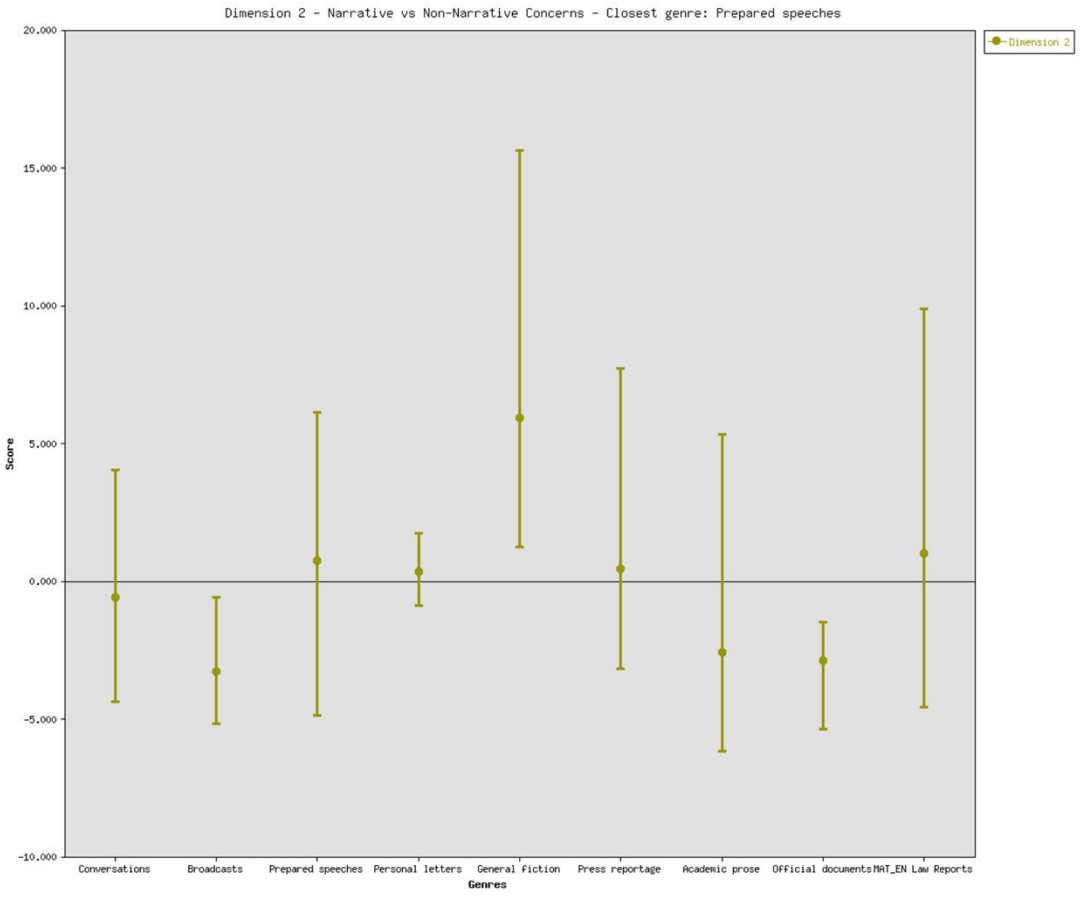
Figure 15: Dimension 1: Involved vs. Informational Focus

In Dimension 2, larger differences between the two genres are observed (Figure 16). While the mean value of law reports' factor scores is slightly positive (0.98), the one of statute law shows a negative value (-5.16). Although the range presented by the values in law reports is higher than the ones in statute law, which more clustered around the mean, this difference in the mean is still noticeable. The *Narrative Focus* side of this dimension is closely related to the positive side in the previous newly created Dimension 5 *Facts-narration*, both composed of features such as *past tense*, *third personal pronouns* or *perfect aspect*. This result confirms a narrative focus in part of the content produced in the genre of law reports, corresponding to the facts in issue or facts section in judgments, dedicated to the explanation of the previous events leading to the trial, as well as the successive intervention of courts in processes with appeals (Álvarez Álvarez, 2008).

In turn, the MAT has considered the closest genre to law reports to be *prepared speeches*, which might be explained by their shared focus in a narration of successive past events (speeches usually attempt to convey a message introducing historical or recent events for society).

Contrarily, statute law is skewed towards the *Non-Narrative Focus of the dimension.* The fact that statute law is more focused on the explanation of the functioning or situations in which the law must be enforced (using a high variety of *nouns*), with a scarce attention to changes produced in time (using prominently *present tenses).*

In this case, the MAT has selected *broadcasts* as the closest genre to statute law. This relationship might be counterintuitive, but this is plausible given both focus on present or atemporal discussions, rather than focusing on past events or the change of the nature of things with time (Biber, 1988).



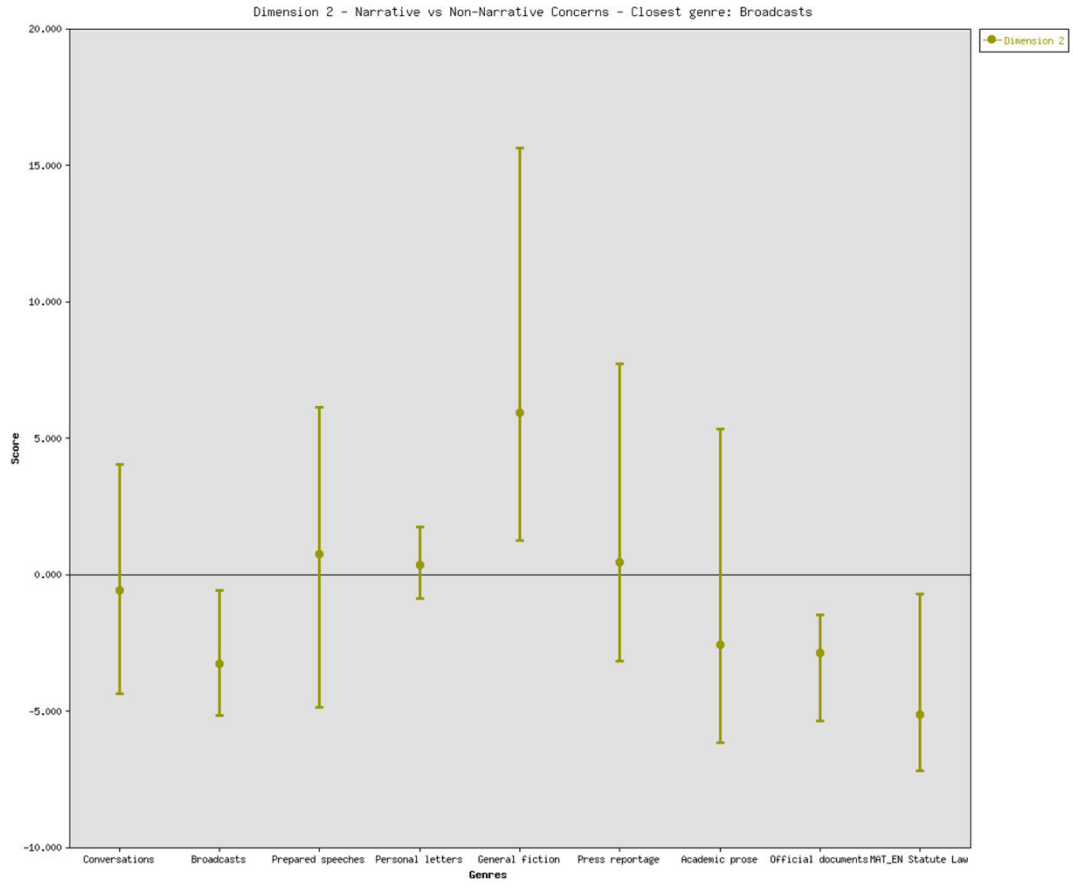


Figure 16: Dimension 2: Narrative vs. Non-Narrative Focus

Dimension 3 considers the extent to what a text makes explicit the contextual information around the core object in issue. In other words, whether it uses strategies or linguistic resources to clarify, repeat or explain the situation in which the statement is uttered (for example, *phrasal coordination*, *nominalisations*, or *wh- relative clauses*). As far as British public legal genres are concerned, the use of this resources is considerable, given the mean values for their factor scores (5.75 and 11.24), but not equal, almost doubling statute law mean value the law reports value. This can be observed in Figure 17.

This significant difference goes in accordance with the highly explicit, explanatory nature of legal provisions, which are to be object of interpretation by professionals of law. The attempt by lawmakers not to leave any aspect of the provision free to wrong or inaccurate interpretation makes them include extremely long sentences including every possible detail, hypothetical situations or potential subjects to the law (Alcaraz, 2007; Sun & Cheng, 2017). This consequently increases considerably its score in this dimension. This aspect is common not only in legislation but in a vast majority of *official documents*, and the MAT therefore considered them the closest genre to statute law.

Law reports, in turn, also present a positive score in the dimension, what makes them an explicit reference genre, but in a lower degree. They include these explanations as a help or guidance for the readers, rather than as an obligation or defence against misleading interpretations. This aspect in discourse is also usual in *academic prose* (Biber, 2006; Hyland, 2005), therefore interpreted by the MAT as the closest genre to law reports.

On the contrary, other familiar or oral genres such as broadcasts, personal letters or conversations are distantly positioned from the British public legal genres.

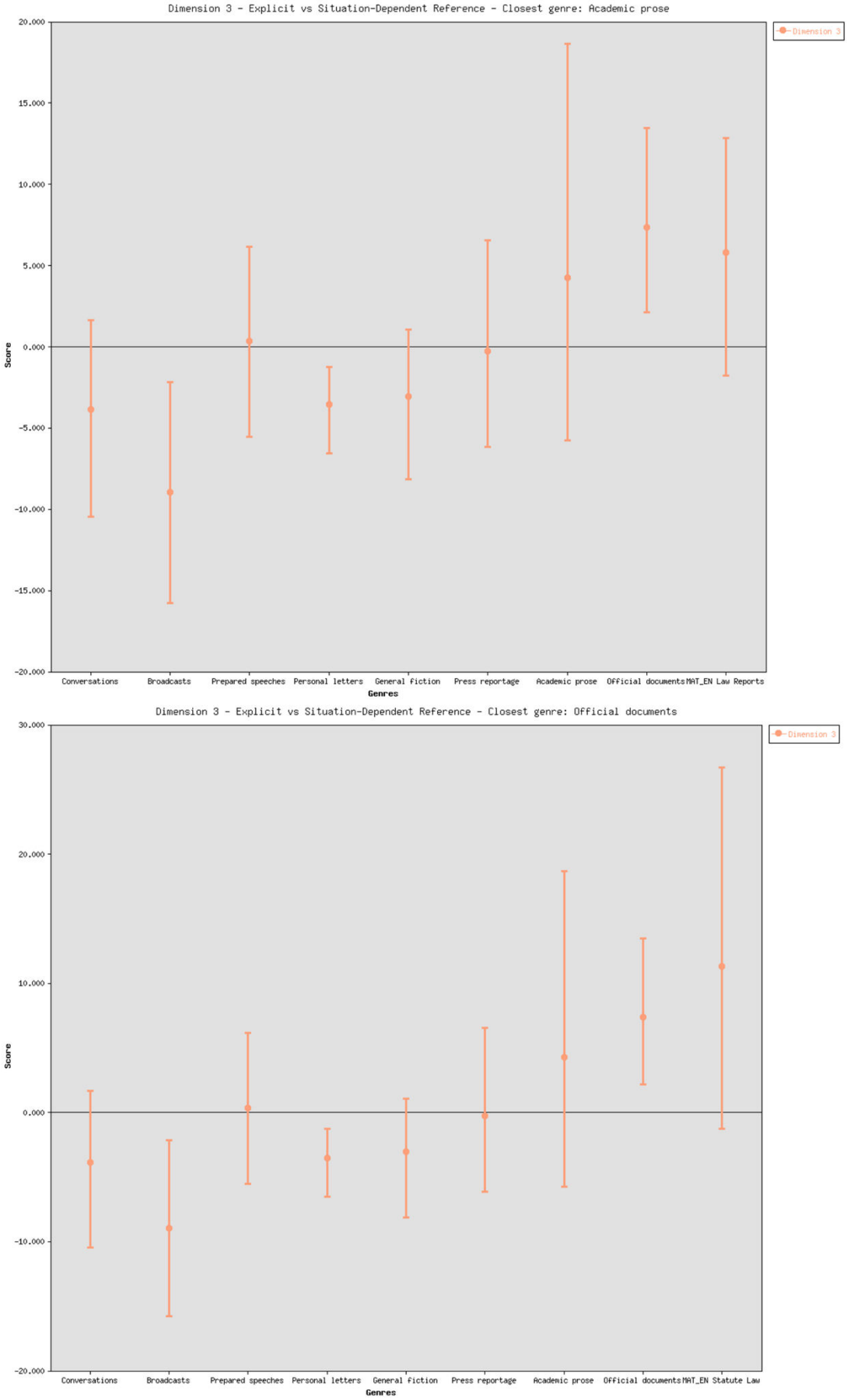


Figure 17: Dimension 3: Explicit vs. Situation Dependent Reference

Dimension 4 shows how evident the resources or strategies used by the utterer to increase persuasion on the discourse are, explicitly marking the author's point of view or their assessment of likelihood or certainty. Figure 18 shows a general high range of scores among every genre, including British legal public genres. These two have a mean value of 3.67 and -2.25, supporting the idea of law reports being more involving, persuasive and closer to the reader documents than statute law, stated in Section 4.3.2.

Having a negative score, statute law clusters with genres such as *broadcasts*, *academic prose* or *official documents*, being the closest one *press reportages*. As a narrative, objective, unguiding and formal written discourse, it is coherent in terms of discursive features that it remains also a non-persuasive discourse.

In turn, law reports, although being a highly specialized, formal genre, remains among the genre with highest scores in the dimension (the most overt-persuasive ones), such as general fiction or prepared speeches, but especially *personal letters*, which is the closest one. This might be found surprising, but it goes in accordance with the reasoning made in 4.3.2, claiming that judicial decisions are characterised by the elaborate orality produced by judges when they answer or argue against other judges through their reasonings in the judicial decisions.

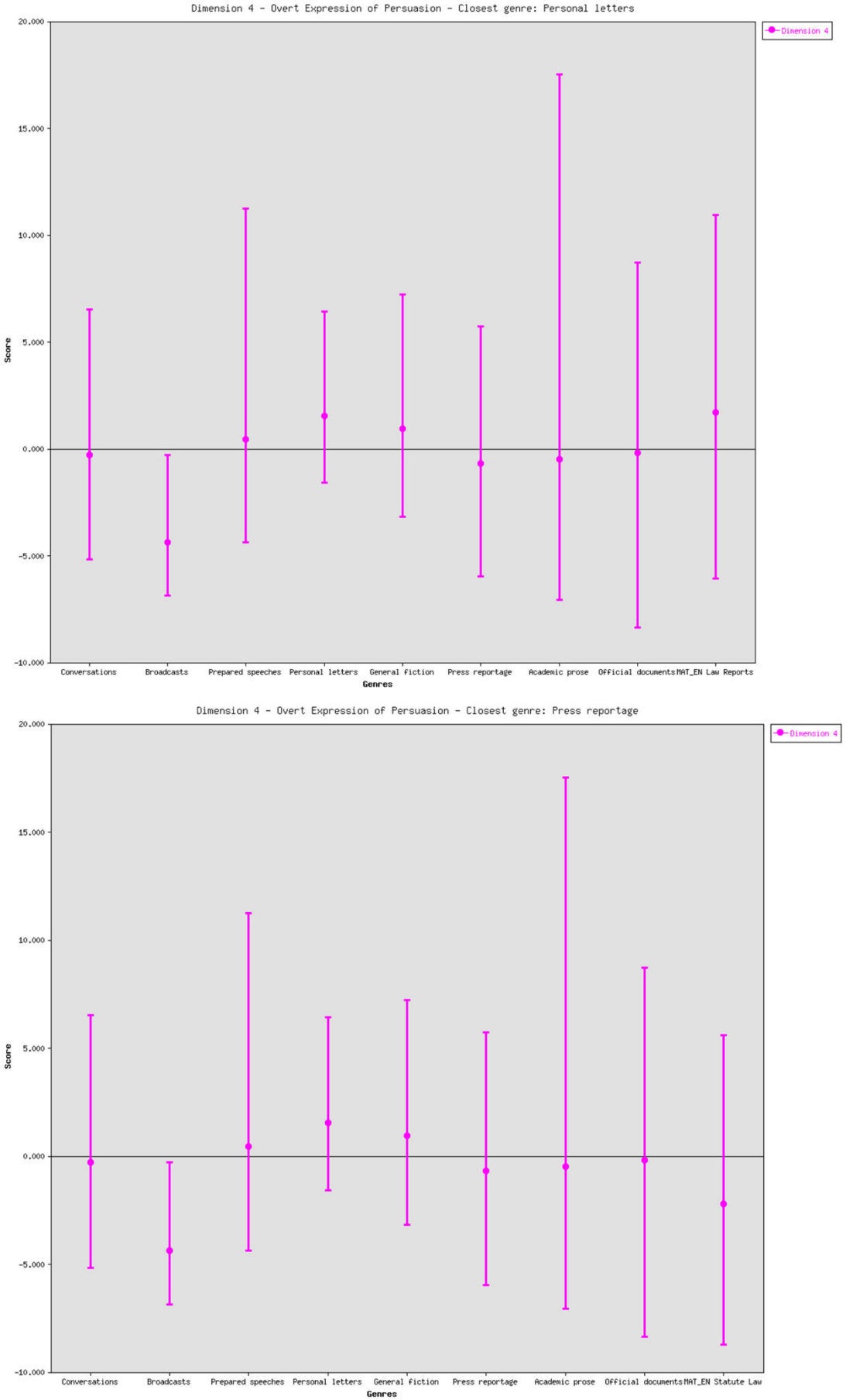


Figure 18: Dimension 4: Overt Expression of Persuasion

Dimension 5 assesses the extent to which a genre provides information in a technical and formal way, focusing on complex and abstract issues. Legal Discourse, though being highly specialised and dealing with complex reasonings on legal concepts and proceedings, is a field of expertise strongly related to real daily life situations, as that is the object of regulation. The complex and abstracts ideas in legal conceptualisations are not more than an abstraction of daily life transactions, conflicts or processes, such as marriage, purchases, rental, private disputes or labour (Alcaraz et al., 2014; Alcaraz & Hughes, 2015; Piszcz & Sierocka, 2020).

As a consequence, these genres include a combination of highly abstract legal reasonings with material narration or description of events, subjects or facts. Thus, both genres have a mean value close to 0, 3.67 and 1.07, remaining rather neutral. Nonetheless, these values show a slight predominance of abstract information in law reports, what is explained by the relevant section including legal reasonings that will be adopted as binding legal precedents by inferior courts, the so-called *ratio decidendi* (Section 1.1). This degree of abstract information contained similar to the one in *academic prose* or *press reportage*, being the closest similar genre, though, *official documents*.

In regard to *statute law*, even if stays in the positive side, the value is so low that it is noticed in Figure 19 how this genre's most similar type of discourse is the one of *press reportage*, also on the lower range of the positive side.

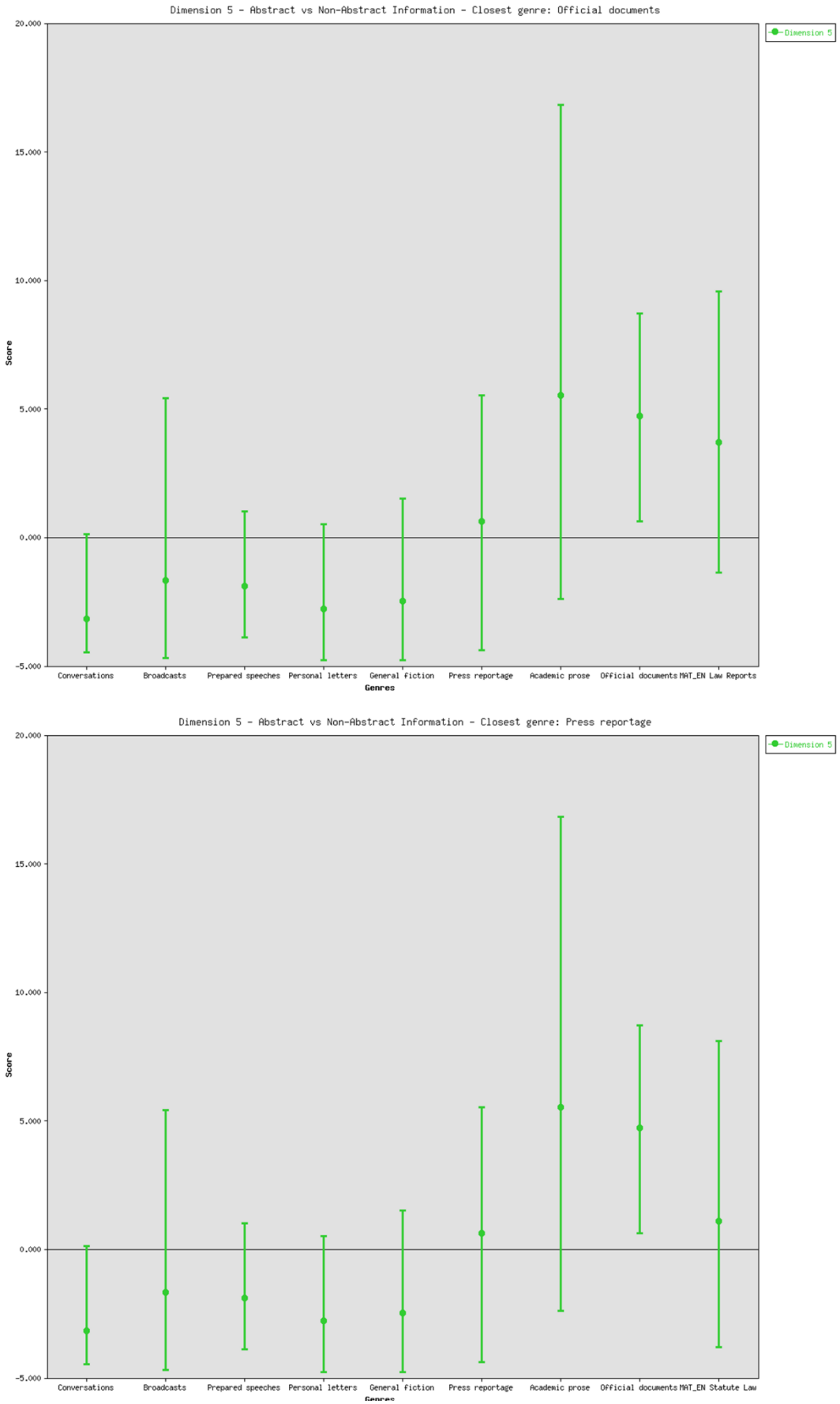


Figure 19: Dimension 5: Abstract vs. Non-Abstract Information

Dimension 6 was not included in the original Biber's MD analysis (Biber, 1988), since it considers a relatively new phenomenon related to the need of immediacy when reporting news or other pieces of information on social networks or digital press. This dimension is presumed to capture a discourse with an essential focus on the provision of information, but, differently from traditional press informative genres, with a general use of strategies, expressions or language resources related to oral discourse rather than informative written genres, such as the use of *THAT clauses as verb complements*, *first person pronouns*, *demonstratives* or *THAT relative clauses in object position*. These are featured in expressions such as *I think that…* or *it is important that…*

Figure 20 shows that both genres have low mean values in this dimension, which is the expected value specialised genres are to have. Nonetheless, *law reports* (even if low) has a positive value (2.25), whereas statute law remains in the negative side of the dimension (-0.26). The plot shows how law reports is positioned on the higher range of the positive side highly communicative but restrictive in time or space genres such as *prepared speeches*. Thus, the lower the time restrictions or informational focus of the genre, the lower the scores: in the range between 1 and -1 values, genres such as *academic prose*, *conversations*, or *statute law* are found. These genres are balanced in the information and time restrictions around the. Lastly, *broadcasts, personal letters, general fiction,* and *official documents* are on the lowest scores, but for different reasons: *broadcasts* and *official documents* are highly informative, but their restrictions are probably very few or none, while *personal letters* and *general fiction* are not restricted at all in the space available, but they are the least informative of the group of genres.

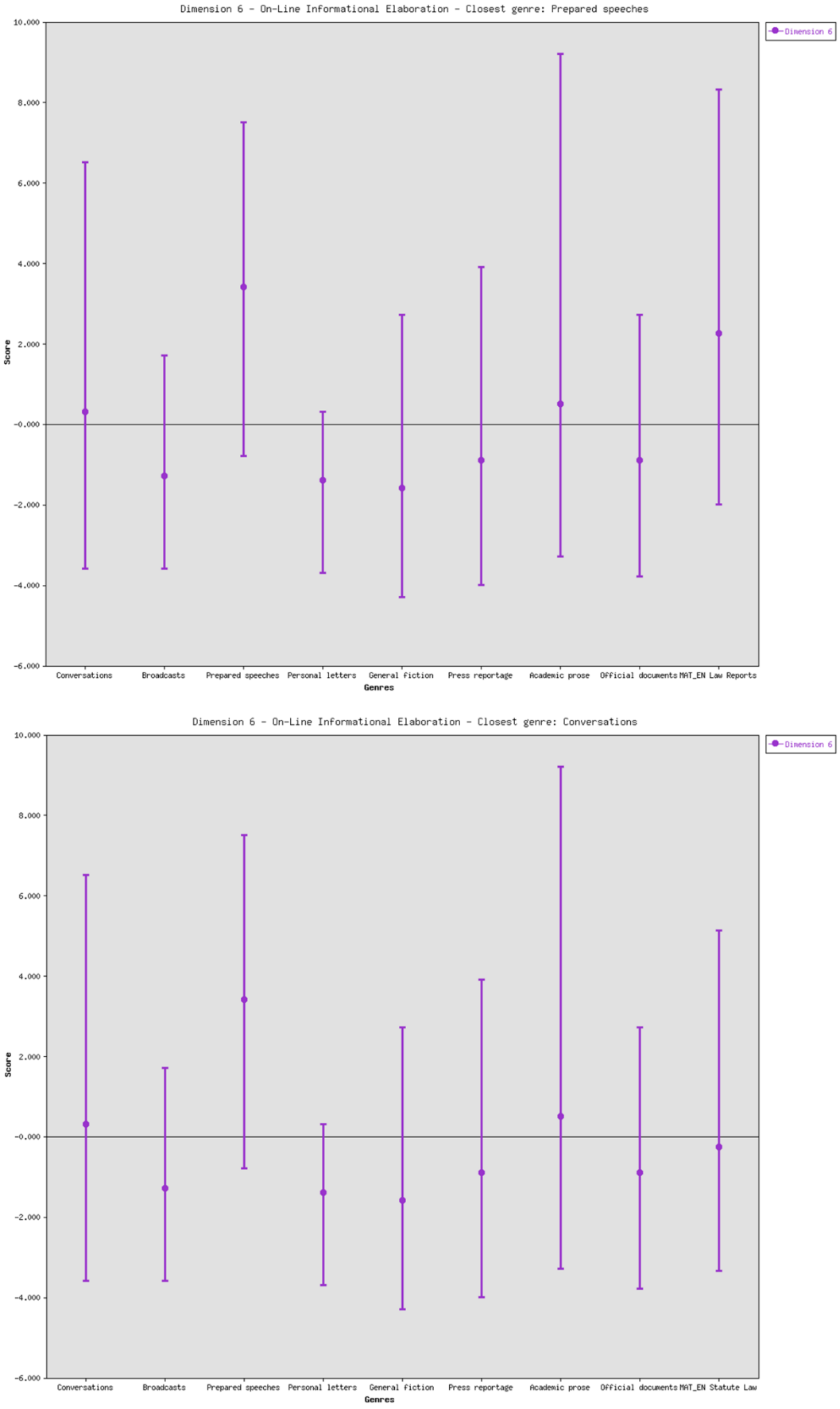


Figure 20: Dimension 6: Online Informational Elaboration

5.3. Conclusions

The MD analysis provided by the Multidimensional Analysis Tagger (Nini, 2019) provides with a complementary broader insight into public legal genres to the one obtain by the *ad hoc* MD analysis in Chapter 4. The majority of conclusions made in the discussion of results in Section 4.3.2 were reinforced with this analysis, after obtaining information about how British Legal Genres are positioned in a common set of textual dimensions for a broader selection of genres. Moreover, this analysis has also provided with the chances of determining the most similar genres outside the legal domain to law reports and statute law.

The combination of these analyses provides with powerful framework to understand the linguistic resources and strategies judges and lawmakers use to achieve their purposes when writing the documents belonging to these genres and the quantitative nature of the analysis make interested researchers able to discern what factors and (more precisely) what variables are most likely to affect the differentiation of British public legal genres. In fact, potential variables to nurture legal regressions for the development of automation processes in the detection or creation of legal decisions and legal provisions are provided. In Figure 21 a summary of the results is provided.

Statute Law	Informational Focus *(Nouns, long words)*
	Non-Narrative Focus *(present tense, attrib. adj.)*
	Explicit Reference *(wh- rel. clauses as S and O, pied piping)*
	Scarce Overt Expression of Persuasion *(Infinitives, prediction modals, suasive verbs)*
	Abstract Information / Style *(conjuncts, agentless passives)*
	Online Informational Elaboration *(that clauses as verb complements)*
Law Reports	(Less) Informational Focus *(Nouns, long words)*
	Narrative Focus *(past tense, 3rd person pron.)*
	(Less) Explicit Reference *(wh- rel. clauses as S and O, pied piping)*
	Overt Expression of Persuasion *(Infinitives, prediction modals, suasive verbs)*
	Abstract Information / Style *(conjuncts, agentless passives)*
	(More) Online Informational Elaboration *(that clauses as verb complements)*

Figure 21: MD analysis computed by the Multidimensional Analysis Tagger (summary)

6. HOW TO SPOT ARGUMENT SCHEMES IN LEGAL DISCOURSE: A CORPUS-DRIVEN STUDY

The exploration, description and analysis of argumentation has been of the concern of philosophers, logicians, computer scientists and linguists. During the 20th century, in an attempt to develop new ways of analysing arguments from a more empirical and informal approach, authors such as Toulmin (1958)), Perelman and Olbrechts-Tyteca (1969) or Walton (2008), developed argument structure models and taxonomies helping ulterior studies further delimitate the types of arguments we find in daily life conversations.

These models are based in the inference relations structuring an argument (being the most basic one of a conclusion drawing on one or several premises that support that conclusion defended by the arguer). Following those taxonomies, mainly the argument schemes proposed by Walton, the prominent studies in Artificial Intelligence have gained interest in Argumentation Theory, with the aim of developing new tools that help the user detect and classify arguments found in any type of speech, such as the OVA+ (Janier et al., 2014; Lawrence et al., 2019).

In turn, linguistics authors have studied so far argumentation merely as one type of discourse, rather than a central aspect to be analysed. Thus, we can find many studies concerning the pragmatic and morphosyntactic features of argumentative discourses in many studies from the approach of genre analysis or discourse analysis (Bhatia, 2014; Biber, 1988; Swales, 1990). Metadiscourse (Hyland, 2005) might be one of the linguistic approaches that has given the most importance to argumentation within the field, describing the so called 'metadiscourse' resources used in argumentative writing to make the text more persuasive and readable for the recipient. He specifically explained metadiscourse in the context of rhetoric, showing examples of the uses of these metadiscourse resources depending on its function. They might be used to appeal the ethos, the pathos, or the logos (that is, the utterer's credibility, emotions and reason) (Hyland,

2005, pp. 65–86). These Hyland's metadiscourse resources might find some common ground to the argumentative indicators explored from the Pragma-dialectics approach (Eemeren et al., 2007) and some scholars from the University of Lugano (Musi & Rocci, 2017). The study of the variation, structure, and features of discourse (and language in general) has been boosted by the development in recent years of corpus linguistics and NLP tools annotating increasingly larger amounts of words accurately with POS, semantic labelling, or parsing.

Argumentation mining emerged with the purpose of automatically detecting, classifying and structuring argumentation in text (Mochales & Moens, 2011) combining the machine learning methods available with the argumentation taxonomies and models proposed so far. Nevertheless, as we will see in later in this paper, these methods still have many difficulties to distinguish arguments from other types of utterances, and to distinguish the different type of arguments there are in a text. For all these reasons, this paper aims at exploring argument indicators that may be used for future argumentation mining models training by finding out whether there are some significant correlations between, metadiscourse resources and morphosyntactic features related to persuasion, and the use of different type of argument schemes in legal texts. To do that, in this paper we will make use of five judgments of the British Law Report Corpus (BLaRC) (Marín & Rea Rizzo, 2012) as legal argumentation is usually seen as the most structured and easy to analyse and see patters in it, which will be parsed by using the NLTK and manually annotated with the legal argument schemes proposed by Walton (2010) following the annotation guidelines presented later.

6.1. Materials and methods

6.1.1. Corpus

As aforementioned, legal discourse is on the one hand one of the most relevant types of argumentation due to its complex and well organised structure, allowing an easier identification of arguments and argumentation schemes, and on the other hand a very different discourse from any other because of its lexical, syntactical and pragmatical peculiarities. We decided therefore to use a legal corpus to undertake this explorative research aiming at finding new ways of spotting argument schemes, as in legal texts a possible correlation between Biber's morphosyntactic features and Hyland's metadiscourse resources with Walton's legal argumentation schemes will be clearer and easier to identify.

We used as a corpus 4 judgments from the British Law Report Corpus (Marín and Camino 2012) which is made up of judicial decisions issued by British courts and tribunals. Our 4 judgment-corpus consists of 36 827 words (Table 21).

Table 21: Types and Tokens in the BLaRC sample for annotation with argument schemes

Judgments	**Tokens**	**Types**
Text 1	6773	1329
Text 2	9774	1677
Text 3	12819	1994
Text 4	7461	1351

This corpus was parsed with NLTK (Bird et al., 2009) to get the values and frequency of 30 morphosyntactic features (Table 22) selected from the ones proposed by Biber & Conrad (2019).

Table 22: Linguistic features selected for the annotation with NLTK

Lexical density	***Subordination/Coordination***	***Statements***
Average word length (AWL)	Cause-Effect (C.E)	Declarative (DECLA)
Average clause length (ACL)	Concession (CONC)	Interrogative (INT)
Type-token ratio (TTR)	Comparison (COMP)	Imperative (IMP)
Content word classes	***Tense***	***Demonstrative***
Nouns (N)	Present tense (PT)	Demonstrative pronouns (DEM)
Verbs (V)	Past tense (PST)	***Person***
Adjectives (ADJ)	***Aspect***	First singular (FS)
Adverbs (ADV)	Simple (SIMP)	Second (SECOND)
Modal classes	Progressive (PROG)	Third singular (TS)
Possibility (POSS)	Perfect (PERF)	First plural (FP)
Necessity (NEC)	***Voice***	Third plural (TP)
Predictive (PRED)	Active (ACT)	***Prepositional***
	Passive (PAS)	Prepositional phrases (PREP)

NLTK or Natural Language Toolkit is a free-to-use Python library that offers a wide number of interfaces to work with human language data. Amongst others, it provides methods for classification, tokenization, stemming, tagging, parsing and semantic reasoning of textual data. A brief description of how we used this library to create the statistics for the morphosyntactic features of the table above is given below:

We first applied a tokenization function to split the text into tokens and then, we further tagged these tokens with Part of Speech (POS) labels. This allowed us to identify the nouns, verbs, adjectives, adverbs, and the modals in general. To discriminate then between the modal classes of possibility, necessity, and prediction we looked for particular indicators for each category. More precisely, indicators like ‘can’, ‘could’, ‘might’, ‘may’ were used to identify the modality of possibility; indicators like ‘must’, ‘should’, ‘ought to’ were used to identify necessity; and indicators like ‘will’, ‘would’ or ‘shall’ were used for the identification of prediction.

For the cause-to-effect relations, we searched for keywords like ‘because’, ‘due to’, ‘consequently’, ‘therefore’ etc, keywords that imply some justification

between two propositions. Concession is identified by keywords like 'if', 'though', 'although' etc. For comparison relations, we used NLTK tags that recognise comparative (larger) or superlative (largest) adjectives. There are also specific NLTK tags about the verbs in present and past tense. For the perfect aspect, we used tags that indicate past participle of verbs and checked which of them are preceded by 'have' or 'has'. Active voice is indicated by the verbs which are in their simple form, as gerunds etc, while for the passive voice, we looked for tags that indicate past participle of verbs, and then checked which of them were preceded by 'is', 'are', 'was', etc.

Demonstrative pronouns were identified by determiner tags of NLTK. For the different persons, we searched for keywords like 'I', 'you', 'he' etc., followed by some verb. Prepositional phrases were also recognised by specific tags.

Finally, after we split text into propositional entities, we identified the interrogative ones as those ended up in a question mark, imperatives were those beginning with some verb and the rest were the declarative ones.

We also obtained the frequency of the metadiscourse resources used in our corpus. The metadiscourse resources are divided into interactive and interactional resources, depending on their way of fulfilling the interpersonal purpose metadiscourse has, as explained in Section 1.2.3.These resources are the following (Hyland, 2005, pp. 50–54).

A. Interactive resources

a. *Transition markers*: conjunctions and adverbial markers. They help the reader interpret pragmatic connections between steps in the argument.

b. *Frame markers*: they signal text boundaries or elements of schematic text structure.

c. *Endophoric markers*: expressions which refer to other parts of the text, often facilitating comprehension and supporting arguments by referring to earlier material or anticipating something yet to come.

d. *Evidentials*: they guide the reader's interpretation and establish an authorial command of the subject (usually represented by references and literature on the subject matter).

e. *Code glosses*: they supply additional information by rephrasing, explaining, or elaborating what has been previously said, to ensure the recipient is able to recover the writer's intended meaning.

B. Interactional resources:

a. *Hedges*: they indicate the writer's decision to recognise alternative viewpoints or approaches and so withhold complete commitment to a proposition, emphasising the subjectivity of a position by presenting information as an opinion rather than a fact.

b. *Boosters*: they are the opposite to hedges, since they allow writers to close down alternatives and express their certainty in what they say.

c. *Attitude markers*: they indicate the writer's affective, rather than epistemic, attitude to propositions, so they convey surprise, agreement, frustration, or importance rather than relevance, truth or reliability.

d. *Self-mention*: this refers to the degree of explicit author presence in the text measured by the frequency of first-person pronouns or possessive adjectives.

e. *Engagement markers*: these devices explicitly address readers, with the aim of either focusing their attention or including them as discourse participants.

6.1.2. Legal argument schemes annotation guidelines and process

The corpus was also annotated with argumentation schemes. To do so, a new set of annotation guidelines for legal argumentation schemes was developed. These guidelines were based on the classification of argumentation schemes by Walton and Macagno (2015), the legal argumentation schemes proposed by Walton (2010), and the annotation guidelines proposed by Lawrence et al. (2020). The guidelines are structured as a key that guides the annotator through the fulfilment or lack of it of simple statements regarding the argument scheme to annotate. Firstly, the guidelines help the annotator choose between three big groups of schemes: source-dependent argument, source-independent arguments, and reasoning arguments. Secondly, the annotator needs to ask themselves whether their argument fits the statement presented or not and choose the scheme according to that. The key to the guidelines is shown in Supplementary document 1.

This process started with reading through the Legal Schemes-specific guidelines to ensure it was annotator appropriate, whilst separating the whole texts into smaller and manageable parts. Once completed, this was shared on platforms accessible to analysts and the annotation process started with gathering a team of analysts, whom we took from our existing pool of annotators. This group then had a debrief and short training session where we read the guidelines thoroughly and collectively went through a couple parts. A spreadsheet and two Slack channels were created, in order to keep track of work completed and give access to the work, respectively. One of these Slack channels was exclusively for encouraging a discussion of the work.

This work was tackled in two similar methods. Firstly, it was made available to all analysts to work on in their own time and secondly, it was annotated in small groups in meeting-specific times. It was close to an even split in these methods, albeit slanted towards the second method: 52 parts completed on analysts own time, and 65 completed during annotation sessions. Both methods had the same ways of completing this work. In both, analysts took parts to complete, which they

then had reviewed by another analyst. This led to a discussion until both analysts are happy with the final analysis, which was then uploaded to the AIFdb (https://www.aifdb.org/search) server. These parts were grouped in a corpus, Legal Schemes Project, in order for us to have ease of access to annotated parts.

The inter-annotator agreement (IAA) was performed to establish the reliability of the Legal Annotation Guidelines as a guide to manually annotate legal argument schemes (Artstein, 2017). This included randomly choosing 10% of the completed corpus and having these parts reannotated. This was done by the same group who did the initial analysis due to their familiarity and training in the data. We ensured that the previous annotators of these parts were different from those doing the reannotation and this work was completed in annotation sessions.

To find out whether there is a possibility for morphosyntactic features and metadiscourse features to be correlated with the appearance or one type of legal argumentation scheme or another, we performed Spearman's correlation with every possible combination of these as independent variables with the legal schemes as dependent variables. Only having 4 observations in our datasets made us unable know if our dataset had a normal distribution, and we could therefore not use Pearson's correlation. Moreover, Spearman's correlation tests with only 4 observation will only give us a clue of the relationship between our variables, but due to the low number of observations the p-value cannot be lower than 0.05 and thus, we will not be able to claim there is any significant correlation between our variables, but at least we can see from an explorative insight if there are any interesting patterns to verify in future studies.

6.2. Results

6.2.1. Inter-annotator agreement test for the evaluation of the guidelines

The absolute and relative frequency of the legal argument schemes found in our corpus by an annotation team hired by the research group following the Annotation Guidelines (Section 6.1.2) are in Table 23. The annotated corpus is available on the online repository aifdb.org with the name 'Legal Schemes 1-4'.

Table 23: Legal Argument Schemes encountered after annotation

Argument Scheme	Judgment 1	Judgment 2	Judgment 3	Judgment 4	Total	%
Default Inference	25	42	73	40	**180**	35,64
Established Rule	18	36	22	12	**88**	17,43
Verbal Classification	8	19	22	25	**74**	14,65
Practical Reasoning	9	19	29	3	**60**	11,88
Analogy	8	25	12	5	**50**	9,90
Position to know	5	6	22	10	**43**	8,51
Commitment	8	8	15	0	**31**	6,14
Full Slippery Slope	4	5	14	1	**24**	4,75
Example	1	7	9	2	**19**	3,76
Sign	2	1	10	4	**17**	3,37
Generic Ad Hominem	3	4	4	0	**11**	2,18

The results of the Inter-annotator agreement reliability test are shown in Table 4. The Kappa value shows that there is a bad agreement between annotators 1 and 2, while z-value and p-value indicate whether the agreement or disagreement between annotators is significant. In our case, the disagreement is not significant as the z-value is above -1.96 while the p-value is above 0.05. In Section 6.3. the possible reasons for this poor agreement between annotators are explained.

Table 24: Cohen's Kappa results

Kappa	**-0.0647**
Z-value	-1.30
p-value	0.191

To better understand these results, we performed a confusion matrix which might show a pattern in the way the annotation fails to agree (Figure 1). The following conclusions can be extracted from the matrix

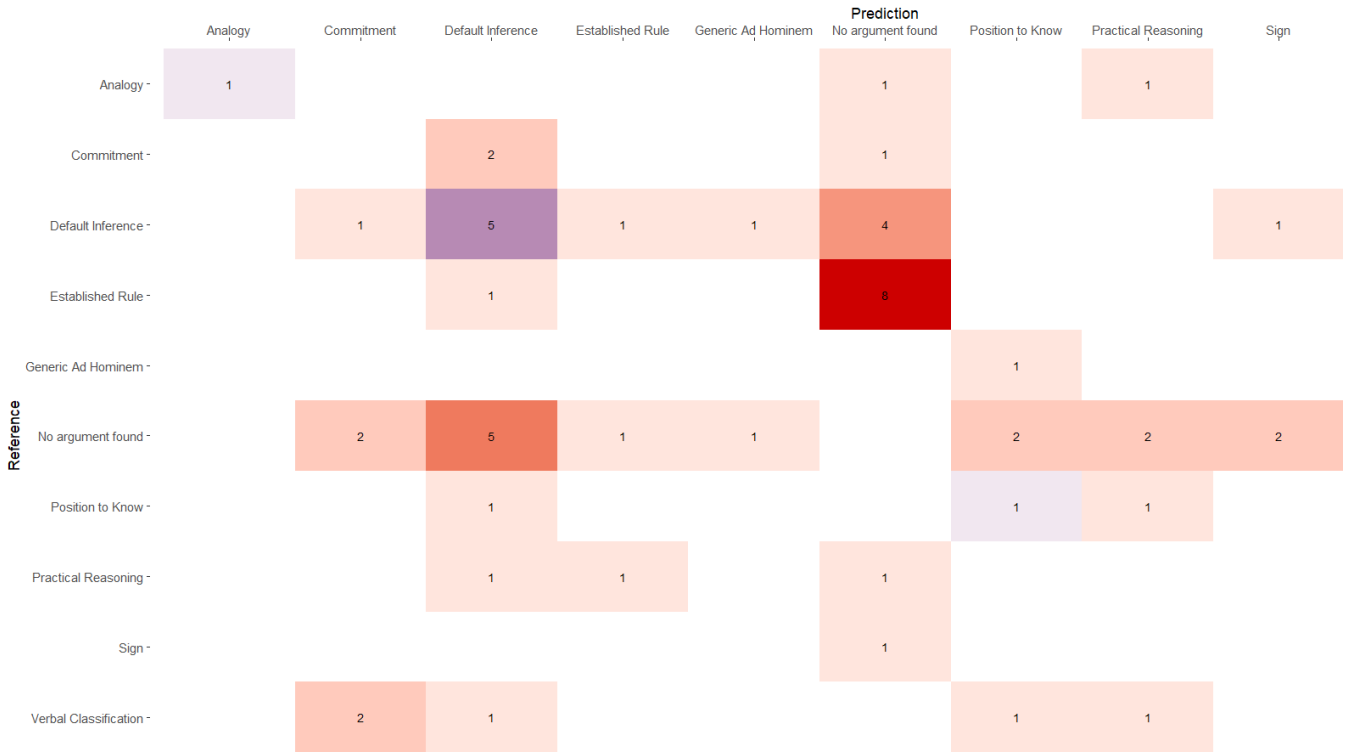


Figure 22: Confusion matrix on inter-annotator agreement

- The schemes annotated as 'Established Rule' are frequently not found in the second round.
- Verbal Classification is confused with several other schemes (Commitment, Position to know and Practical Reasoning).
- Only Analogy and Position to know have been annotated in the same way in the two rounds (but only on one occasion).
- Annotators in the first and the second round have annotated many schemes as ‘Default Inference’, not being able to classify the scheme in one of the categories provided by the guidelines.

6.2.2. Spearman's correlation tests results

In the following pages you can find Figure 23, Figure 24 and Figure 25, where scatterplots show the most relevant correlations between morphosyntactic features or metadiscourse resources and legal argumentation schemes. In Table Table 25 significant correlations obtained are shown in descending order according to the p value. (From most significant to less significant correlations)

This provides us with an idea of some patterns that might be present in the language used in argumentation schemes in legal discourse. According to our results, these correlation patterns might be consistent in legal argumentation:

(a) Verbal Classification argumentation schemes are positively correlated with Third Plural (Plot N) and Transition Markers (Plot O)

(b) Analogy argumentation schemes are negatively correlated with Declarative statements (Plot E), Progressive Aspect (Plot J) and the Type-Token Ratio (TTR) (Plot P)

(c) Position to Know argumentation schemes are positively correlated with TTR (Plot P), Demonstrative pronouns (Plot F), Declarative statements (Plot E), Active Voice (Plot A), Simple Aspect (Plot M), and Adjectives (Plot B)

(d) Commitment argumentation schemes are negatively correlated with Necessity modals (Plot G), Boosters (Plot C) and Endophoric Markers (Plot Q)

(e) Full Slippery Slope argumentation schemes are positively correlated with Predictive modals (Plot I), Concessive Subordination (Plot D), and negatively correlated with Endophoric Markers (Plot Q)

(f) Example argumentation schemes are positively correlated with Past Tenses (Plot K)

(g) Generic Ad Hominem argumentation schemes are negatively correlated with Boosters (Plot C), Necessity modals (Plot G), Transition Markers (Plot O) and Possessive verbs (Plot H).

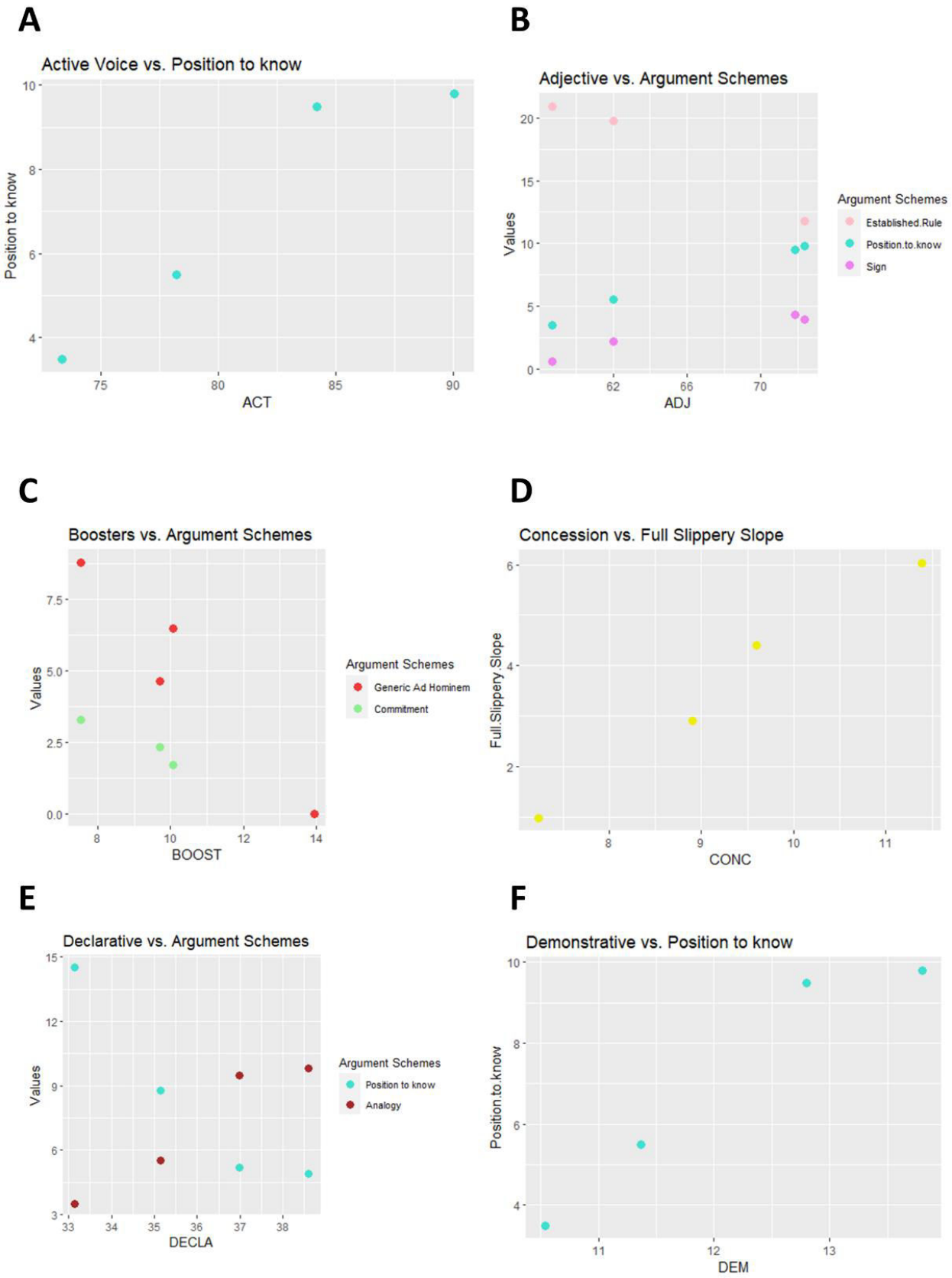


Figure 23: Correlation tests on argument schemes I

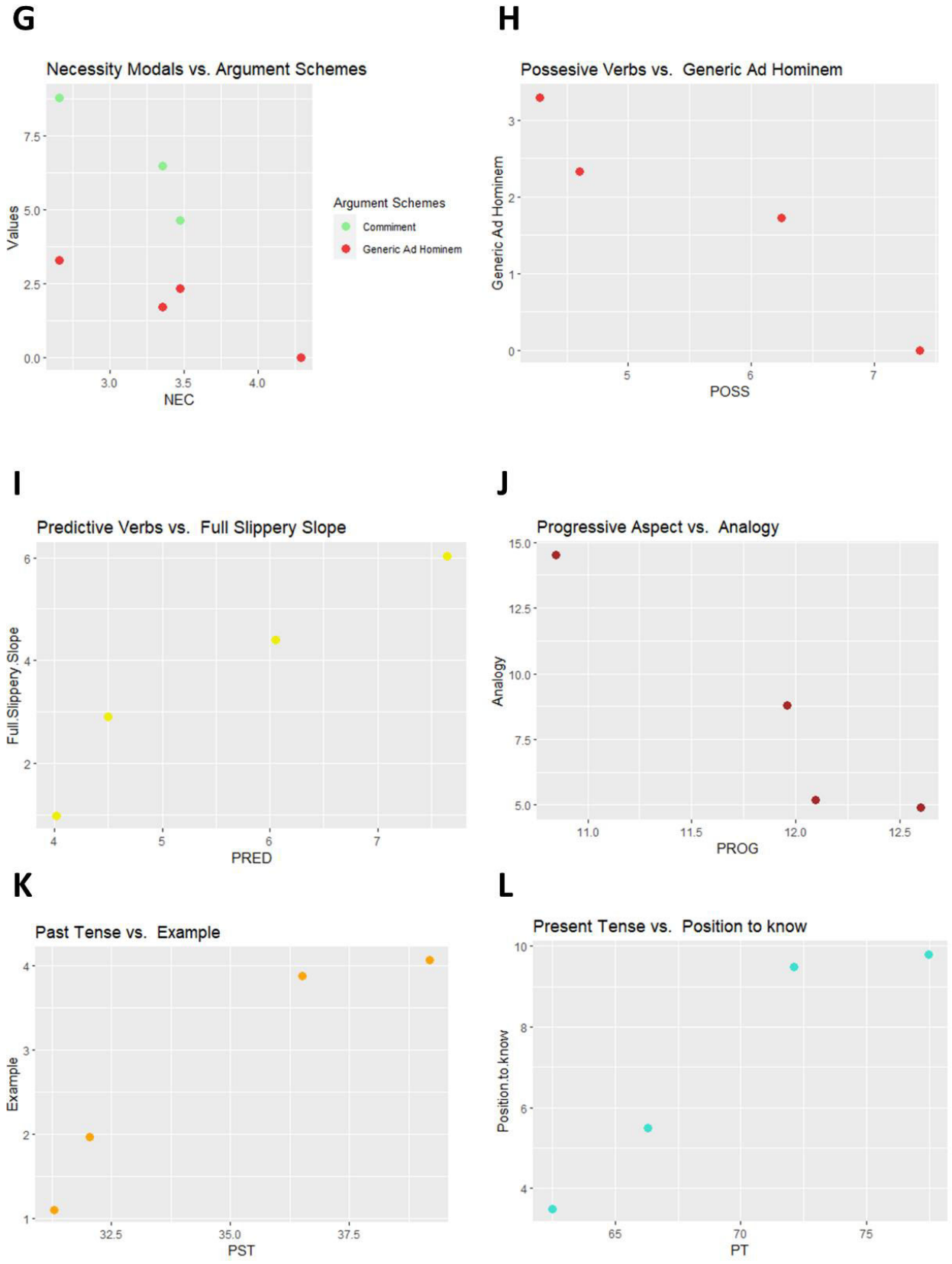


Figure 24: Correlation tests on argument schemes II

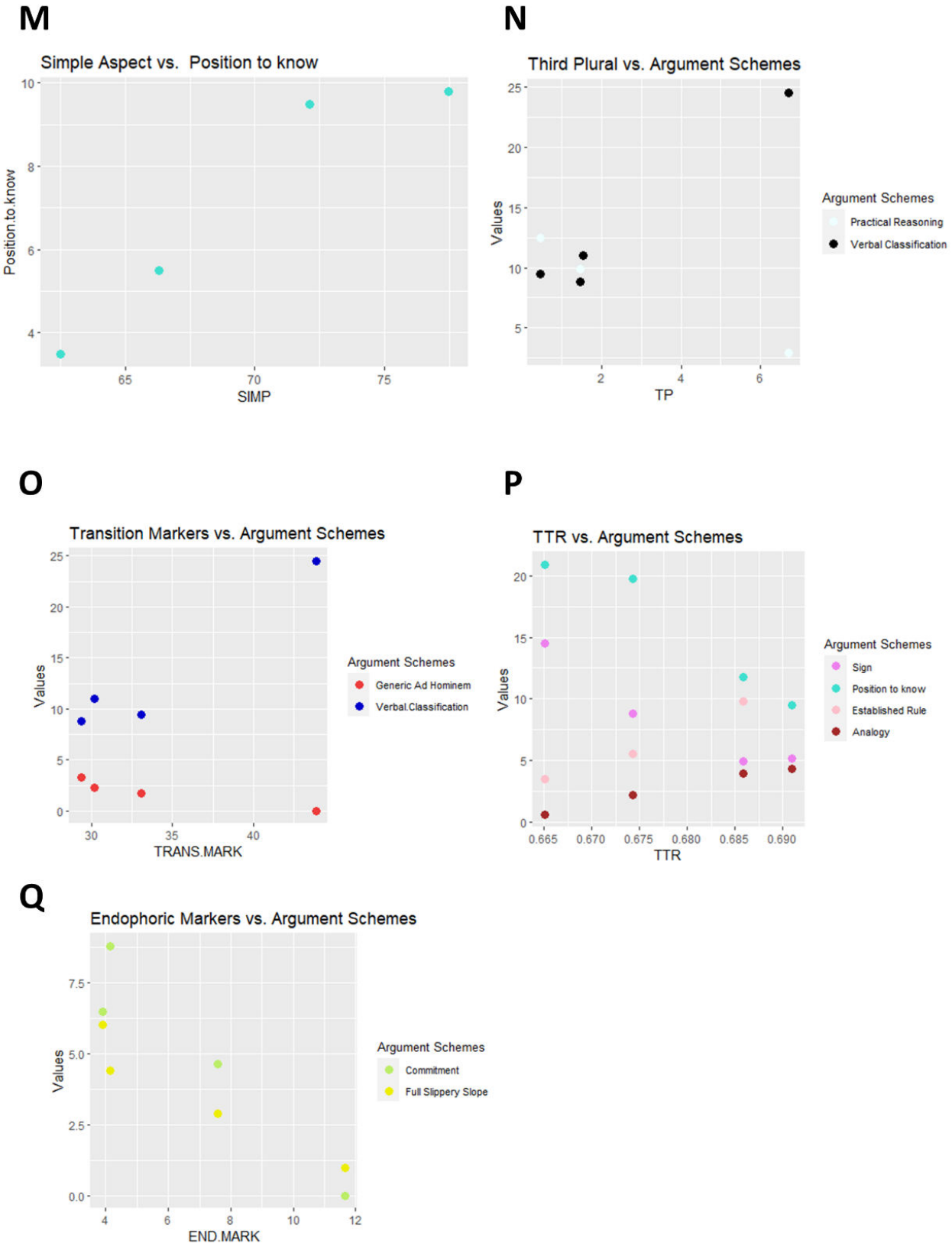


Figure 25: Correlation tests on argument schemes III

Table 25: Significant Pearson's correlations

Legal Arg. Schemes	MPS / Metadisc.	Correlation (r)	Significance (p)
Analogy	PROG	-0.9566554946	0.043344505
Analogy	TTR	-0.9560407305	0.043959270
Analogy	DECLA	-0.9505576342	0.049442366
Commitment	NEC	-0.9832357119	0.016764288
Commitment	BOOST	-0.9681563032	0.031843697
Commitment	END.MARK	-0.9570236617	0.042976338
Established Rule	ADJ	-0.9750168211	0.024983179
Established Rule	TTR	-0.9699192297	0.030080770
Example	PST	0.9585097392	0.041490261
Full Slippery Slope	PRED	0.9656109142	0.034389086
Full Slippery Slope	END.MARK	-0.9550589287	0.044941071
Full Slippery Slope	CONC	0.9909284364	0.009071564
Generic Ad Hominem	BOOST	-0.9918850402	0.008114960
Generic Ad Hominem	NEC	-0.9619114493	0.038088551
Generic Ad Hominem	TRANS.MARK	-0.9587447440	0.041255256
Generic Ad Hominem	POSS	-0.9544474731	0.045552527
Position to know	TTR	0.9734280639	0.026571936
Position to know	DEM	0.9699998891	0.030000111
Position to know	DECLA	0.9660265622	0.033973438
Position to know	ACT	0.9563050843	0.043694916
Position to know	PT	0.9556470133	0.044352987
Position to know	SIMP	0.9556470133	0.044352987
Position.to.know	ADJ	0.9972293166	0.002770683
Practical Reasoning	TP	-0.9913176852	0.008682315
Sign	TTR	0.9941818892	0.005818111
Sign	ADJ	0.9721983895	0.027801610
Verbal Classification	TP	0.9826274299	0.017372570
Verbal Classification	TRANS.MARK	0.9629150286	0.037084971

6.3. Discussion

The Legal Schemes Annotation Guidelines can be considered a useful tool to help the annotator identify the legal scheme their facing when annotating a legal corpus, as, in our corpus of four judgments, annotators encountered a broad variety of schemes, even if there were many which were classified in the 'Default Inference' category, as they did not fit in any other. Nevertheless, the values resulting from the calculation of the Cohen's Kappa indicate us that the Guidelines are not reliable enough to structure and annotate a legal corpus with Legal Schemes, at least in the hands of an annotation team with the same characteristics as ours (no previous expertise or academic knowledge in the legal field or in the annotation of legal texts, but with expertise in annotating other type of texts).

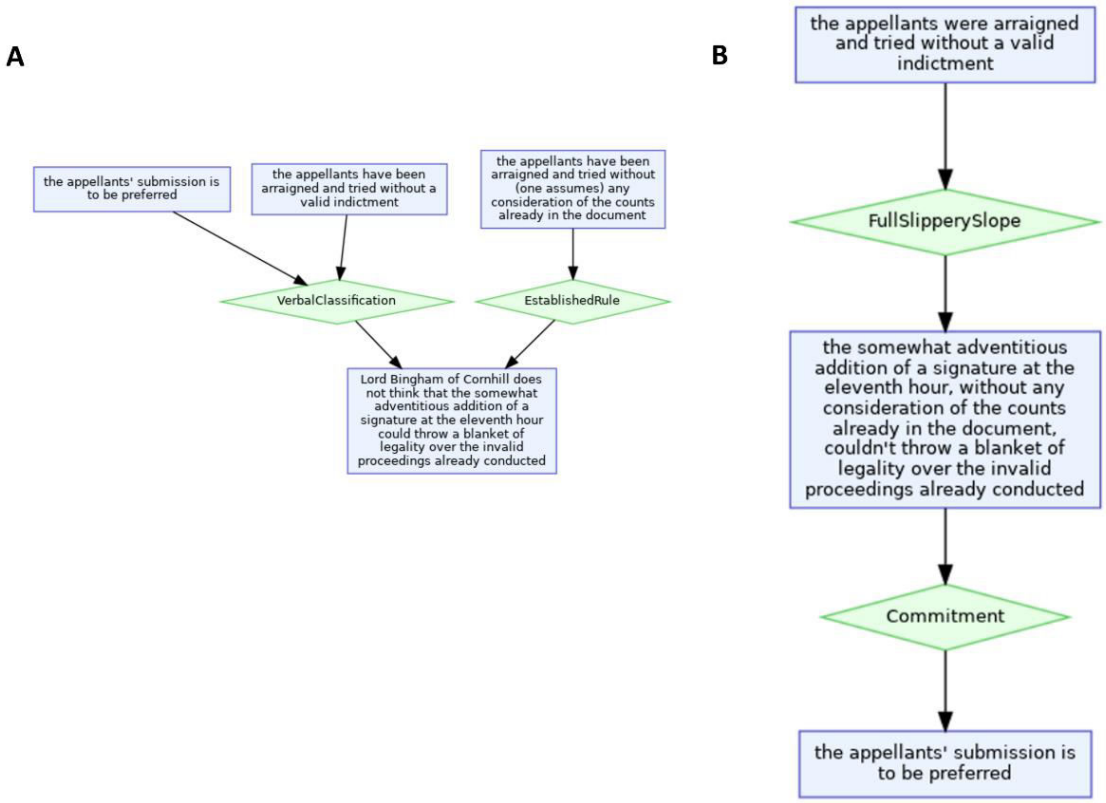


Figure 26: Two annotations of the same legal scheme I

The main possible reasons for the results of this IAA test are the following: (1) the annotation team's lack of expertise / knowledge on the legal domain, (2) scarce indications for identifying premise (P) and conclusions (C) in an argument, resulting in argumentation maps with a different structure (Figure 26), (3) annotation team with lack of previous experience in the annotation of a legal corpus.

When the annotation process started, some annotators had difficulties when annotating the corpus, such as not understanding parts of the text, not knowing the reason why judges were continuously rephrasing or making quotes in the judgments, or whether they should omit arguments contained in the quotes. Moreover, the fact that the Guidelines created offered very general and scarce indications on how to detect an argument scheme made the first and second annotator of the same chunk of text create different argumentation maps, resulting even in, for example, schemes having a Premise, that for the second annotator is a Conclusion, and a Conclusion, which is precisely is the Premise in the second annotation.

Moreover, if we wanted to better assess the reliability of the Guidelines exclusively for the scheme's categorisation (once the argumentation map is structured), rather than the whole argumentation process, the second annotator could have been given the argumentation map produced by the first annotator.

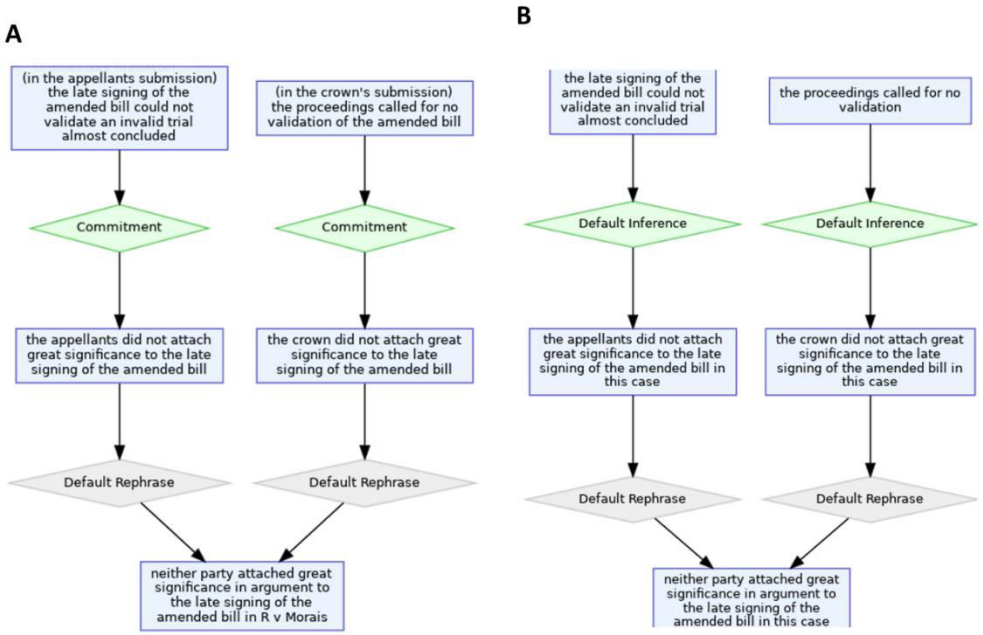


Figure 27: Two annotations of the same legal scheme II

We can further compare the way the team members annotated the argumentation schemes having a look at other argumentation maps in the database created. For example, in Figure 27, the former argumentation map (A) contains two argumentation schemes classified as 'Commitment'. Moreover, in (B) we can see that annotators were not able to classify the schemes (so they selected the 'Default Inference' option), although they did agree in the structure of the argument (that is, the premise and the conclusion were identified in the same way) created by the annotators in A. They even agreed in considering that the last chunk of text (blue boxes) is a rephrase of the conclusions of the two argumentation schemes.

In Figure 28, there is once more disagreement between the former (A) and the latter (B) annotator: in A, the premise is the fact that suicide was not a possibility that is reasonably foreseeable, and that is why the suicide fell outside the employer's duty (they did not have the obligation to prevent that since it was unforeseeable). Nevertheless, annotator B considers that the premise is the fact that addressing a suicide is outside the employer's duty, and that is precisely the cause for it being unforeseeable for him (conclusion). This disagreement in the structure of the argument might have caused the different classification for the argumentation schemes, which is labelled as 'Verbal Classification' in A, and as 'Position to Know' in B.

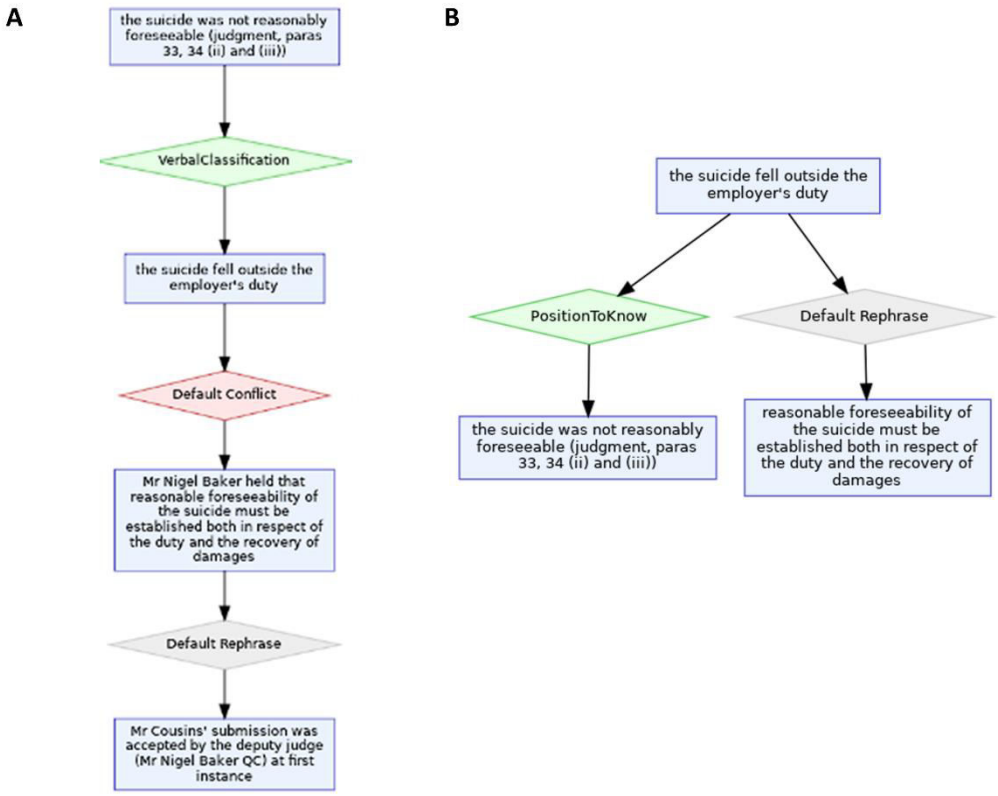


Figure 28: Two annotations of the same legal scheme III

Figure 8 shows another argumentation map which has been structured with the same premises and conclusions by annotators A and B, but the classification of the argumentation scheme is different. A considers that this argumentation scheme

is a Verbal Classification. As such, the premise would have the logic structure of 'a has property f' and 'for all x, can be considered to have property g, if has property f', so in the conclusion, a has property g.

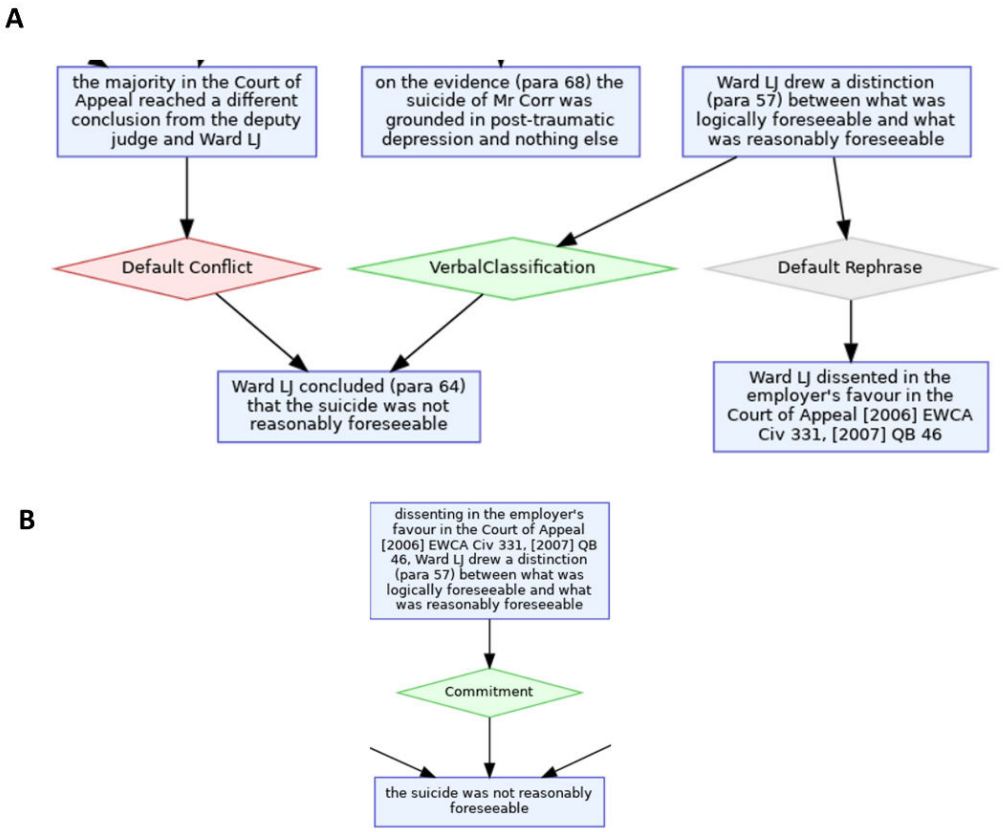


Figure 29: Two annotations of the same legal scheme IV

Moreover, if we consider this argument as an Argument from Commitment, we will conclude x because x was claimed previously by the judge (the distinction between the two types of what it is foreseeable, and the judge commits to that claim. The problem here is that both classifications are somewhat correct but at the same time they do not fit completely in any of these categories. The solution in this case would be being able to ascertain whether the premise depends on something external to the argument itself (in this case, the judge's claim and commitment, or rather the judge's argument itself).

Regarding the Pearson's correlation results, these potential linguistic and metadiscourse structure around each legal scheme can be seen if we have a look at specifical examples of excerpts where these schemes were found by the annotators: According to the correlations, Verbal Classification might be encountered when having transition markers and third person in our texts, as we can see in Excerpt 12.

Excerpt 12: Verbal Classification - TP and Transition Markers

***The case of R v Draz was** more immediately germane to the present case since **the second and third questions posed** for consideration (para 65) **were** whether **the judge had been** correct to conclude, when following the procedure under paragraph 7 of Schedule 3 to the Crime and Disorder Act 1998, that **it was unnecessary** for an indictment to be preferred and, if an **indictment should have been** preferred, whether **the absence of a signed indictment was** fatal to the validity of the proceedings. **Paragraph 7 of Schedule 3 to the 1998 Act opened**:* *[...]* *(a) **a person has been sent** for trial under section 51 of this Act but has not been arraigned; and* *(b) **the person is charged on** an indictment which (following amendment of the indictment, or as a result of an application under paragraph 2 above, or for any other reason), includes no offence that is triable only on indictment* *(3) **The court shall** cause to be read to the accused each count of the indictment that charges an offence triable either way.*

In turn, **adjectives**, present tense, and *simple aspect* are supposed to be found in Position to know schemes, such as in Excerpt 13. This scheme is also correlated with demonstrative pronouns, declarative statements, and active voice. Position to know is also correlated positively with the TTR: this excerpt has 153 tokens and 95 types, so we get around a 0.6 type-token ratio, which is a moderate level.

Excerpt 13: Position to know - Adjectives, Present Tense and Simple Aspect

*Mr Corr, the respondent's husband, was injured at work by the negligence of the **appellant** company, his employers. The accident he suffered could easily have killed him but in the event inflicted on him **serious and disfiguring** injuries to his head but left him alive. It is **easy** to understand that the repercussions of an injury of that character may have an **enduring** effect on the mental state of the victim, continuing after the **physical** effects are spent. So, it was with Mr Corr. He became clinically **depressed**, **bad-tempered,** and suffered from nightmares. He was treated with electro-convulsive therapy. All of this was, it is accepted, a result of the accident. Mr Corr also began to entertain thoughts of suicide. This, it is accepted, was a symptom of his **clinical** depression. On 23 May 2002, nearly six years after the accident, Mr Corr did commit suicide. In doing so he acted deliberately, **aware** of the consequences and with the intention of killing himself. The action which has now reached your Lordships' House is the action brought by his widow, Mrs Corr, under the Fatal Accidents Act 1976.*

In Example schemes, we are supposed to find past tenses preferably according to the correlations (Excerpt 14).

Excerpt 14: Example - Past tense

*These proceedings **were begun** by Mr Corr in June 1999, shortly before expiry of the three year limitation period, claiming damages for the physical and psychological injuries which he had suffered. The proceedings **were amended** after his death to substitute his widow and personal representative as claimant. She claims for the benefit of Mr Corr's estate pursuant to the Law Reform (Miscellaneous Provisions) Act 1934 and also for herself as a dependant of the deceased under the 1976 Act. The first of these claims has not been contentious. The second is a claim to recover the financial loss attributable to Mr Corr's suicide, and that alone is in issue in this appeal.*

6.4. Conclusions

This study had several limitations abovementioned that led to an inter-annotator reliability test not strong enough to assure that the new annotation guidelines created for legal schemes are reliable enough to be later used, at least with a group of annotators of the same characteristics (no previous experience or knowledge annotating highly specialised legal texts).

Nevertheless, the Pearson's correlation results show some interesting significant correlations providing us with a first insight on how linguistic and metadiscourse resources might be useful indicators to identify (legal) argumentation schemes. This research creates a framework for future studies interested in further exploring legal argumentation from the linguistics insight, since it has found some patterns that are worth further studying with larger data, which could lead to significant correlations in the patterns already found.

In addition, the new annotation guidelines for legal schemes elaborated for this study can be considered a starting point to develop more specifical and improved guidelines for non-expert annotators in the field, giving more steps to properly identified the conclusions and premises in an argument. In short, this research establishes the grounds for a new line of research trying to identify whether the presence of discourse markers is an indicator of an argument scheme or another and provides with a new legal corpus of annotated argumentation schemes.

FINAL CONCLUSION

In Section 2, three general purposes were established, and these were materialised by the conduction of more specifical tasks. This is outlined with more detailed in Figure 3. As a final conclusion to this PhD dissertation, whether each of the purposes has been fulfilled and to what extent is verified. Thus, the results obtained, and the limitations encountered are exposed. Finally, an outline of the consequences of the achievements and future potential lines of research are proposed.

- *Purpose 1: Development of new tools and resources for the study of Legal English*

Several R libraries were of great assistance in the process of corpus processing, data wrangling and data visualisation. R programming language and its libraries are powerful tools that researchers in linguistics, but their used is still not generalised due to the difficult learning process of learning to use programming languages. This dissertation provides with guidelines and R script for researchers attempting to replicate this study using R.

The first achievement of this dissertation was the compilation of a new legal corpus, the British Statute Law Corpus, which consists of a reliable representation of British legal provisions or statute law. This corpus will be made available online for the general public and will provide with a powerful dataset for future studies interested in the discourse of British legal provisions.

The Multidimensional Analysis Tagger is a convenient tool that automatically computes the complete process in a MD analysis from the corpus annotation process, the linguistics features extraction and the textual dimensions obtention. This dissertation provides further evidence that this tool is reliable, fast and easy to use.

- *Purpose 2: Proposal of new methods for the study of Legal English*

MD analysis is a well-established methodology in Corpus Linguistics, but not especially in the study of Legal English. There are only a few very recent studies that have applied this methodology focusing in some specifical aspects or genres of legal English (Granados-Meroño, 2023; Huang & Sang, 2024; Sun & Cheng, 2017) and only one study attempting to obtain a general insight of American Legal English by applying MD analysis (Goźdź-Roszkowski, 2011).

Moreover, these analyses tend to only perform an *ad hoc* MD analysis with small or moderate in size corpora, not locating the scores of the legal genres in comparison to genres from other fields or contexts. The studies in Chapters 4 and 5 do that, using the power of fine-tuning provided by R libraries 'psy' and '`sych' obtaining a comprehensive insight into the structure, differences and similarities between the two most relevant genres in British Public Law, and in contrast to the wide range of genres analysed in Biber's original MD analysis (Biber, 1988).

Moreover, the data obtained in these analyses provided with valuable information about the correlative structure of the 67 linguistic features considered, allowing the selection of the most relevant variables for the computation of future regressions that constitute the first steps towards the creation of specialised AI tools for the production, summarisation and detection of legal genres.

- *Purpose 3: Development of new approaches for the study of Legal English*

The combination of the insights from discourse studies and argumentation theory and technology have been combined to develop one of the first attempts to explore the influence of linguistic variables in the construction of argumentation schemes. Thus, the results in the study from Chapter 6 constitute a first approach to understanding the relationship between the discursive and logic structures in the construction of arguments.

These achievements do not mean, though, that there have not been any limitations to the development of these studies. Firstly, the corpus compiled represents the genre of legal provisions, therefore completing together with the several corpora available, specifically the BLaRC, the landscape of public legal genres. However, there exists a considerable gap in the corpus compilation and the study of private legal genres, such as contracts, last wills or deeds, due to the difficulty of obtaining samples of them. This is a task that future researchers should complete in order to have a complete understanding of legal discourse, which is produced not only by legal professionals, but also by non-experts citizens that are users of the tools provided by the legal framework.

Consequently, the MD analyses developed in this dissertation, though clarifying and insightful, are still lacking the inclusion of private legal genres as aforementioned.

Finally, the study in Chapter 6 is certainly a big step in the study of the relationship between linguistic and logic structures in argumentation, but only a first approach that needs further exploration, further empiric studies and new insights on the object of study to better understand the variables affecting argumentation.

References

Alamri, B. (2023). A Multidimensional Comparative Analysis of MENA and International English Research Article Abstracts in Applied Linguistics. *SAGE Open*, *13*(1), 21582440221145669. https://doi.org/10.1177/21582440221145669

Alcaraz, E. (2007). *El inglés jurídico: Textos y documentos* (6.ª). Ariel.

Alcaraz, E., & Hughes, B. (2002). *El español jurídico*. Ariel.

Alcaraz, E., & Hughes, B. (2015). *Legal translation explained* (Vol. 4). Routledge.

Alcaraz, E., Hughes, B., & González-Jover, A. G. (2014). *El español jurídico*. Ariel.

Álvarez Álvarez, S. (2008). Elementos cohesivos en el lenguaje jurídico: Análisis contrastivo de las sentencias judiciales en lengua inglesa y española. *La traducción del futuro: mediación lingüística y cultural en el siglo XXI, Vol. 1, 2008 (La traducción y su práctica), ISBN 978-84-477-1026-3, págs. 407-418*, 407–418. https://dialnet.unirioja.es/servlet/articulo?codigo=5660324

Artstein, R. (2017). Inter-annotator Agreement. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 297–313). Springer Netherlands. https://doi.org/10.1007/978-94-024-0881-2_11

Austin, J. L. (1975). *How to Do Things with Words*. Clarendon Press.

Benoit, K., & Matsuo, A. (2020). *spacyr: Wrapper to the 'spaCy' 'NLP' Library* (Version 1.2.1) [Computer software]. https://CRAN.R-project.org/package=spacyr

Benoit, K., Obeng, A., Watanabe, K., Matsuo, A., Nulty, P., & Müller, S. (2021). *readtext: Import and Handling for Plain and Formatted Text Files* (Version 0.81) [Computer software]. https://CRAN.R-project.org/package=readtext

Berber Sardinha, T., & Veirano Pinto, M. (Eds.). (2014). *Multi-Dimensional Analysis, 25 years on*. John Benjamins Publishing Company. https://benjamins.com/catalog/scl.60

Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Longman.

Bhatia, V. K. (2014). *Analysing Genre: Language Use in Professional Settings*. Taylor & Francis.

Biber, D. (1988). Variation across Speech and Writing. In *Variation across Speech and Writing*. https://doi.org/10.1017/cbo9780511621024

Biber, D. (2006). University Language. In *Scl.23*. John Benjamins Publishing Company. https://benjamins.com/catalog/scl.23

Biber, D., Connor, U., & Upton, T. A. (2007). Discourse on the Move. In *Scl.28*. John Benjamins Publishing Company. https://benjamins.com/catalog/scl.28

Biber, D., & Conrad, S. (2019). *Register, Genre, and Style*. Cambridge University Press. 10.1017/9781108686136

Biber, Douglas. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biel, Ł. (2017). Lexical bundles in EU law: The impact of translation process on the patterning of legal language. In *Phraseology in Legal and Institutional Settings*. Routledge.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*.

Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In *The Routledge Handbook of Corpus Linguistics*. Routledge.

Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics*, *18*(2), 229–253. https://doi.org/10.1075/ijcl.18.2.03bre

Cao, D. (2016). Translating Law. In *Translating Law*. https://doi.org/10.1080/09076760903073996

Clarke, I. (2022). A Multi-Dimensional Analysis of English tweets. *Language and Literature*, *31*(2), 124–149. https://doi.org/10.1177/09639470221090369

Eemeren, F. H. van, Garssen, B., Krabbe, E. C. W., Henkemans, F. A. S., Verheij, B., & Wagemans, J. H. M. (2014). *Handbook of Argumentation Theory*. Springer Netherlands.

Eemeren, F. H. van, Houtlosser, P., & Henkemans, A. F. S. (2007). *Argumentative Indicators in Discourse: A Pragma-Dialectical Study*. Springer Science & Business Media.

Ehret, K., & Taboada, M. (2021). Characterising Online News Comments: A Multi-Dimensional Cruise Through Online Registers. *Frontiers in Artificial Intelligence*, *4*. https://doi.org/10.3389/frai.2021.643770

Engberg, J. (2020). Comparative Law for Legal Translation: Through Multiple Perspectives to Multidimensional Knowledge. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *33*(2), 263–282. https://doi.org/10.1007/s11196-020-09706-9

Feteris, E. T. (2012). The role of the judge in legal proceedings: A Pragma-dialectical analysis. *Journal of Argumentation in Context*, *1*(2), 234–252. https://doi.org/10.1075/jaic.1.2.05fet

Feteris, E. T. (2017). The identification of prototypical argumentative patterns in the justification of judicial decisions: *Journal of Argumentation in Context*, *6*(1), 44–58. https://doi.org/10.1075/jaic.6.1.03fet

Garofalo, G. (2009). *Géneros discursivos de la justicia penal: Un análisis contrastivo español-italiano orientado a la traducción*. Franco Angeli.

Giampieri, P. (2024). Key n-Grams in EU Directives and in the UK National Legislation on Consumer Contracts. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *37*(1), 59–75. https://doi.org/10.1007/s11196-023-10087-y

Goźdź-Roszkowski, S. (2011). *Patterns of Linguistic Variation in American Legal English: A Corpus-based Study*. Peter Lang.

Goźdź-Roszkowski, S. (2020). Move Analysis of Legal Justifications in Constitutional Tribunal Judgments in Poland: What They Share and What They Do Not. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *33*(3), 581–600. https://doi.org/10.1007/s11196-020-09700-1

Goźdź-Roszkowski, S. (2021). Corpus Linguistics in Legal Discourse. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *34*(5), 1515–1540. https://doi.org/10.1007/s11196-021-09860-8

Granados-Meroño, D. (2023). Judgments of the English and Spanish Supreme Courts: A corpus-based approach to the legal English and Spanish discourse using multi-dimensional analysis. *Journal of Research Design and Statistics in Linguistics and Communication Science*, *7*(1), 21–68. https://doi.org/10.1558/jrds.22453

Granados-Meroño, D., & Orts, M. A. (2021). Traducir la corrupción: Análisis traductológico de un auto judicial de la «operación púnica». In *Aspectos forenses de la traducción e interpretación: Jurídica, judicial y policial, 2021, ISBN 9788413591391, págs. 93-117* (pp. 93–117). Colex. https://dialnet.unirioja.es/servlet/articulo?codigo=7866050

Grover, C., Hachey, B., & Hughson, I. (2004). The HOLJ Corpus. Supporting Summarisation of Legal Texts. In S. Hansen-Schirra, S. Oepen, & H. Uszkoreit (Eds.), *Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora* (pp. 47–54). COLING. https://aclanthology.org/W04-1907

Hajlaoui, N., Kolovratnik, D., Väyrynen, J., Steinberger, R., & Varga, D. (2014). DCEP -Digital Corpus of the European Parliament. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/pdf/943_Paper.pdf

Hinton, M., & Wagemans, J. H. M. (2022). Evaluating Reasoning in Natural Arguments: A Procedural Approach. *Argumentation*, *36*(1), 61–84. https://doi.org/10.1007/s10503-021-09555-1

Huang, Y., & Sang, Z. (2024). Linguistic variation in supreme court oral arguments by legal professionals: A novel multi-dimensional analysis. *Discourse Studies*, 14614456231221075. https://doi.org/10.1177/14614456231221075

Hyland, Ken. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.

Janier, M., Lawrence, J., & Reed, C. (2014). OVA+: An argument analysis interface. In S. Parsons, N. Oren, C. Reed, & F. Cerutti (Eds.), *Computational Models of Argument* (pp. 463–464). IOS Press. https://doi.org/10.3233/978-1-61499-436-7-463

Joaristi, L., & Lizasoain, L. (2008). Análisis factorial clásico y análisis factorial de información total: Análisis de pruebas de matemáticas de Primaria (5º y 6º cursos) y Secundaria obligatoria. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, *14*(2), Article 2. https://doi.org/10.7203/relieve.14.2.4191

Kalinowski, T., Ushey, K., Allaire, J. J., RStudio, Tang [aut, Y., cph, Eddelbuettel, D., Lewis, B., Keydana, S., Hafen, R., library, M. G. (TinyThread, & http://tinythreadpp.bitsnbites.eu/). (2023). *reticulate: Interface to 'Python'* (Version 1.28) [Computer software]. https://CRAN.R-project.org/package=reticulate

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). *The Sketch Engine*. Universite de Bretagne-Sud. https://is.muni.cz/publication/560635/en/The-Sketch-Engine/Kilgarriff-Rychly-Smrz-Tugwell

Lawrence, J., & Reed, C. (2015). Combining Argument Mining Techniques. *Proceedings of the 2nd Workshop on Argumentation Mining*, 127–136. https://doi.org/10.3115/v1/W15-0516

Lawrence, J., Visser, J., & Reed, C. (2019). An Online Annotation Assistant for Argument Schemes: Proceedings of the 13th Linguistic Annotation Workshop. *Proceedings of the 13th Linguistic Annotation Workshop*, 100–107. https://doi.org/10.18653/v1/W19-4012

Lawrence, J., Visser, J., Walton, D., & Reed, C. (2020). *A decision tree for annotating argumentation scheme corpora: 3rd European Conference on Argumentation*. 97–114. http://ecargument.org/

Le Foll, E. (2024). Textbook English. In *Scl.116*. John Benjamins Publishing Company. https://benjamins.com/catalog/scl.116

Llisterri, J., & Torruella Casañas, J. (1999). Diseño de corpus textuales y orales. In *Filología e informática: Nuevas tecnologías en los estudios filológicos, 1999, ISBN 84-89790-41-8, págs. 45-81* (pp. 45–81). Seminario de Filología e Informática. https://dialnet.unirioja.es/servlet/articulo?codigo=595883

Marín, M. J., & Rea Rizzo, C. (2012). Structure and Design of the British Law Report Corpus (BLRC): A Legal Corpus of Judicial Decisions from the UK. *Journal of English Studies, 10*, 131–145. https://doi.org/10.18172/jes.184

Matulewska, A. (2014). A review of 'Patterns of linguistic variation in American Legal English. A corpus based study' by Stanislaw Gostanisław Goźdź-Roszkowski. *Comparative Legilinguistics*, *19*, 135–138. https://doi.org/10.14746/cl.2014.19.07

Mochales, R., & Moens, M.-F. (2008). *Study on the Structure of Argumentation in Case Law* (p. 20). https://doi.org/10.3233/978-1-58603-952-3-11

Mochales, R., & Moens, M.-F. (2009). Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the International Conference on Artificial Intelligence and Law* (p. 107). https://doi.org/10.1145/1568234.1568246

Mochales, R., & Moens, M.-F. (2011). Argumentation Mining. *Artificial Intelligence and Law*, *19*(1), 1–22. https://doi.org/10.1007/s10506-010-9104-x

Musi, E., & Rocci, A. (2017). Evidently epistential adverbs are argumentative indicators: A corpus-based study. *Argument & Computation*, *8*(2), Article 2.

Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In *Multi-Dimensional Analysis: Research Methods and Current Issues* (pp. 67–94). Bloomsbury Academic.

Ooms, J. (2023). *pdftools: Text Extraction, Rendering and Converting of PDF Documents* (Version 3.3.3) [Computer software]. https://CRAN.R-project.org/package=pdftools

Orts, M. Á. (2015). Power distance and persuasion: The tension between imposition and legitimation in international legal genres. *Journal of Pragmatics*, *Journal of Pragmatics (2016), pp. 1–16*. https://doi.org/10.1016/j.pragma.2015.11.009

Orts, M. Á. (2016). Opacity in International Legal Texts: Generic Trait or Symbol of Power? *Revista Alicantina de Estudios Ingleses*, *28*, 119–145. https://doi.org/10.14198/raei.2015.28.07

Parodi, G. (2003). Lingüística de corpus y análisis multidimensional: Exploración de la variación en el corpus PUCV-2003. *Revista Española de Lingüística*, *35*(1), 45–76.

Pearson, J. (1998). Terms in Context. In *Scl.1*. John Benjamins Publishing Company. https://benjamins.com/catalog/scl.1

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. University of Notre Dame Press.

Piszcz, A., & Sierocka, H. (2020). The Role of Culture in Legal Languages, Legal Interpretation and Legal Translation. *International Journal for the Semiotics of Law = Revue Internationale De Semiotique Juridique*, *33*(3), 533–542. https://doi.org/10.1007/s11196-020-09760-3

Rodríguez-Puente, P., & Hernández-Coalla, D. (2023). A new tool for analysing recent changes in English legal discourse. *ICAME Journal*, *47*(1), 109–117. https://doi.org/10.2478/icame-2023-0006

Rossini Favretti, R., Tamburini, F., & Martelli, E. (2007). Words from Bononia Legal Corpus. In W. Teubert (Ed.), *Text Corpora and Multilingual Lexicography* (pp. 11–30). John Benjamins Publishing Company. https://doi.org/10.1075/bct.8.03ros

Ruiz Moneva, M. A. (2013). Cognition and context of legal texts: Spanish and English judgments compared. *Revista de Lingüística y Lenguas Aplicadas*, *8*(0). https://doi.org/10.4995/rlyla.2013.1245

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Shakir, M. (2024). An exploratory investigation of functional variation in South Asian online Englishes. *English Language & Linguistics*, 1–30. https://doi.org/10.1017/S1360674324000017

Sun, Y., & Cheng, L. (2017). Linguistic variation and legal representation in legislative discourse: A corpus-based multi-dimensional study. *International Journal of Legal Discourse*, *2*(2), 315–339. https://doi.org/10.1515/ijld-2017-0017

Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.

Toulmin, S. E. (1958). *The Uses of Argument*. University Press. https://doi.org/10.1017/CBO9780511840005

UCREL. (1987). *Free CLAWS web tagger* [Computer software]. Lancaster University. http://ucrel-api.lancaster.ac.uk/claws/free.html

Vass, H. (2017). The Role of Hedging in Balancing Power and Persuasion in the Judicial Context: The case of majority and dissenting opinions. In *Power, Persuasion and Manipulation in Specialised Genres: Providing Keys to the Rhetoric of Professional Communities*. Peter Lang. https://pureportal.coventry.ac.uk/en/publications/the-role-of-hedging-in-balancing-power-and-persuasion-in-the-judi

Walton, D. (2006). *Fundamentals of Critical Argumentation*. Cambridge University Press.

Walton, D. (2010). *Legal Argumentation and Evidence*. Penn State Press.

Walton, D., & Macagno, F. (2015). A classification system for argumentation schemes. *Argument & Computation*, *6*(3), 219–245. https://doi.org/10.1080/19462166.2015.1123772

Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*. Cambridge University Press.

Webber, B. (2011). Discourse Structures and Language Technologies. *Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA 2011)*, 12–16. https://aclanthology.org/W11-4603

Wojtasik-Dziekan, E. (2020). Analysis of the Semantic Scope of Two Korean Terms Equivalent to English Court. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *33*(3), 657–671. https://doi.org/10.1007/s11196-020-09693-x

Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes*, *28*(4), 421–450. https://doi.org/10.1111/j.1467-971X.2009.01606.x

Xiao, R. Z. (2008). Theory-driven corpus research: Using corpora to inform aspect theory. In A. Ludeling & M. Kyto (Eds.), *Corpus Linguistics: An International Handbook* (Vol. 1). Mouton de Gruyter.

Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

Ziemski, M., Junczys-Dowmunt, M., & Pouliquen, B. (2016). The United Nations Parallel Corpus v1.0. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 3530–3534). European Language Resources Association (ELRA). https://aclanthology.org/L16-1561

CODE

Code 1: R Code for Corpus processing

```
#0. Install and load the following packages

library(pdftools)library(dplyr)

library(stringr)

#A. Working with one single document

#1.a. Provide R with the URL or path to your document
pdf_path <- "C:\\Users\\Daniel\\Desktop\\Texts Workshop CILC23\\58.pdf"

#2.a. Extract the text using R
txt_output <- pdftools::pdf_text(pdf_path) %>%
  paste(sep = " ") %>%
  stringr::str_replace_all(fixed("\n"), " ") %>%
  stringr::str_replace_all(fixed("\r"), " ") %>%
  stringr::str_replace_all(fixed("\t"), " ") %>%
  stringr::str_replace_all(fixed("\""), " ") %>%
  paste(sep = " ", collapse = " ") %>%
  stringr::str_squish() %>%
  stringr::str_replace_all("- ", "")

#3.a. Inspect the text
str(txt_output)

#4.a. save one single text to a .txt file
write(txt_output, file="58.txt")

#B. Working with multiple documents

#1.b. Provide R with the URL or path to your documents
dirpath<-"C:/Users/Daniel/Desktop/Texts Workshop CILC23"

#2.b. Create a function allowing us to create plain txt files from several PDF
documents.


convertpdf2txt <- function(dirpath){
  files <- list.files(dirpath, full.names = T)
  x <- sapply(files, function(x){
    x <- pdftools::pdf_text(x) %>%
      paste(sep = " ") %>%
      stringr::str_replace_all(fixed("\n"), " ") %>%
      stringr::str_replace_all(fixed("\r"), " ") %>%
      stringr::str_replace_all(fixed("\t"), " ") %>%
      stringr::str_replace_all(fixed("\""), " ") %>%
      paste(sep = " ", collapse = " ") %>%
      stringr::str_squish() %>%
      stringr::str_replace_all("- ", "")
    return(x)
  })
}
```

```
#3.b. Apply the function created to your imported texts
txts <- convertpdf2txt(dirpath)
str(txts)

#4.b. Add names to your txt files
names(txts) <- paste("textworkshopcilc23-", 1:length(txts), sep = "")

#5.b. Save result to disc
lapply(seq_along(txts), function(i)writeLines(text = unlist(txts[i]),
                                              con = paste("C:\\Users\\Daniel\\De
sktop\\Texts Workshop CILC23",
                                                          names(txts)[i],
                                                          ".txt",
                                                          sep = "")))
```

Code 2: R Code for corpus cleaning and basic textual data analysis

```
#0. Install and load the following packages
library(reticulate)

library(spacyr)

library(Rcpp)

library(quanteda)

library(dplyr)

#1. Import texts and create a corpus
Judgments<-readtext("C:\\Users\\Daniel\\Desktop\\Texts Workshop CILC23")
Judgmentscorpus<-corpus(Judgments)

#2. Tokenize text removing numbers and punctuation
Judgmentstokens <- tokens(Judgmentscorpus,
                          what = "word",
                          remove_numbers = TRUE,
                          remove_punct = TRUE)

#3. Remove stop words
library(stringr)

## Warning: package 'stringr' was built under R version 4.3.2

Judgmentstokens_cleaned <- tokens_remove(Judgmentstokens,
                                         stopwords("english"))

#Alternative way to remove punctuation
punctuation <- c(",", ".", "!", "?", ";", ":", "-", "'", "\"", "(", ")", "[",
"]", "{", "}")
Judgmentstokens_cleaned <- tokens_remove(Judgmentstokens_cleaned,
                                         pattern = paste0("\\",
                                                          punctuation,
                                                          collapse = "|"))


#4. Convert tokens to a document-feature matrix (DFM)
Judgmentsdfm <- dfm(Judgmentstokens_cleaned)


#5.Create a wordcloud
library(quanteda.textplots)
set.seed(100)
textplot_wordcloud(Judgmentsdfm, min_freq = 6, random_order = FALSE,
                   rotation = .25,
                   colors = RColorBrewer::brewer.pal(8, "Dark2"))

topfeatures(Judgmentsdfm, 50)

#6. Display 10 most common words in a barplot
library(quanteda.textstats)

library(tidyr)
# calculate term frequency
tf <- textstat_frequency(Judgmentsdfm)
```

```
# view top 10 most frequent words
top_words <- head(tf, 10)
print(top_words)

# create a bar chart of top 10 most frequent words
barplot(top_words$frequency, names.arg = top_words$feature,
        xlab = "Word", ylab = "Frequency",
        main = "Top 10 Most Frequent Words")

#7. Calculate Lexical Diversity
library(quanteda)
# calculate lexical density
unique_words <- sum(Judgmentsdfm > 0)
total_words <- sum(Judgmentsdfm)
lexdensity <- unique_words / total_words

# display results
cat(paste0("Lexical Density: ", round(lexdensity * 100, 2), "%"))
```

Code 3: R code for data preparation for FA and FA computation

```
#1.1. Normality and homogeneity tests
library(psych)
library(psy)
library(nortest)

lillie.test(unlist(variables_preplegalcorpus[2:68]))

fligner.test(variables_preplegalcorpus[2:68], n=1941)

library(corrplot)

#1.2. Correlation matrix
matrix<-cor(variables_preplegalcorpus[2:68])
corrplot(matrix, method="number", type="upper")

#2.1. Scree plot
eigenvalues<-eigen(matrix)
scree.plot(variables_preplegalcorpus[2:68])
totalvariance<-eigenvalues$values/sum(eigenvalues$values)*100

#2.2. Parallel Analysis
fa.parallel(variables_preplegalcorpus[2:68], n.obs=1941, fa="fa", fm =
"minres")

#3.1. Factor Analysis. We choose promax as it allows correlation
between factors (significant cross-loadings), suitable for large
datasets.
resultadosFA<-factanal(variables_preplegalcorpus[2:68], cor =
"matrix", factors = 6, scores = c("regression"), rotation="promax")
FactorscoresFA<-data.frame(resultadosFA$scores) #Factor scores

print(resultadosFA, digits=2, cutoff=0.35, sort=TRUE)

#4. Create a data frame with the results
loading_df_wide<-as.data.frame(unclass(resultadosFA$loadings))
loading_df_wide <- loading_df_wide %>%
  rownames_to_column(var = "Variable")



#5. Factor scores for resultados FA
factor_loadings <- resultadosFA$loadings
factanalscores_law_report <- as.matrix(variables_preplawreport[2:68])
%*% factor_loadings

#3.2 FA Method 2
Modelo_promax<-fa(variables_preplegalcorpus[2:68], nfactors = 6,
rotate = "promax", scores = "regression")
Modelo_promax$communality #We check communality
```

DATA

Data 1: Linguistic features - abbreviations and description

ABBREVIATION	VARIABLE NAME	DESCRIPTION
AMP	Amplifiers	Words increasing intensity (e.g., 'very', 'extremely').
ANDC	Independent clause coordination	'And' connecting independent clauses.
AWL	Word length	Average word length in letters.
(X.) BEMA	'Be' as main verb	Occurrences where 'be' acts as a main verb.
(X.) BYPA	By-passives	Passive constructions including 'by'.
CAUS	Causative adverbial subordinators	Words like 'because' indicating causality.
CONC	Concessive adverbial subordinators	Words like 'although', 'though'.
COND	Conditional adverbial subordinators	Words like 'if', 'unless'.
CONJ	Conjuncts	Conjunctions like 'therefore', 'moreover'.
(X.) CONT	Contractions	Count of contracted forms (e.g., "n't", "'ll").
DEMO	Demonstratives	Non-pronoun uses of 'this', 'that', etc.
DEMP	Demonstrative pronouns	Pronouns such as 'this', 'that', 'these', 'those'.
DPAR	Discourse particles	Words like 'well', 'anyway' used in discourse.
DWNT	Downtoners	Words like 'almost', 'barely', 'slightly'.
EMPH	Emphatics	Words adding emphasis (e.g., 'just', 'really').
EX	Existential 'there'	Instances of 'there' as an existential marker.
FPP1	First-person pronouns	Pronouns referring to the speaker (e.g., 'I', 'we').
GER	Gerunds	Verbs ending in '-ing' used as nouns.
HDG	Hedges	Expressions indicating uncertainty (e.g., 'maybe', 'sort of').
INPR	Indefinite pronouns	Pronouns like 'someone', 'anything'.
JJ	Attributive adjectives	Adjectives modifying nouns.
NEMD	Necessity modals	Modal verbs indicating necessity (e.g., 'must', 'should').
NN	Total other nouns	Count of common nouns excluding nominalizations and gerunds.
NOMZ	Nominalizations	Nouns ending in '-tion', '-ment', '-ness', or '-ity'.
OSUB	Other adverbial subordinators	Words introducing adverbial clauses (e.g., 'since', 'while').
(X.) PASS	Agentless passives	Passive constructions without an agent.

ABBREVIATION	VARIABLE NAME	DESCRIPTION
(X.) PASTP	Past participial clauses	Clauses using past participles.
PEAS	Perfect aspect	Instances of 'have' followed by a past participle.
PHC	Phrasal coordination	Coordinated phrases of the same category.
PIN	Total prepositional phrases	Count of all prepositional phrases.
(X.) PIRE	Pied-piping relative clauses	Relative clauses where the WH-word follows a preposition.
PIT	Pronoun 'it'	Occurrences of the pronoun 'it'.
PLACE	Place adverbials	Adverbs indicating location (e.g., 'abroad', 'behind').
POMD	Possibility modals	Modal verbs indicating possibility (e.g., 'can', 'might').
PRED	Predicative adjectives	Adjectives used after 'be' or similar verbs.
(X.) PRESP	Present participial clauses	Clauses using present participles.
(X.) PRIV	Private verbs	Verbs indicating cognition (e.g., 'think', 'believe').
PRMD	Predictive modals	Modal verbs indicating prediction (e.g., 'will', 'shall').
(X.) PROD	Pro-verb 'do'	Instances of 'do' used as a main verb.
(X.) PUBV	Public verbs	Verbs indicating speech acts (e.g., 'say', 'declare').
RB	Total adverbs	Count of all adverbs.
(X.) SERE	Sentence relatives	Clauses modifying entire sentences, typically starting with 'which'.
(X.) SMP	'Seem' or 'appear' verbs	Instances of 'seem' or 'appear'.
(X.) SPAU	Split auxiliaries	Auxiliary verbs separated from the main verb by an adverb.
(X.) SPIN	Split infinitives	Infinitives interrupted by an adverb.
SPP2	Second-person pronouns	Pronouns referring to the addressee (e.g., 'you').
(X.) STPR	Stranded prepositions	Prepositions appearing at the end of clauses.
(X.) SUAV	Suasive verbs	Verbs suggesting persuasion (e.g., 'suggest', 'recommend').
SYNE	Synthetic negation	Instances of 'no' modifying a noun.
THAC	That adjective complements	Clauses introduced by 'that' following an adjective.
(X.) THATD	Subordinator 'that' deletion	Instances where 'that' is omitted.
THVC	That verb complements	Clauses introduced by 'that' following a verb.

ABBREVIATION	VARIABLE NAME	DESCRIPTION
TIME	Time adverbials	Adverbs indicating time (e.g., 'yesterday', 'soon').
TO	Infinitives	Instances of 'to' marking an infinitive verb.
TOBJ	That relative clauses (object position)	Relative clauses where 'that' replaces an object.
TPP3	Third-person pronouns	Pronouns referring to others (e.g., 'he', 'they').
TSUB	That relative clauses (subject position)	Relative clauses where 'that' replaces a subject.
TTR	Type-token ratio	Ratio of unique words to total words.
VBD	Past tense	Verbs in the past tense (e.g., 'went', 'saw').
VPRT	Present tense	Present tense verbs (e.g., 'goes', 'runs').
(X.) WHCL	WH-clauses	Subordinate clauses introduced by WH-words.
(X.) WHOBJ	WH relative clauses (object position)	Relative clauses starting with 'who', 'which' as objects.
(X.) WHQU	Direct WH-questions	Questions starting with 'what', 'where', etc.
(X.) WHSUB	WH relative clauses (subject position)	Relative clauses starting with 'who', 'which' as subjects.
(X.) WZPAST	Past participial WHIZ relatives	WH-relative clauses with past participial verbs.
(X.) WZPRES	Present participial WHIZ relatives	WH-relative clauses with present participial verbs.
XX0	Analytic negation	Instances of 'not' or 'n't'.

Data 2: Descriptive statistical values of linguistic features in the BLaRC

FACTORS	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
AWL	1,229	4.68	0.15	4.67	4.68	0.15	4.16	5.21	1.05	0.12	0.06	0.00
TTR	1,229	194.54	16.63	196.00	194.93	16.31	122.00	246.00	124.00	-0.25	0.13	0.47
AMP	1,229	0.11	0.09	0.10	0.10	0.07	0.00	0.85	0.85	2.12	8.33	0.00
ANDC	1,229	0.42	0.24	0.38	0.40	0.21	0.00	2.05	2.05	1.24	3.14	0.01
X.BEMA.	1,229	1.66	0.37	1.64	1.65	0.34	0.37	3.05	2.68	0.32	0.44	0.01
X.BYPA.	1,229	0.21	0.10	0.20	0.21	0.09	0.00	0.93	0.93	1.06	3.91	0.00
CAUS	1,229	0.10	0.08	0.09	0.09	0.07	0.00	0.58	0.58	1.34	2.89	0.00
CONC	1,229	0.07	0.06	0.06	0.06	0.04	0.00	0.58	0.58	2.05	9.98	0.00
COND	1,229	0.28	0.15	0.28	0.28	0.13	0.00	1.11	1.11	0.68	1.31	0.00
CONJ	1,229	0.46	0.17	0.45	0.45	0.15	0.00	1.43	1.43	0.62	1.41	0.00
X.CONT.	1,229	0.01	0.05	0.00	0.00	0.00	0.00	0.68	0.68	8.69	93.93	0.00
DEMO	1,229	1.30	0.36	1.26	1.28	0.31	0.12	3.06	2.94	0.57	1.19	0.01
DEMP	1,229	0.53	0.20	0.51	0.51	0.18	0.00	1.67	1.67	0.88	2.44	0.01
DPAR	1,229	0.01	0.02	0.00	0.00	0.00	0.00	0.30	0.30	6.52	61.11	0.00
DWNT	1,229	0.16	0.08	0.16	0.16	0.07	0.00	0.72	0.72	0.73	2.34	0.00
EMPH	1,229	0.21	0.11	0.20	0.21	0.10	0.00	0.73	0.73	0.84	1.73	0.00
EX	1,229	0.32	0.17	0.29	0.30	0.15	0.00	1.34	1.34	1.25	3.21	0.00
FPP1	1,229	0.86	0.53	0.79	0.81	0.42	0.00	4.45	4.45	1.64	5.51	0.02
GER	1,229	0.63	0.40	0.54	0.57	0.30	0.00	4.68	4.68	2.44	13.17	0.01
HDG	1,229	0.00	0.01	0.00	0.00	0.00	0.00	0.23	0.23	6.08	59.58	0.00
INPR	1,229	0.04	0.04	0.03	0.03	0.04	0.00	0.56	0.56	3.24	26.90	0.00
JJ	1,229	4.57	0.99	4.50	4.53	0.96	1.67	9.13	7.46	0.40	0.38	0.03
NEMD	1,229	0.30	0.17	0.28	0.29	0.15	0.00	1.19	1.19	1.08	2.19	0.00
NN	1,229	23.20	2.38	23.03	23.09	2.18	17.25	37.37	20.12	0.64	1.64	0.07
NOMZ	1,229	3.93	1.17	3.82	3.88	1.13	0.46	9.23	8.77	0.53	0.63	0.03
OSUB	1,229	0.13	0.08	0.11	0.12	0.06	0.00	0.57	0.57	1.17	2.64	0.00
X.PASS.	1,229	1.67	0.40	1.63	1.65	0.37	0.48	4.91	4.43	0.88	3.80	0.01
X.PASTP.	1,229	0.07	0.05	0.06	0.06	0.04	0.00	0.57	0.57	1.94	9.86	0.00
X.PEAS.	1,229	1.03	0.43	0.95	1.00	0.37	0.00	3.71	3.71	1.01	2.16	0.01
PHC	1,229	0.54	0.26	0.51	0.52	0.22	0.00	1.91	1.91	1.20	2.80	0.01
PIN	1,229	12.65	1.09	12.64	12.65	1.05	9.29	16.90	7.61	0.01	0.08	0.03
X.PIRE.	1,229	0.22	0.12	0.20	0.21	0.10	0.00	0.97	0.97	0.93	2.41	0.00
PIT	1,229	1.15	0.41	1.12	1.12	0.37	0.00	3.13	3.13	0.79	1.69	0.01
PLACE	1,229	0.16	0.12	0.14	0.15	0.09	0.00	1.24	1.24	2.36	12.27	0.00
POMD	1,229	0.52	0.20	0.51	0.52	0.18	0.00	1.39	1.39	0.36	0.47	0.01
PRED	1,229	0.82	0.26	0.80	0.81	0.22	0.00	2.09	2.09	0.69	1.87	0.01
X.PRESP.	1,229	0.11	0.07	0.10	0.10	0.06	0.00	0.59	0.59	1.31	3.61	0.00
X.PRIV.	1,229	1.29	0.35	1.25	1.28	0.33	0.00	2.59	2.59	0.35	0.40	0.01
PRMD	1,229	0.56	0.26	0.53	0.54	0.24	0.00	1.95	1.95	0.80	1.78	0.01
X.PROD.	1,229	0.06	0.06	0.05	0.05	0.04	0.00	0.69	0.69	2.77	15.56	0.00
X.PUBV.	1,229	0.88	0.33	0.84	0.86	0.30	0.00	2.91	2.91	0.91	1.97	0.01
RB	1,229	2.15	0.51	2.13	2.14	0.50	0.48	4.20	3.72	0.29	0.36	0.01
X.SERE.	1,229	0.11	0.07	0.10	0.10	0.07	0.00	0.47	0.47	0.83	0.80	0.00
X.SMP.	1,229	0.13	0.10	0.10	0.11	0.07	0.00	1.03	1.03	1.99	7.91	0.00
X.SPAU.	1,229	0.37	0.14	0.36	0.36	0.13	0.00	1.34	1.34	0.71	2.52	0.00

FACTORS	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
X.SPIN.	1,229	0.00	0.01	0.00	0.00	0.00	0.00	0.10	0.10	3.86	18.04	0.00
SPP2	1,229	0.09	0.27	0.01	0.04	0.01	0.00	6.30	6.30	12.33	248.40	0.01
X.STPR.	1,229	0.06	0.05	0.05	0.05	0.04	0.00	0.38	0.38	1.61	4.57	0.00
X.SUAV.	1,229	0.65	0.25	0.63	0.64	0.22	0.00	2.02	2.02	0.89	2.25	0.01
SYNE	1,229	0.30	0.15	0.28	0.29	0.12	0.00	1.34	1.34	1.25	3.89	0.00
THAC	1,229	0.09	0.06	0.08	0.08	0.06	0.00	0.38	0.38	1.16	2.10	0.00
X.THATD.	1,229	0.19	0.11	0.16	0.17	0.09	0.00	1.17	1.17	1.80	7.69	0.00
THVC	1,229	0.77	0.29	0.73	0.75	0.25	0.00	2.69	2.69	1.13	3.51	0.01
TIME	1,229	0.25	0.15	0.22	0.24	0.12	0.00	1.14	1.14	1.31	3.16	0.00
TO	1,229	1.85	0.45	1.85	1.84	0.42	0.61	3.36	2.75	0.27	0.31	0.01
TOBJ	1,229	0.34	0.14	0.33	0.33	0.12	0.00	0.97	0.97	0.80	1.20	0.00
TPP3	1,229	2.08	1.30	1.79	1.93	1.05	0.00	10.11	10.11	1.47	3.47	0.04
TSUB	1,229	0.08	0.08	0.06	0.07	0.06	0.00	0.60	0.60	1.93	5.80	0.00
VBD	1,229	3.98	1.57	3.77	3.88	1.56	0.57	10.59	10.02	0.60	0.09	0.04
VPRT	1,229	3.55	1.15	3.48	3.50	1.10	0.28	7.87	7.59	0.41	0.32	0.03
X.WHCL.	1,229	0.08	0.06	0.07	0.07	0.06	0.00	0.55	0.55	1.82	7.25	0.00
X.WHOBJ.	1,229	0.09	0.07	0.08	0.09	0.06	0.00	0.61	0.61	1.40	4.53	0.00
X.WHQU.	1,229	0.01	0.02	0.00	0.01	0.00	0.00	0.15	0.15	2.65	10.28	0.00
X.WHSUB.	1,229	0.25	0.13	0.24	0.24	0.12	0.00	0.87	0.87	0.95	2.15	0.00
X.WZPAST.	1,229	0.33	0.15	0.31	0.31	0.13	0.00	1.25	1.25	1.10	3.36	0.00
X.WZPRES.	1,229	0.29	0.13	0.27	0.28	0.12	0.00	1.14	1.14	1.04	2.72	0.00
XX0	1,229	0.92	0.27	0.90	0.91	0.25	0.00	3.13	3.13	0.78	4.36	0.01

Data 3: Descriptive statistical features in the BSLC

FACTORS	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
AWL	714	4.85	0.39	4.74	4.77	0.18	4.18	6.25	2.07	2.14	4.44	0.01
TTR	714	159.78	26.98	157.50	158.88	25.95	89.00	256.00	167.00	0.36	0.10	1.01
AMP	714	0.00	0.02	0.00	0.00	0.00	0.00	0.41	0.41	12.32	204.59	0.00
ANDC	714	0.65	0.24	0.61	0.63	0.18	0.00	1.98	1.98	1.19	3.39	0.01
X.BEMA.	714	0.69	0.32	0.69	0.69	0.28	0.00	1.97	1.97	0.15	0.37	0.01
X.BYPA.	714	0.16	0.11	0.14	0.15	0.09	0.00	1.08	1.08	1.82	8.46	0.00
CAUS	714	0.01	0.02	0.00	0.00	0.00	0.00	0.22	0.22	4.89	38.07	0.00
CONC	714	0.00	0.01	0.00	0.00	0.00	0.00	0.19	0.19	13.89	254.33	0.00
COND	714	0.34	0.23	0.33	0.33	0.24	0.00	1.21	1.21	0.42	0.11	0.01
CONJ	714	0.14	0.11	0.14	0.13	0.09	0.00	0.89	0.89	1.53	6.34	0.00
X.CONT.	714	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	18.78	351.01	0.00
DEMO	714	1.23	0.48	1.21	1.21	0.37	0.06	3.62	3.56	0.73	2.72	0.02
DEMP	714	0.09	0.08	0.08	0.08	0.09	0.00	0.50	0.50	1.27	2.80	0.00
DPAR	714	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11	24.66	632.43	0.00
DWNT	714	0.08	0.09	0.06	0.06	0.06	0.00	0.83	0.83	2.97	16.98	0.00
EMPH	714	0.12	0.10	0.11	0.11	0.07	0.00	0.72	0.72	1.83	5.90	0.00
EX	714	0.07	0.14	0.03	0.04	0.04	0.00	1.66	1.66	5.22	38.19	0.01
FPP1	714	0.12	0.12	0.10	0.11	0.12	0.00	0.76	0.76	1.28	2.33	0.00
GER	714	0.76	0.58	0.61	0.68	0.41	0.00	5.91	5.91	2.24	10.29	0.02
HDG	714	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	16.02	306.61	0.00
INPR	714	0.01	0.03	0.00	0.01	0.00	0.00	0.30	0.30	4.75	29.56	0.00
JJ	714	4.61	1.46	4.38	4.46	1.13	0.96	9.95	8.99	0.91	0.99	0.05
NEMD	714	0.29	0.26	0.25	0.26	0.30	0.00	1.77	1.77	1.05	1.60	0.01
NN	714	27.88	5.47	26.25	27.13	3.62	19.19	53.43	34.24	1.24	1.09	0.20
NOMZ	714	7.61	1.93	7.53	7.61	1.81	1.96	13.99	12.03	0.01	-0.04	0.07
OSUB	714	0.04	0.05	0.02	0.03	0.03	0.00	0.50	0.50	3.38	17.67	0.00
X.PASS.	714	1.22	0.49	1.25	1.23	0.43	0.03	3.17	3.14	-0.01	0.70	0.02
X.PASTP.	714	0.06	0.08	0.04	0.05	0.04	0.00	0.92	0.92	3.52	25.14	0.00
X.PEAS.	714	0.28	0.18	0.26	0.26	0.14	0.00	1.02	1.02	1.01	1.80	0.01
PHC	714	1.06	0.99	0.72	0.84	0.44	0.12	4.74	4.62	2.10	3.79	0.04
PIN	714	13.29	1.25	13.36	13.31	1.06	2.93	20.25	17.32	-0.50	7.67	0.05
X.PIRE.	714	0.33	0.21	0.34	0.33	0.18	0.00	1.42	1.42	0.78	2.30	0.01
PIT	714	0.43	0.26	0.40	0.41	0.22	0.00	2.61	2.61	1.75	9.36	0.01
PLACE	714	0.09	0.19	0.06	0.06	0.09	0.00	3.84	3.84	12.45	227.24	0.01
POMD	714	0.59	0.31	0.59	0.59	0.32	0.00	1.70	1.70	0.33	0.00	0.01
PRED	714	0.27	0.19	0.28	0.26	0.18	0.00	1.35	1.35	0.58	1.43	0.01
X.PRESP.	714	0.11	0.10	0.09	0.10	0.07	0.00	0.96	0.96	2.14	9.73	0.00
X.PRIV.	714	0.61	0.33	0.65	0.61	0.30	0.00	1.98	1.98	-0.03	-0.04	0.01
PRMD	714	0.15	0.22	0.09	0.10	0.10	0.00	1.99	1.99	3.87	20.35	0.01
X.PROD.	714	0.03	0.06	0.01	0.02	0.01	0.00	0.55	0.55	3.93	23.11	0.00
X.PUBV.	714	0.23	0.18	0.21	0.21	0.16	0.00	1.14	1.14	1.23	2.46	0.01
RB	714	0.74	0.36	0.74	0.73	0.31	0.00	4.42	4.42	1.69	14.93	0.01
X.SERE.	714	0.04	0.06	0.03	0.03	0.04	0.00	0.47	0.47	3.21	14.02	0.00
X.SMP.	714	0.03	0.04	0.01	0.02	0.01	0.00	0.43	0.43	3.95	23.27	0.00
X.SPAU.	714	0.07	0.07	0.06	0.06	0.06	0.00	0.53	0.53	2.26	10.18	0.00

FACTORS	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
X.SPIN.	714	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.07	9.12	101.45	0.00
SPP2	714	0.00	0.03	0.00	0.00	0.00	0.00	0.68	0.68	21.58	509.80	0.00
X.STPR.	714	0.09	0.10	0.08	0.08	0.09	0.00	0.86	0.86	2.31	9.62	0.00
X.SUAV.	714	0.36	0.24	0.34	0.34	0.24	0.00	1.68	1.68	0.78	1.69	0.01
SYNE	714	0.06	0.07	0.04	0.04	0.06	0.00	0.82	0.82	3.45	23.55	0.00
THAC	714	0.01	0.02	0.00	0.00	0.00	0.00	0.14	0.14	4.32	24.86	0.00
X.THATD.	714	0.06	0.06	0.05	0.05	0.06	0.00	0.49	0.49	2.14	8.07	0.00
THVC	714	0.09	0.09	0.07	0.08	0.10	0.00	0.48	0.48	1.22	1.47	0.00
TIME	714	0.08	0.14	0.06	0.06	0.07	0.00	2.23	2.23	9.32	124.95	0.01
TO	714	1.30	0.48	1.30	1.30	0.45	0.04	3.36	3.32	0.25	0.75	0.02
TOBJ	714	0.06	0.06	0.05	0.05	0.07	0.00	0.41	0.41	1.53	3.32	0.00
TPP3	714	0.23	0.18	0.18	0.20	0.12	0.00	1.53	1.53	2.48	10.38	0.01
TSUB	714	0.07	0.08	0.05	0.05	0.07	0.00	0.59	0.59	2.30	7.77	0.00
VBD	714	0.46	0.27	0.41	0.43	0.21	0.00	2.57	2.57	2.05	8.35	0.01
VPRT	714	3.61	1.11	3.85	3.76	0.73	0.20	6.19	5.99	-1.29	1.71	0.04
X.WHCL.	714	0.03	0.04	0.02	0.02	0.03	0.00	0.34	0.34	3.04	15.07	0.00
X.WHOBJ.	714	0.06	0.06	0.05	0.05	0.06	0.00	0.52	0.52	2.09	7.96	0.00
X.WHQU.	714	0.00	0.01	0.00	0.00	0.00	0.00	0.11	0.11	10.94	147.98	0.00
X.WHSUB.	714	0.20	0.15	0.19	0.18	0.15	0.00	1.20	1.20	1.22	3.76	0.01
X.WZPAST.	714	0.60	0.26	0.59	0.59	0.21	0.00	2.01	2.01	0.27	1.69	0.01
X.WZPRES.	714	0.51	0.33	0.44	0.46	0.18	0.00	2.50	2.50	1.96	5.51	0.01
XX0	714	0.38	0.22	0.40	0.37	0.21	0.00	1.35	1.35	0.17	0.34	0.01

Data 4: Factor scores for the new dimensions in the BLaRC (descriptive values)

FACTORS	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
1	1229	61.39	6.47	61.63	61.52	6.62	39.86	80.70	40.83	-0.22	-0.03	0.18
2	1229	66.74	6.69	67.21	66.92	6.83	42.33	85.77	43.44	-0.29	-0.04	0.19
3	1229	21.05	4.68	21.40	21.22	4.67	1.53	33.71	32.18	-0.37	0.10	0.13
4	1229	-17.71	2.36	-17.84	-17.78	2.34	-25.51	-8.96	16.55	0.29	0.17	0.07
5	1229	9.24	2.82	9.48	9.35	2.39	-0.48	18.98	19.47	-0.38	0.70	0.08
6	1229	8.74	1.37	8.73	8.75	1.41	4.04	12.55	8.51	-0.05	-0.11	0.04

Data 5: Factor scores for the new dimensions in the BSLC (descriptive values)

FACTORS	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
1	714	36.44	10.33	36.16	36.72	9.31	7.43	67.76	60.33	-0.17	0.18	0.39
2	714	45.10	11.06	44.94	45.25	9.88	13.52	79.94	66.42	-0.07	0.20	0.41
3	714	10.97	10.79	12.63	12.01	9.18	-23.91	34.21	58.12	-0.85	0.42	0.40
4	714	-10.18	5.31	-10.47	-10.68	4.14	-23.83	6.42	30.25	0.89	1.31	0.20
5	714	1.04	4.66	1.85	1.38	4.29	-15.14	11.72	26.85	-0.68	0.08	0.17
6	714	5.88	2.19	6.10	6.06	1.78	-0.93	11.56	12.49	-0.73	0.81	0.08

Data 6: Factor scores for predefined dimensions in the BLaRC (descriptive values)

DIM.	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
1	1,229	-10.27	3.42	-10.49	-10.42	3.04	-20.85	4.62	25.47	0.53	1.13	0.10
2	1,229	0.98	2.28	0.67	0.81	2.12	-4.59	9.84	14.43	0.80	0.82	0.07
3	1,229	5.75	1.81	5.70	5.71	1.72	-1.82	12.79	14.61	0.25	0.66	0.05
4	1,229	1.68	2.54	1.64	1.64	2.58	-6.11	10.92	17.03	0.23	0.22	0.07
5	1,229	3.67	1.47	3.61	3.63	1.41	-1.39	9.54	10.93	0.23	0.54	0.04
6	1,229	2.25	1.30	2.14	2.18	1.19	-2.00	8.31	10.31	0.72	1.53	0.04

Data 7: Factor scores for predefined dimensions in the BSLC (descriptive values)

DIM.	N	MEAN	SD	MEDIAN	TRIMMED	MAD	MIN	MAX	RANGE	SKEW	KURTOSIS	SE
1	714	-16.34	4.21	-15.42	-15.66	2.82	-29.61	-7.48	22.13	-1.45	1.89	0.16
2	714	-5.16	0.91	-5.22	-5.21	0.82	-7.22	-0.75	6.47	0.71	1.26	0.03
3	714	11.24	4.13	10.22	10.53	2.27	-1.32	26.64	27.96	1.81	3.63	0.15
4	714	-2.25	2.94	-1.90	-2.12	2.90	-8.77	5.57	14.34	-0.36	-0.53	0.11
5	714	1.07	1.50	1.17	1.13	1.25	-3.81	8.07	11.88	-0.09	1.53	0.06
6	714	-0.26	1.21	-0.27	-0.26	0.93	-3.35	5.12	8.47	0.34	1.97	0.05

SUPPLEMENTARY DOCUMENTS

Supplementary document 1: Legal Argumentation Schemes Annotation Guidelines

I. Argumentation detection and annotation in OVA+

Look for any statement looking like a Conclusion (C) and try to link it to its corresponding Premises (P). Be careful and don't be mistaken by any description of facts and its explanation. (See key concepts in argumentation below). Once you have annotated (C) linked to its (P) in OVA+, move on to 1 to select one the following Walton's argumentation schemes.

1. Argument relies on source's opinion or character: *Go to T1) Source-dependent arguments*
2. Argument does not depend on source's opinion or character: *Go to 3 below*
3. Argument is about classification or legal rules (Usually following the structure, if x happens, then…): *Go to T2) Schemes for applying rules to cases*
4. Argument focuses on the outcome for a course of action: *Go to T3) Reasoning*

II. Argument schemes identification

T1 – Source-dependent arguments

1. a. Argument relies on a source's character: 2
   b. Argument relies on source's opinion: 3
2. a. Argument relies on source's good character: Other/non specified
   b. Argument relies on source's bad character; conclusion draws on the lack of credibility of a person (which is attacked by the arguer): *Ad hominem* argument
3. a. Argument establishes a source's opinion; conclusion draws on the (change in the) commitment of a person to some values/beliefs/statements/goals: Argument from Commitment
   b. Argument is based on an existing opinion: Argument from Position to Know

T2 – Source-independent arguments: applying rules to cases

4. a. Conclusion is about the applicability of a (legal) rule: 5
   b. Conclusion draws on the application of a rule established/stated in the premises (i.e., if x is a when G, as now G takes place, x is a): Argument from verbal classification
5. a. Argument discusses/decides on the interpretation/applicability/narrowness of a specific law (might be an article, a section, a paragraph, etc in an act/regulation): Argument from an established rule
   b. Argument does not discusses/decides on the interpretation/applicability/narrowness of a specific law (might be an article, a section, a paragraph, etc in an act/regulation): 6
6. a. Argument refers to a characteristic sign. Premises draw on some particular (empirical) finding that, given the context, strongly leads to the conclusion (i.e., you saw some feathers in the window, you assume there was a bird there) – Argument from sign and abduction argument
   b. Argument is based on comparison: 7
7. a. Case at issue is similar to compared case(s): Argument from analogy
   b. Argument generalises from a particular instance: Argument from example

T3 – Reasoning

8. a. Conclusion promotes a positive outcome. The arguer makes use of hypotheses to determine what would have been the intention of the lawmaker in that case. In other words, the arguer creates a scenario of decision making in which the conclusion is necessary as the best alternative: Practical reasoning
   b. Conclusion prevents a negative outcome: 9
9. a. Conclusion is in favour of a course of action: Other/non specified
   b. Conclusion is against a course of action: 10
10. a. Chain of events would lead to bad outcome: Slippery Slope Argument

Action's direct outcome is good: Other/non specified