



---

**Universidad de Valladolid**  
**Facultad de Filosofía y Letras**

**Grado en Estudios Ingleses**

**A corpus-based study of impersonal grammatical  
patterns across two different text types of EU legal acts**

**Ana Mercedes Franco Calvo**

**Tutora: Raquel Fernández Fuertes**

**Departamento de Filología Inglesa**

**Curso: 2024-2025**

## Agradecimientos

*A Raquel, mi tutora de TFG, por guiarme, por enseñarme tantísimo y por tenerme tanta paciencia. No lo habría logrado sin tu ayuda.*

*Mamá, gracias por creer siempre en mí.*

*Javi, mi amor, gracias... “god only knows what I’d be without you”*

*Oli, Nico, lo conseguimos!!! Os adoro.*

## ABSTRACT

This dissertation is framed within the broad field of applied linguistics and focuses on the use of impersonal grammatical patterns—specifically, non-finite clauses and passive constructions—in two types of EU legal texts: regulations and directives. The main objective of this research is to determine the extent to which these impersonal structures can be considered distinctive and prevalent features of legal language. To achieve this goal, a corpus linguistics approach has been adopted as the most appropriate methodological framework for testing the initial hypothesis. The methodology combines the systematic and customized collection of data with a thorough and detailed analysis of the grammatical patterns under study, allowing for both quantitative and qualitative analyses. The results confirm the initial hypothesis: impersonal constructions are indeed consistent, distinctive, and prevalent features of legal discourse. Their widespread use highlights their essential role in promoting objectivity, neutrality, and general applicability—fundamental principles of legal communication.

**Keywords:** Impersonal structures, Grammatical pattern, Legal language, Text types, Frequency.

## RESUMEN

Este trabajo se enmarca en el campo de la lingüística aplicada y tiene como objetivo analizar el uso de estructuras gramaticales impersonales —en particular, cláusulas no finitas y construcciones pasivas— en dos tipos de textos jurídicos producidos por la Unión Europea: reglamentos y directivas. El objetivo principal es determinar si estas construcciones pueden considerarse rasgos distintivos y recurrentes del lenguaje jurídico. Para ello, se adopta un enfoque metodológico basado en la lingüística del corpus, que permite realizar un análisis tanto cuantitativo como cualitativo. La metodología empleada combina la recopilación sistemática de datos con un examen detallado de los patrones gramaticales seleccionados. Los resultados obtenidos respaldan la hipótesis inicial: las construcciones impersonales son elementos característicos, consistentes y predominantes en el discurso jurídico. Su uso frecuente contribuye a reforzar la objetividad, la neutralidad y la aplicabilidad general de los textos legales, principios fundamentales que rigen la comunicación jurídica en contextos institucionales y normativos.

**Palabras clave:** Estructuras impersonales, Patrón gramatical, Lenguaje jurídico, Tipos de texto, Frecuencia.



## Table of contents

|                                                                                       |    |
|---------------------------------------------------------------------------------------|----|
| 1. Introduction.....                                                                  | 1  |
| 2. Theoretical framework.....                                                         | 4  |
| 2.1. Language for specific purposes: the linguistic properties of legal language..... | 4  |
| 2.1.1. Characteristics of legal language.....                                         | 6  |
| 2.1.2. Types of European legislations .....                                           | 7  |
| 2.2. Corpus linguistics .....                                                         | 8  |
| 2.3. Impersonal structures .....                                                      | 9  |
| 2.3.1. Non-finite clauses .....                                                       | 10 |
| 2.3.2. Passive voice structures .....                                                 | 12 |
| 3. Methodology .....                                                                  | 15 |
| 3.1. Research question .....                                                          | 15 |
| 3.2. Research design .....                                                            | 17 |
| 3.3. Corpus compilation.....                                                          | 18 |
| 3.3.1. Data selection criteria .....                                                  | 18 |
| 3.3.2. Corpus size and characteristics.....                                           | 19 |
| 3.4. Corpus data processing and analytical procedures .....                           | 22 |
| 3.4.1. Tools and software .....                                                       | 22 |
| 3.4.2. Impersonal grammatical patterns and operational definitions.....               | 24 |
| 3.5. Final remarks .....                                                              | 27 |
| 4. Results.....                                                                       | 28 |
| 4.1. Overall approach to the corpus .....                                             | 28 |
| 4.2. Analysis of impersonal patterns across text types .....                          | 30 |
| 4.2.1. Non-finite clauses .....                                                       | 31 |
| 4.2.2. Passive structures.....                                                        | 32 |
| 5. Conclusion .....                                                                   | 33 |
| 6. Bibliographical references .....                                                   | 36 |
| 7. Appendix.....                                                                      | 38 |
| 7.1. Corpus 0.....                                                                    | 38 |
| 7.2. Compiled corpus.....                                                             | 38 |

## 1. Introduction

The present undergraduate dissertation is framed on the generic theoretical framework of applied linguistics to the study of English. This choice was made based on the significant opportunities this subject field offers to inquire into the study of grammatical, syntactic, or lexical patterns. Once the generic theoretical framework was established, it became imperative to determine the subfield of specialization. English for legal purposes, a specialized form of language for specific purposes, offers significant opportunities for the study of grammatical patterns of use. Furthermore, corpus linguistics offers substantial opportunities for developing practical research studies by creating customized corpora for a thorough search. However, the decision to research on grammatical patterns of use across legal language was mainly motivated by personal concerns. Firstly, a personal interest in the legal domain, given the consistent linkage of my professional career to the legal sector. Secondly, the recognition of legal language as a subject field particularly conducive to the analysis of linguistic features, characterized by an extensive corpus of texts and a distinct, purpose-driven language. For these reasons, this dissertation is focused on English for legal purposes and its specific use of language applied to the study of grammatical patterns. For this dissertation, a specific corpus has been meticulously custom-built to analyze the frequency of use of the most common impersonal structures in two different text types of EU legal acts: regulations and directives.

The *Grammar of Spoken and Written English* by Biber et al. (2021) serves as the principal theoretical framework sustaining this dissertation. As a comprehensive, corpus-based reference grammar, it systematically studies lexico-grammatical patterns across four major registers: conversation, fiction, news, and academic prose. However, legal language—particularly the language of legal acts—is not explicitly addressed within this framework.

This dissertation aims to contribute to the study of grammatical variation by examining specific impersonal grammatical patterns—namely, non-finite clauses and passive constructions—within two distinct types of EU legal texts: regulations and directives. Employing a corpus-based methodology, the study analyzes the frequency and distribution of these patterns to uncover register-specific tendencies in legal drafting.

The findings of this research have practical implications for both linguistic and legal domains. By shedding light on the structural features of legal language, the dissertation offers valuable insights for students and professionals in law, linguistics, and translation. Ultimately, it aspires to support clearer and more effective legal drafting and interpretation.

The regular and continuous process of lawmaking constitutes a vast and constantly expanding subject field of study, profoundly affecting the lives of all citizens. Consequently, the proper understanding of legal language is of outstanding importance, as it eases the adept management of many quotidian affairs, including the policy surrounding cookies within web searches, data protection policies and other everyday matters. It is noteworthy that all selected legal acts have a direct impact on everyday life, as they all establish a compliance legal framework across EU borders. To narrow down the scope of this study, the legal texts have been carefully selected, and the corpus is composed of binding legal acts concerning consumers' privacy and personal data protection policies, as well as free movement through EU borders.

This dissertation is organized into seven sections. Section 1 is the present introduction. Section 2 is devoted to establishing the generic theoretical framework, which conceptualizes the domains of knowledge from language for specific purposes, to corpus linguistics and impersonal structures, namely non-finite clauses and passive structures. Therefore, the theoretical background is twofold since it focuses on impersonal grammatical patterns across a specialized language, i.e. legal language. Section 3 focuses on the methodological side of the study, addressing all the required processes from a practical approach. These processes include the establishment of the selecting criteria for each legal text, the download, editing, compilation, as well as gathering all the housekeeping information to track each specific text. Section 4, dedicated to results, endeavors to accomplish qualitative analysis by applying precise tools of grammatical

analysis to provide a precise answer to the established research question. Section 5 comes to a consistent and coherent conclusion after a thorough analysis of the data gathered. Section 6 lists the bibliographic references upon which this dissertation is supported. Finally, section 7 holds the appendix, which includes access information and the link to the compiled corpus.



## 2. Theoretical framework

In the preceding four academic years, the various subject fields of study in the undergraduate program in English Studies at the Universidad de Valladolid (henceforth UVA) have furnished me with a substantial understanding of language use and the potential applications of applied linguistics. In recent years, the study of specialized language within legal and institutional contexts has gained prominence, particularly through the application of corpus linguistics. This dissertation explores the use of impersonal structures across two different text types of EU legislation; i.e. regulations and directives, employing a corpus-based approach to analyze their frequency and distribution. Impersonal structures can be defined as syntactic constructions that allow for the expression of accounts without explicit reference to an agent. Given the formal and authoritative nature of EU legal acts, impersonal structures seem to play a crucial role in keeping objectivity, precision, and depersonalization in legislative discourse. The central goal of this dissertation is to find out the extent to which these impersonal structures are a prevalent feature of the two text types of EU legal acts under study, namely regulations and directives.

To organize the theoretical framework, this section is structured into three main subsections. Firstly, language for specific purposes with a specific focus on legal language. Secondly, corpus linguistics as methodological approach for linguistic analysis. Thirdly, grammatical patterns for impersonal structures, i.e., non-finite clauses and passive structures can be considered the backbone of this dissertation.

### 2.1. Language for specific purposes: the linguistic properties of legal language

The term language for specific purposes refers to the study of linguistic features unique to specialized subject fields such as law, medicine, or business. Within this framework, English for legal purposes or legal language can be described as a sub-branch of language for specific purposes, focusing on the linguistic, grammatical, semantic, and discursive

elements of legal texts. As Paltridge et al. (2012) have noted, the study of language for specific purposes involves the teaching and analysis of English within a specific professional or academic domain, emphasizing its functional and contextualized use. Significantly, advancements in technology, driven by the rapid and widespread adoption of digital technologies, have had a profound impact on the expansion of numerous scientific fields, including medicine, engineering, mathematics, physics, law, and linguistics. These technological advancements have been particularly pronounced in the domain of language studies, particularly in applied linguistics, therefore leading to substantial developments in all specialized languages.

Traditionally, grammar studies have been approached from two main perspectives: prescriptive and descriptive. The prescriptive approach seeks to establish norms for how language should be used, adhering to established grammatical rules. In contrast, the descriptive approach aims to analyze how language is actually used, focusing on the relationships among clauses and smaller grammatical units. According to Biber et al. (2007) a common thread runs through these two approaches: both share a rigorous description of grammatical features and lexical characteristics of language. However, they fail to address the patterned use of grammatical features within large datasets of texts. In this context, recent developments in corpus linguistics have positioned corpora as essential tools in applied linguistic research, enabling the systematic analysis of grammatical patterns across large, authentic datasets.

In this context, the research into linguistic properties, grammatical or lexical patterns, specialized terminology, or discursive strategies has been enhanced by the vast opportunities afforded by corpus linguistics. Presently, this subfield is undergoing a period of rapid development, with researchers focusing on the creation of digital tools that ease the identification of linguistic patterns within vast data sets. This ongoing development has enabled the expansion of research studies in applied linguistics, as evidenced by this undergraduate dissertation that focuses on impersonal structures across two different text types of legal acts. It is also noteworthy that contemporary applied linguistics increasingly integrates both theoretical and empirical approaches. This dual orientation—combining grammatical theory with data-driven analysis—has significantly enriched the field, as reflected in the growing body of research focused on grammatical

patterns of use. As a result, applied linguistics has emerged as a leading discipline for the research of linguistic features within specialized languages.

### 2.1.1. Characteristics of legal language

The production of legal discourse is inextricably linked to the use of language, which serves as the primary medium for the drafting of legal texts. Nevertheless, the development of effective language must not be confined to the mere achievement of communicative goals; it should also be regarded as a repository of knowledge in itself. As Ainsworth (2014:19) asserts, "law cannot be understood without an appreciation of the nature of language." To that extent, in the context of EU legislation, linguistic choices may carry particular significance, potentially influencing the interpretation and application of legal provisions across multiple jurisdictions. Consequently, there is a necessity for in-depth research on specific grammatical patterns across legal text types to achieve a comprehensive understanding of the pervasive linguistic nature of legal language.

Furthermore, like all specialized languages, legal language is characterized by specific linguistic features, including nominalization, specific terminology or the use of complex syntactic structures, like impersonal constructions. The latter is of particular relevance to this dissertation, as impersonal structures contribute to the formal and abstract nature of legislative texts. Impersonality in legal discourse is a key feature that ensures objectivity and general applicability. Unlike ordinary language, which frequently uses personal pronouns and active voice constructions, legal language often employs impersonal constructions to avoid ambiguity and maintain neutrality. In EU legislative texts, impersonality manifests through various grammatical patterns, including non-finite clauses and passive constructions. These structures contribute to the authoritative tone and universality of legal provisions.

Accordingly, impersonal structures might be a prevalent feature in legal discourse, since they allow for the expression of general rules without assigning responsibility to specific individuals or entities. These structures seem to play a distinct role in framing legal obligations, recommendations, and prohibitions within EU legislative texts. A

classification of these impersonal structures reveals the distinction between non-finite clauses (examples 1-3) and passive voice (example 4), as illustrated by the following examples from the grammar by Biber et al. (2021)

1. *It enables the farmer to maintain uniform and near constant conditions in the house.* (ACAD) (687)
2. *Understanding how a planet generates and gets rid of its heat is essential if we are to understand how that planet works.* (ACAD) (201)
3. *Taken in the order shown they provide propulsive jets increasing mass flow and increasing jet velocity.* (ACAD) (202)
4. *[The proposal] was approved by the Project Coordinating Team.*” (ACAD) (473)

In the case of the non-finite clauses, three subtypes have been identified: *to* clauses (example 1), *-ing* clauses (example 2) and *-ed* clauses (example 3). These will be discussed in more detail in section 2.3.1. below, while passive constructions will be thoroughly described in section 2.3.2.

Another facet to take into consideration when describing legal language would be to establish the communicative situation in which legal discourse occurs, displaying different degrees of abstraction and formalization depending on the communicative contexts. According to Pearson's (1998) classification, legal acts could be regarded as an expert-to-non-expert communicative setting, given that the texts are redacted by experts addressing a non-expert audience, such as the general public and other legal actors, including judges or lawyers, and other professionals and entities subject to the laws.

### 2.1.2. Types of European legislations

The objectives outlined in the EU treaties are actualized through a variety of legal documents, some of which are binding, while others are not. Furthermore, the application of these acts may be limited to all EU countries or be restricted to a select few. Table 1 below, derived from the official website of the EU (EUR-LEX, n.d.), presents a list of EU legal acts. It is also noteworthy that the table exclusively displays the various subtypes

of regulations and directives, as this dissertation is focused exclusively on these two specific text types.

Table 1. Types of EU legal acts

| Text types of legal acts | Binding force   |
|--------------------------|-----------------|
| Regulation               | Legally binding |
| Delegated Regulation     | Legally binding |
| Implementing Regulation  | Legally binding |
| Directive                | Legally binding |
| Delegated Directive      | Legally binding |
| Implementing Directive   | Legally binding |

Since this dissertation centers exclusively on two types of legal acts—regulations and directives—the following definitions are focused on these types of legal texts. Regulations are binding legislative acts and directly applicable in all EU Member States. Whereas, directives set binding objectives to be achieved by the Member States to which they are addressed. The Member States are free to choose the manner they see fit to achieve those aims. Such disparities in the applicability of laws may indicate a discrepancy in the use of impersonal structures across different text types.

## 2.2. Corpus linguistics

Corpus linguistics has undergone significant quantitative development in the first decades of the 21st century, largely because of the rapid advances in computer, digital, and media technology. According to McEnery & Wilson (2001:449), “[t]he term corpus should properly only be applied to a well-organized collection of data, collected within the boundaries of a sampling frame designed to allow the exploration of a certain linguistic feature (or set of features) via the collected data.” Corpus linguistics provides a powerful empirical approach to studying linguistic patterns in large datasets of authentic texts. This methodological framework is particularly valuable in legal linguistics, as it allows for the systematic analysis of frequent grammatical patterns across different text types of legal discourse. With this definition of corpus linguistics in mind, this dissertation aims to enrich knowledge in the field of legal language by providing the frequency of use of specific impersonal grammatical patterns, namely non-finite clauses and passive

structures, in two different text types of EU legal acts, regulations and directives. Corpus-based approaches, on the other hand, offer objective insights into linguistic tendencies by examining large-scale text collections. By analyzing a corpus of EU legislative texts, this study aims to uncover how impersonal structures are employed across different legal text types, considering their syntactic features and pragmatic functions.

An extensive amount of literature has been published on the topic of corpus linguistics, underscoring the relevance and prominent role the subfield has gained in recent years. Accordingly, corpus linguistics can be described as a branch of applied linguistics that employs large, structured collections of texts—known as corpora—to conduct qualitative and systematic linguistic research. By analyzing authentic language data, this approach enables researchers to uncover grammatical patterns of use across registers and domains, including specialized languages, such as legal discourse. Hence, it can also be conceptualized as a linguistic discipline that aims to identify specific patterns of use associated with lexical or grammatical features. Additionally, corpus linguistics has been defined by Jones & Waller (2015:9) as "a methodology that allows researchers to make clearer and better descriptions of language which can inform or develop our understanding of language in use." Taking into account these various definitions, corpus linguistics can be regarded as a methodological approach that eases the study of the frequency of use of linguistic features in specialized languages, or in any other variation of language. This methodology leverages digital tools that streamline the data collection process, enabling the extraction of quantitative linguistic data from either purpose-built text datasets or existing corpus resources. These data then serve as the basis for conducting qualitative analyses, allowing researchers to infer patterns of language use within specific contexts or registers.

### 2.3. Impersonal structures

The so-called impersonal structures can be described as syntactic constructions that allow for the expression of accounts without explicit reference to an agent. These structures may serve multiple purposes in legal texts, including depersonalization, generalization, and an emphasis on legal authority.

### 2.3.1. Non-finite clauses

As previously stated, the theoretical framework of this dissertation is mostly supported by the grammar handbook written by Biber et al. (2021). This grammar is a descriptive work, providing an exhaustive description of grammatical phenomena in the English language. It is also empirical in nature, as it employs a corpus-based approach to analyze lexical or grammatical patterns in four different registers. The empirical nature of this grammar is reflected in its systematic analysis of how specific linguistic features differ in frequency, function or distribution across four different registers: conversation (CONV), fiction (FICT), news (NEWS), and academic prose (ACAD). This register-based approach underscores the importance of context in shaping grammatical choices and highlights the patterned nature of language use across different communicative domains.

Given that this dissertation focuses on one of the registers analyzed in Biber et al.'s grammar (2021)—namely, academic prose—this subsection aims to illustrate the prevalence of impersonal grammatical structures within this specific register. To this end, quantitative data and figures are presented to assess the frequency of such structures. The subsequent analysis seeks to extrapolate these findings in order to evaluate the frequency and distribution of these grammatical patterns across the two types of EU legal texts under study: regulations and directives.

Non-finite clauses are a type of subordinate clause in which the verb appears in its non-finite form, either in to infinitive, bare infinitive, past participle, or gerund form. There are three main types of non-finite clauses, each containing a different type of verb phrase: *to* clauses, *-ing* clauses and *-ed* clauses. As any subordinate clause, non-finite clauses play a syntactic role in the main sentence they are part of. They can occur either as subjects or in post-predicative positions, as shown in (5) and (6), respectively.

5. ***To attempt to forecast the effects of changing regulations on a national scale is very difficult.*** (ACAD: 654).
6. ***They [hope] to heighten their appreciation of nature.*** (ACAD: 698)

Figure 1 below shows the most common realization of major clauses considering their structure. An analysis of this table reveals a notable distinction in the versatility of different clause types, particularly in regard to their syntactic function. Specifically, *-ing* and *to* clauses exhibit a greater degree of flexibility compared to *-ed* clauses, which are constrained by a more limited array of syntactic roles, merely two: direct objects or adverbials. However, all three structures are more frequent in post-predicative positions.

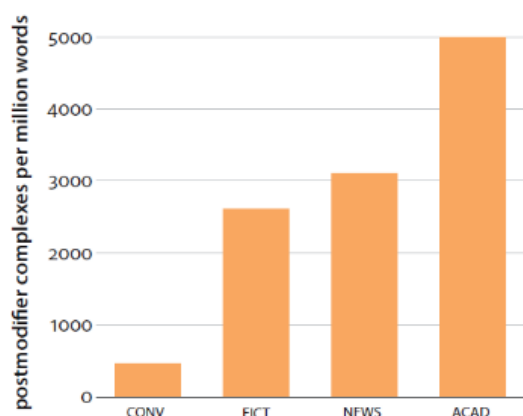
Figure 1. Common realizations of major clause elements

| Clause element       | Structural realization |                     |                   |
|----------------------|------------------------|---------------------|-------------------|
|                      | <i>ing</i> -clause     | <i>inf.</i> -clause | <i>ed</i> -clause |
| subject              | •                      | •                   |                   |
| verb phrase          |                        |                     |                   |
| subject predicative  | •                      | •                   |                   |
| direct object        | •                      | •                   | •                 |
| indirect object      |                        |                     |                   |
| prepositional object | •                      |                     |                   |
| object predicative   |                        | •                   |                   |
| adverbial            | •                      | •                   | •                 |

As Biber et al. (2021:636) have stated, non-finite clauses are considered, among other grammatical structures, as postmodifiers complexes which are a combination of post modifying elements. These structures are by far the most common in academic prose, as illustrated in figure 2 below.



Figure 2. Frequency of postmodifiers complexes across registers



### 2.3.2. Passive voice structures

Passive structures are characterized by more than a mere variation in the word order of a sentence; rather, they determine the functional structure of a clause entirely. Two forms of passive constructions exist: long passives, in which the agent is expressed in a by-phrase, and short passives, in which the agent of the action is left unknown. As Biber et al. (2021:475) remark in their grammar one of the primary functions of the passive voice “is that it demotes the agent of the verb (often the person doing the action of the verb) while giving topic status to the affected patient (the entity being acted on ...).” Therefore, the exclusion of the agent of the verb through the use of the passive voice may be a primary justification for the prevalence of passive voice structures in legal production. The use of this functional feature of the passive may aid in establishing objectivity and neutrality. The omission of the subject reinforces the mandatory nature of the provision without specifying who must submit the request. As it can be seen in the following example extracted from the corpus built up to support this dissertation:

7. *"the implementation and application of this Directive **should be made** in full compliance with the principles relating to the protection of personal data."*

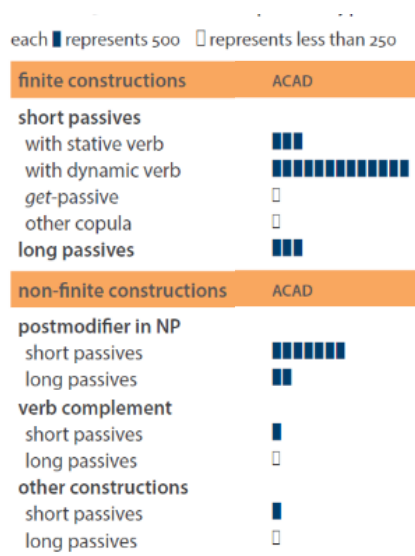
In addition, passive structures also demonstrate a high degree of complexity, given their capacity to be employed exclusively by transitive or ditransitive verbs. They are

typically formed with the auxiliary to be followed by a past participle verb form. As with non-finite clauses, passive voice constructions are most frequently used in academic prose than in any of the other registers examined at Biber's et al. grammar. Indeed, some passages of academic prose are completely written in the passive voice, as illustrated by the example (8) below:

8. *“Three communities on a brackish marsh of the Rhode River, a sub-estuary of the Chesapeake Bay, **were exposed** to elevated carbon dioxide concentrations for two growing seasons beginning in April 1987. The study site and experimental design **are described** in Curtis et al. (1989a). One community **was dominated** by the perennial carbon 4 grass spartina patens...” (ACAD:475).*

The prevalent use of passive structures among academic prose is also reinforced by the ensuing figure 3, which shows that the occurrence of long and short passives is higher in academic prose than in any other register.

Figure 3. Distribution of passive structures across registers



In essence, the theoretical framework of this dissertation is twofold. Firstly, it has tried to establish that *non-finite* clauses and passive structures are prevalent grammatical features associated with a specific register within Biber et al.'s (2021) grammar book,

namely academic prose. Secondly, it has established a range of grammatical patterns and verb types associated with impersonal structures. The depiction of these structures will serve as the foundation for the research on the specifically compiled corpus on which this undergraduate dissertation is grounded. In conclusion, this section endeavors to provide the most accurate and comprehensive theoretical framework upon which the methodological section will be sustained, to explore the extent to which impersonal grammatical patterns are prevalent in the two types of legal texts under analysis, namely EU regulations and directives.

### 3. Methodology

#### 3.1. Research question

As developed in subsection 2.1 of the theoretical framework about language for specific purposes, the present dissertation is grounded in the extensive subject field of applied linguistics, which encompasses disciplines such as general linguistics, corpus linguistics, discourse analysis, terminology and lexicography. Consequently, it was imperative to set up a narrow scope for the study to achieve the established goal of this dissertation. In this regard, it is noteworthy that specialized languages offer a suitable subject field for conducting linguistic or grammatical research. My personal interest in the domain of juridical sciences and law was crucial to establish legal language as the frame subfield to perform qualitative linguistic analysis. Following the establishment of the domain and subdomains of study, it was imperative to find the grammatical scope of the research to ease the development of a precise research question. Following exhaustive reflection, it was decided that the so-called impersonal structures and the study of their prevalent use across legal texts would constitute the narrowest scope of this dissertation.

Impersonal structures can be defined as syntactic constructions that allow for the expression of accounts without explicit reference to an agent (see the detailed account in subsection 2.3 about impersonal grammatical patterns). These structures may serve multiple purposes in legal texts, including depersonalization, generalization, and an emphasis on legal authority. This dissertation undertakes a comprehensive analysis of two distinct categories of impersonal structures: passive constructions and non-finite clauses. These structures will be studied across two different text types, namely EU regulations and directives. Regulations are defined as legally binding documents that are directly applicable in all Member States. Directives delineate objectives that Member States are obligated to achieve, while permitting the implementation of national measures. The selection of these two text types was based on their divergent degrees of prescriptiveness and interpretative flexibility. While both EU regulations and directives employ impersonal structures, it is acknowledged that some variations may exist due to their distinct legal functions. Regulations, given their direct applicability, may show a high

tendency to employ passive constructions. On the other hand, directives, which need national transposition, may exhibit a reduced frequency of these grammatical patterns.

Therefore, the objective of this research is to contribute to a more profound understanding of how linguistic varieties might reflect differing nuances within EU legal texts, depending for instance on the communicative purpose of each text type. In order to accomplish this aim with precision, the corpus has been meticulously compiled using a rigorous research design. The selection criteria for the compiled data have been carefully outlined to achieve both the necessities of the grammatical research, and theoretical quality standards. Thus, the corpus compiled for this dissertation comprises solely two distinct types of legal texts, encompassing a selection of official EU legislative texts from the EUR-Lex database, thereby ensuring the credibility of the source and the reliability of the texts.

According to Breeze (2013), there is a lack of research available on the frequency of grammatical patterns across legal texts. However, the article goes on to explore the potential for further research in this area, emphasizing the deontic nature of legal acts and the how certain grammatical structures may more effectively convey the normative sense of legal texts. It also highlights the distinctive nature of specific modal verb constructions, which are employed to convey either obligation, commission, or permission. Consequently, there appears to be significant scope for further research to ascertain data regarding the frequency of use of specific grammatical patterns within legal text types with greater precision.

Additionally, corpus-based studies have witnessed a notable expansion in recent years, largely due to the versatility this research approach offers in examining grammatical patterns of use. As Biber et al. (2006:4) have noted, “corpus-based studies are characterized by their empirical nature, analyzing actual patterns of use in natural samples of texts specifically compiled for this purpose.” Additionally, they use computer tools to conduct analyses and extract results. Most notably, they use a combination of quantitative and qualitative analysis methods to formulate conclusions. Consequently, it was determined that the corpus-based approach was the most suitable method for designing the corpus on which this dissertation is based. This approach was implemented to ensure the representativeness and linguistic consistency of the corpus on which this

dissertation delves to obtain results about the prevalent use of impersonal grammatical patterns across legal texts.

In light of these considerations, the following research question has been formulated: To what extent impersonal grammatical patterns, namely non-finite clauses and passive structures, show a prevalent frequency of use and distribution across two distinct text types of EU legislation: regulations and directives?

Therefore, the main objective of this dissertation is to contribute to a better understanding of how these impersonal structures are employed across different text types and considering their distinct communicative goal. However, it is noteworthy to remark that there are many grammatical patterns or linguistic features that could be subject to further research, such as the frequency of impersonal pronouns, existential *there* constructions, or the spread use of nominalized terms to depersonalize actions within legal texts.

### 3.2. Research design

This dissertation research employs a corpus-based analysis to investigate the frequency of use of impersonal structures across two different types of EU legal acts. As it has been previously stated, corpus-based analysis, as an interdisciplinary field, employs a combination of quantitative and qualitative research methods. This mixed-methods approach allows for both quantitative analyses through the compilation of a large collection of natural texts (corpus) and qualitative analysis of grammatical patterns of use within these texts. Consequently, the corpus-based methodology proves extensive suitability for conducting linguistic research, given its ability to integrate both quantitative linguistic data with qualitative analysis.

Additionally, advances in computer and digital technologies have enabled substantial improvements in corpus-based research methods. Today's advanced digital tools ease systematic comparative research by enabling the analysis of grammatical patterns and linguistic features across diverse forms of linguistic production. Besides, an expanded corpus can be achieved by incorporating more texts.

Corpus-based analysis has appeared as a valuable tool in the field of linguistic research, offering three key advantages: transparency, reproducibility, and efficiency. In essence, the corpus-based part of this dissertation is characterized by its transparency, which eases re-analysis and review. The Corpus 0 excel file with the database that has been created for the present research can be accessed via [TFG Corpus 0 Database Ana M. Franco Calvo](#) (see the Appendix section for further details). It is also imperative that the process be reproducible, since all processes and results obtained are replicable by any research that adheres to the methodological processes outlined in the following subsections. Consequently, it shows a high degree of efficiency, as it lends credibility to the findings and ensures the reliability of the results. Evidently, given the rationales, the corpus-based approach appears to be the best methodological technique for research design.

### 3.3. Corpus compilation

#### 3.3.1. Data selection criteria

Once the research question was defined and the research design outlined, the next step was to establish the selection criteria for the compilation of the texts, considering the required register. It could be stated that legal acts, more precisely EU regulations and directives, show a considerable degree of formality and technical language, or terminology. Consequently, the register can be regarded as formal and technical.

The selection criteria are thus founded on several parameters that align with the characteristics of formal registers. It is important to note that the selected texts are sourced from the Official Journal of the European Union webpage, EUR-LEX (n.d.). This distinguishes it from other corpora as it is an institutional corpus, exclusively composed of legal acts published by the mentioned publication. Secondly, the selected text types are in force regulations and directives enacted by the European Union. Thirdly, the length of the texts was also an important issue to consider, since the corpus aims to provide quantitative data about both types of legal texts.

Another essential aspect to consider when compiling the texts was the purpose or subject matter of the legal acts. Given the necessity to narrow the scope of the study, the selection of the texts was made on the grounds of the subject domains. In the case of regulations, the subjects encompassed are predominantly those related to science, information, education and culture, information technology, telecommunications and data processing, and the environment, consumers and health protection, among others. Similarly, the selected directives pertain to subjects such as information technology, telecommunications, data processing, and freedom of movement of people. In consideration of the comparative nature of the research, it is considered appropriate that the selection criteria for both text types address analogous subjects.

In summary, the empirical and methodological character of both corpus-based and grammatical research needed a meticulous description of procedures to ease the arduous task of compiling, collecting, and saving a substantial amount of text documents. Furthermore, it was imperative to establish a comprehensive and systematic set of patterns associated with impersonal structures. This was accomplished to facilitate the analytical study of grammatical patterns in the corpus employing the text analysis digital tool, Sketch Engine (2014). Consequently, the set of procedures established for the creation of this corpus has been determined by the necessities required to meet the two facets of the study: the corpus-based approach and grammatical pattern research. A more thorough elucidation of this information will be provided in the subsections that follow.

### 3.3.2. Corpus size and characteristics

The corpus is composed of 52 distinct legal texts, including 26 regulations and 26 directives. Collectively, these texts amount to 1.078.865 tokens and 919.381 words (22.721 sentences). Figure 4 below, extracted from the corpus, offers a comprehensive overview of the corpus size, encompassing the total counts of tokens, words, sentences, and documents. Additionally, it retrieves data regarding lexicon size and information concerning sub-corpora size. It is noteworthy that the total length of each sub-corpora appears to be relatively balanced, with directives accounting for 51,47% of the entire corpus, while regulations account for 48,53%.



Figure 4. Corpus information

Corpus TFG - AFC - UVA  user/anamercedes.franco/corpus\_tfg\_afc\_uva • created May 4, 2025 at 7:19:22 AM

Corpus TFG

COUNTS 

|           |           |
|-----------|-----------|
| Tokens    | 1,078,865 |
| Words     | 919,381   |
| Sentences | 22,721    |
| Documents | 52        |

SUBCORPUS SIZES

| Subcorpus   | Tokens  | %      |
|-------------|---------|--------|
| Directives  | 555,262 | 51.467 |
| Regulations | 523,603 | 48.533 |

LEXICON SIZES 

|                                                                                             |        |
|---------------------------------------------------------------------------------------------|--------|
| word?                                                                                       | 13,189 |
| tag                                                                                         | 59     |
| lempos?                                                                                     | 8,461  |
| pos                                                                                         | 9      |
| lemma                                                                                       | 7,638  |
| lempos_lc  | 8,069  |
| lemma_lc   | 7,154  |
| lc         | 11,188 |

The corpus is characterized by its monomodal, monolingual, and diachronic nature, comprising texts written exclusively in English and spanning from 2000 to 2024, although, since it is an open corpus, it could get expanded with more texts. As previously stated, all texts have been meticulously sourced from the Official Journal of the European Union, thus establishing its status as an institutional corpus. The credibility of the source ensures the integrity and reliability of the corpus.

The initial step in the compilation process entailed the creation of a Corpus 0 document, an Excel spreadsheet that houses all the pertinent information regarding each legal text. The Corpus 0 document is an essential component of the process, as it eases the storage of textual information, this database is divided into different working sheets; the first one is devoted to regulations, the second one to directives, the last one keeps information about the grammatical tags from Sketch Engine (2014). The sheets devoted to either regulations or directives contain the same fields, arranged in the same order to give the database consistency and coherence. This methodology eases the access to filter information about any specific document or field within Corpus 0. Consequently, it is imperative that this Excel document is maintained in a manner that ensures the storage of as much non-linguistic information as possible regarding each text, so it is easily traceable.

Once one legal text has been chosen following the established selection criteria, the first step is to complete the information about the text within Corpus 0. To identify each text document, a set of parameters has been established to name each document. An example on how these parameters have been used to name text documents compiled in the corpus appears in (9) below:

9.       D01\_17072000\_31EC\_EURLEX\_On electronic commerce  
           D: text type: R\_ Regulations & D\_ Directives  
           01: Correlative number of each text document within the corpus  
           17072000: DDMMYYYY; publication date  
           EURLEX: Source of the text  
           On electronic commerce: Abbreviation of the name of the legal act

Following the recordkeeping process of the text in the Corpus 0 database, the document must be stored. The ideal format for the effective preservation and management of documentation is the Notepad application from the Office Package suite of software programs since this tool eases the efficient edition of the texts. To guarantee the corpus's quality, the texts have undergone a thorough editing process. This process entailed the removal of sections considered to be less pertinent, which were determined to be non-contributory to the grammatical analysis, appendixes, blank spaces, and other non-linguistic information that was considered irrelevant for the present research. In the case that any of the selected laws have undergone an amendment, the modification is incorporated following the original text. It is noteworthy that all the procedures involved in the compiling process of the corpus are extremely tedious, laborious and time consuming, therefore, it is imperative to adhere to a systematic methodology.

### 3.4. Corpus data processing and analytical procedures

#### 3.4.1. Tools and software

The fast-speed development of digital technology has fostered a substantial increase in the development of digital tools designed for conducting linguistic research. However, most of these text-analysis tools are not freely available. However, according to the agreement of use between UVA and Sketch Engine (2014), all community members of the university are entitled to use the tool free of charge. This factor has also contributed to the selection of this software, in preference to other available options. Comprehensive details on accessing the compiled corpus are provided in the appendix (see Section 7).

Sketch Engine (2014) is a powerful corpus manager and text analysis software. It is designed to manage extensive text collections, empowering users such as lexicographers, linguists, translators, and language learners to conduct complex linguistic queries. Sketch Engine (2014) has been demonstrated to be a reliable tool due to its versatility. The software's intuitive design facilitates ease of use, and the software offers a user service that enables inquiries to optimize its use. Furthermore, Sketch Engine (2014) offers substantial storage capacity, with a maximum of 4.000.000 words, which can be expanded upon request.

This corpus manager tool supports multiple languages and scripts, offering functionalities such as word frequency analysis, keyword extraction, and co-occurrence analysis. Additionally, Sketch Engine (2014) offers advanced features such as word sketches, which provide a detailed usage overview of a word through concordance lines, and thesaurus generation, which identifies synonyms based on contextual similarity. The following is a brief description of each of the dashboard functionalities.

The Word Sketch search option returns outcomes for any lemma or word in the corpus. The use of a particular word may lead to identify other frequently co-occurring words that function either as post-modifiers or subjects, depending on the grammatical category of the searched item. This type of search option allows to lay the groundwork for further analysis of lexico-grammatical patterns. The figures below, taken from Sketch Engine

(2014), provide highly visual examples of the functionalities associated with the Word Sketch searching feature. Figure 5 shows the most common words that occur with the verb *to be*, including subjects, postmodifiers, prepositional phrases, and adjectives. Whereas Figure 6, displays the most common verbs occurring with the verb *to be*.

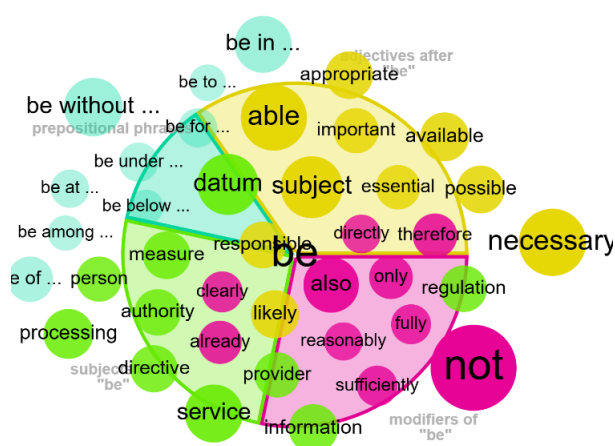
Figure 5. Modifiers of *to be*

Figure 6. Frequent verbs co-occurring with *to be*



The Word Sketch Difference tool enables the user to make a comparison between the collocations of two words. One potential application of the present tool is the comparison of two nouns within a given corpus to ascertain which one occurs with greater frequency, as well as the identification of all the words that are collocations of both nouns. The Thesaurus returns the synonyms and similar words of a term; within each synonymous word it appears a pull-down menu with the available options for each term. Additionally, the N-Gram search allows for the search of any multi-word expressions, up to 6 items, which entails the study of collocations and co-occurring words. In addition, the Keyword search option extracts individual key words (including tokens) or multi-word expressions.

Finally, the concordance search option provides examples of the use of any given word in the corpus. This feature eases the identification of complex syntactic patterns by enabling the selection of the advanced concordance search option for Corpus Query Language (CQL). This language is a specialized code employed in Sketch Engine (2014) to find intricate grammatical or lexical patterns. Given the dissertation's emphasis on complex grammatical patterns, the CQL search functionality has been proven to be the

most precise search option for finding impersonal structures within the corpus. Both the grammatical patterns and the CQL script established for each impersonal structure will be further developed in the ensuing subsection.

### 3.4.2. Impersonal grammatical patterns and operational definitions

Subsection 2.3 of the theoretical framework covers impersonal structures, such as non-finite clauses and passive voice constructions. These structures are examined across two distinct text types of EU legal acts, namely regulations and directives. Therefore, it is imperative to ascertain the grammatical patterns associated with these impersonal structures. Consequently, the grammar by Biber et al. (2021) has been consulted once more as the reference book to describe the grammatical patterns associated with this kind of impersonal structure. Furthermore, as Edger et al. (2020:25) have observed, “operational definitions are indispensable for the analysis of linguistic features or patterns in a corpus-based study.” Consequently, a fundamental aspect of the methodological process entails the establishment of precise and accurate operational definitions for each impersonal grammatical pattern to guarantee the extraction of precise and accurate data.

#### Non-finite clauses

There are three main types of non-finite clauses, each containing a different type of verb phrase: *to* clauses, *-ing* clauses and *-ed* clauses. They can occur either as subjects or in post-predicative positions. As pointed in the grammar by Biber et al. (2021:691) “Pattern 1: verb + *to*-clause is the most-common pattern for infinitive clauses.” Although no specific grammatical patterns have been identified for *-ed* clauses, they are described by their grammatical function either as direct object or adverbials. Therefore, *-ed* clauses typically occur either immediately after a verb or as a part of an adverbial phrase. The broadening of the scope of these grammatical patterns was deemed appropriate for the purpose of ascertaining their prevalence across the corpus and thus to achieve a thorough and comprehensible qualitative analysis. Example (10) displays the grammatical patterns and operational definitions for non-finite clauses.

# 10. Grammatical patterns for non-finite clauses:

## To clauses:

*Pattern 1:* verb + *to* clause

*Pattern 2:* verb + DP + *to* clause

*Pattern 3:* *to* clause

## –ing clauses:

*Pattern 4:* verb + –ing clause

*Pattern 5:* –ing clause

*Pattern 6:* –ing clause + DP + *to* clause

*Pattern 7:* –ing clause + past participle

## –ed clauses:

*Pattern 8:* verb + past participle

*Pattern 9:* past participle + verb

The following table displays the operational definitions for each pattern established in example (10) above, along with the CQL equivalent, necessary for the identification of non-finite clauses across the corpus.

Table 2. Description of operational definitions for non-finite clauses

| Non-finite clauses |                                     |                                                   |
|--------------------|-------------------------------------|---------------------------------------------------|
| To clauses         |                                     |                                                   |
|                    | Operational definition              | CQL tagging:                                      |
| <i>Pattern 1</i>   | verb + <i>to</i> clause             | [tag="VV.*"][tag="TO.*"][tag="VV.*"]              |
| <i>Pattern 2</i>   | verb +DP + <i>to</i> clause         | [tag="VV.*"][tag="NN.*"][tag="TO.*"][tag="VV.*"]  |
| <i>Pattern 3</i>   | <i>to</i> clause                    | [tag="TO.*"][tag="VV.*"]                          |
| –ing clauses       |                                     |                                                   |
|                    | Operational definition              | CQL tagging:                                      |
| <i>Pattern 4</i>   | verb + –ing clause                  | [tag="VV.*"][tag="VVG.*"]                         |
| <i>Pattern 5</i>   | –ing clause                         | [tag="VVG.*"]                                     |
| <i>Pattern 6</i>   | –ing clause + DP + <i>to</i> clause | [tag="VVG.*"][tag="NN.*"][tag="TO.*"][tag="VV.*"] |
| <i>Pattern 7</i>   | –ing clause + past participle       | [tag="VVG.*"][tag="VVN*"]                         |
| –ed clauses        |                                     |                                                   |
|                    | Operational definition              | CQL tagging:                                      |
| <i>Pattern 8</i>   | verb + past participle              | [tag="VV.*"][tag="VVN*"]                          |
| <i>Pattern 9</i>   | past participle + verb              | [tag="VVN*"][tag="VV.*"]                          |

## Passive voice structures

As discussed in figure 3 of subsection 2.3.2. devoted to passive structures, the most prevalent passive structures in academic prose are short passives or *be* passives. Additionally, finite constructions may be preceded by semi-modals and other auxiliaries. Finally, short passives in non-finite constructions are predominantly used to modify determiner phrases (DP). In summary, five distinct grammatical patterns have been identified for passive structures, as illustrated in (11). These include short or *be* passives, modal passives, perfect passives, and two different types of long passives, the latter type has a DP as head of the subject.

### 11. Grammatical patterns of passive voice constructions:

*Pattern 1P*: short or *be* passives: *be* + past participle

*Pattern 2P*: modal passives: modal + *be* + past participle

*Pattern 3P*: perfect passive: *have* + *been* + past participle.

*Pattern 4P*: long passives: *be/have/modal* + past participle + *to* clause

*Pattern 5P*: DP long passives: DP + *be/have/modal* + past participle + *to* clause

Table 4 retrieves the operational definition for each grammatical pattern associated with passive voice structures, as well as the CQL tagging necessary to perform the precise inquiry within the corpus.

Table 4. Description of operational definitions for passive structures

| Passive structures |                                                                                     |                                                                                  |
|--------------------|-------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Pattern No.        | Operational definition                                                              | CQL tagging:                                                                     |
| <i>Pattern 1P</i>  | short passive or <i>be</i> passive:<br><i>be</i> + past participle                  | [tag="VB.*"] [tag="VVN.*"]                                                       |
| <i>Pattern 2P</i>  | modal Passives:<br>modal + <i>be</i> + past participle                              | [tag="MD"] [tag="VB.*"] [tag="VVN.*"]                                            |
| <i>Pattern 3P</i>  | perfect passive<br><i>have</i> + <i>been</i> + past participle                      | [word="have has had"] [word="been"] [tag="VVN.*"]                                |
| <i>Pattern 4P</i>  | long passives:<br><i>be/have/modal</i> + past participle + <i>to</i> clause         | [tag="VH.* MD"]? [tag="VB.*"] [tag="VVN"] [tag="TO.*"] [tag="VV.*"]              |
| <i>Pattern 5P</i>  | DP long passives:<br>DP + <i>be/have/modal</i> + past participle + <i>to</i> clause | [tag="NN.*"] [tag="VH.* MD"]? [tag="VB.*"] [tag="VVN"] [tag="TO.*"] [tag="VV.*"] |

### 3.5. Final remarks

The process of documenting the methodology section of this dissertation has proven to be a laborious and arduous undertaking. One of the most challenging aspects of the study has been the compilation of a sufficient number of texts to achieve proper representativeness of the corpus. It is noteworthy that the corpus compiled shows sufficient size to perform a qualitative analysis, thus allowing the generalizability of the extracted results. As previously stated, all legal texts that constitute the corpus are available for download, saving or printing at no cost in all official languages of the European Union. These texts have been extracted from the Official Journal of the European Union webpage, EUR-LEX (n.d.), thus rendering the corpus an institutional one.

In summary, the field of applied linguistics is characterized by a multidisciplinary perspective on the study of language. This interdisciplinary approach provides researchers with a multifaceted framework for conducting and understanding research, thus contributing significantly to the advancement of knowledge in the language sciences. All in all, given the empirical and methodological character of this discipline, it has been imperative to develop a methodological section that provides thorough, descriptive, accurate, and precise information about the process of compiling a corpus to undertake linguistic research.



## 4. Results

### 4.1. Overall approach to the corpus

As delineated in subsection 3.3.2 Corpus size of the methodology section, the present corpus comprises 1.078.865 tokens and 919.381 words, amounting to a total of 22.721 sentences. A preliminary analysis of the corpus indicates that it holds the necessary scope and scale to undertake the research concerning impersonal grammatical patterns across two different types of EU legal acts, which constitutes the primary focus of the present undergraduate dissertation. It should be noted that, despite the use of American English as the standard language for this dissertation, the numerical data presented in this section are expressed in the metric system, using the European system of punctuation for numerical data (1.000.000,00 for one million, instead of 1,000,000.00). This discrepancy can be attributed to the use of the metric system in the Excel files from which the data have been extracted.

A first overall approach to the data is presented in tables 5 and 6. Table 5 offers an overview of the data in terms of the two text types (i.e., regulations and directives), providing the total amount of tokens. Table 6 shows the total number of impersonal structures (i.e., non-finite and passive structures) across the entire corpus.

Table 5. Distribution of the corpus by text type

|             | Percentage | Total number of tokens |
|-------------|------------|------------------------|
| Regulations | 49%        | 523.603                |
| Directives  | 51%        | 555.262                |
| Total       | 100%       | 1.078.865              |

A quantitative analysis reveals that the two text types under study account for a similar number of tokens. Regulations constituted 49% of the sample, while directives accounted for 51%.

Table 6. Total amount of impersonal structures by text type

| Text type  | <i>Non-finite</i> clauses | Passive structures |
|------------|---------------------------|--------------------|
| Regulation | 19.674                    | 8.894              |
| Directives | 18.865                    | 8.654              |
| Total      | 38.539                    | 17.548             |

Table 6 shows the total amount of impersonal structures across the two different text types, with a total of 38.539 occurrences for non-finite clauses, while passive structures account for a total amount of 17.548 occurrences. Both types of impersonal structures show an almost equal distribution across the two text types.

A second overall approach to the data is provided in tables 7 and 8 below, showing the total distribution of each type of non-finite clause and passive structures across the entire corpus, without distinguishing by text type.

Table 7. Distribution of non-finite clauses by type across the corpus

| Type of non-finite clause | Pattern No.      | No. of hits in the whole corpus | %     |
|---------------------------|------------------|---------------------------------|-------|
| <i>to</i> clauses         | <i>Pattern 1</i> | 2.266                           | 6%    |
|                           | <i>Pattern 2</i> | 471                             | 1%    |
|                           | <i>Pattern 3</i> | 13.130                          | 34%   |
| <i>-ing</i> clauses       | <i>Pattern 4</i> | 610                             | 1 %   |
|                           | <i>Pattern 5</i> | 20.751                          | 54%   |
|                           | <i>Pattern 6</i> | 193                             | 0,6 % |
|                           | <i>Pattern 7</i> | 109                             | 0,4 % |
| <i>-ed</i> clauses        | <i>Pattern 8</i> | 388                             | 1 %   |
|                           | <i>Pattern 9</i> | 621                             | 2 %   |
| Total                     |                  | 38.539                          | 100 % |

Table 7 shows the total distribution of each type of non-finite clause. Two of the nine patterns described exhibit a significantly higher frequency than the rest. These two patterns correspond with *-ing*, and *to* clauses, respectively. *Pattern 5* presents the highest frequency with 20.751 hits, while *pattern 3* stands as the second more frequent with 13.130 hits, both accounting for almost 88% of all non-finite clauses.

Table 8. Distribution of passive structures by type across the corpus

| Type of passive structure | <i>Pattern</i>    | No. of hits in the whole corpus | %    |
|---------------------------|-------------------|---------------------------------|------|
| <i>Be</i> passive         | <i>Pattern 1P</i> | 11.342                          | 65 % |
| Modal passive             | <i>Pattern 2P</i> | 4.071                           | 23 % |
| Perfect passive           | <i>Pattern 3P</i> | 659                             | 4 %  |
| Long passive              | <i>Pattern 4P</i> | 1.107                           | 6 %  |
| DP long passive           | <i>Pattern 5P</i> | 369                             | 2 %  |
| Totals                    |                   | 17.548                          | 100% |

Table 8 presents the distribution of passive structures across the corpus. It is noteworthy that two patterns stand out from the rest: *pattern 1P* corresponds to *be* passives with 11.342 hits, while *pattern 2P* corresponds to modal passives with 4.071 hits. Both subtypes of passives structures account for 88% of the total of passives of the entire corpus.

In the following subsection, the analysis focuses on the different patterns for each impersonal structure across the two text types, paying special attention to those patterns that account for a higher frequency.

#### 4.2. Analysis of impersonal patterns across text types

As established in section 3.1. the research question of this dissertation pretends to determine to what extent impersonal structures, (i.e., non-finite clauses and passive structures) can be considered prevalent grammatical patterns across two different text types of EU legal acts (i.e. regulations and directives). The analysis of following subsections intends to provide an answer to this question, by analyzing the frequency of use of each type of impersonal pattern across the two text types under study, i.e. regulations and directives.

#### 4.2.1. Non-finite clauses

As described in example (10) of subsection 3.4.2., three subtypes of non-finite clauses are distinguished: *to* clauses, *-ing* clauses and *-ed* clauses. Non-finite clauses are categorized into several patterns based on their syntactic structure. The frequency and distribution of these patterns across regulations and directives are presented in table 9.

Table 9. Distribution of non-finite clauses by operational definition and text type

| Non-finite clauses                         |                         |               |            |               |            |
|--------------------------------------------|-------------------------|---------------|------------|---------------|------------|
| Operational definition                     | Pattern No.             | Regulations   |            | Directives    |            |
|                                            |                         | No. of hits   | %          | No. of hits   | %          |
| verb + <i>to</i> clause                    | <i>Pattern 1</i>        | 1.105         | 6%         | 1.161         | 6%         |
| verb + DP + <i>to</i> clause               | <i>Pattern 2</i>        | 240           | 1%         | 231           | 1%         |
| <b><i>to</i> clause</b>                    | <b><i>Pattern 3</i></b> | <b>6.384</b>  | <b>32%</b> | <b>6.746</b>  | <b>36%</b> |
| verb + <i>-ing</i> clause                  | <i>Pattern 4</i>        | 347           | 2%         | 263           | 1%         |
| <b><i>-ing</i> clause</b>                  | <b><i>Pattern 5</i></b> | <b>10.853</b> | <b>55%</b> | <b>9.898</b>  | <b>52%</b> |
| <i>-ing</i> clause + DP + <i>to</i> clause | <i>Pattern 6</i>        | 131           | 1%         | 62            | 0%         |
| <i>-ing</i> clause + past participle       | <i>Pattern 7</i>        | 70            | 0%         | 39            | 0%         |
| verb + past participle                     | <i>Pattern 8</i>        | 208           | 1%         | 180           | 1%         |
| past participle + verb                     | <i>Pattern 9</i>        | 336           | 2%         | 285           | 2%         |
| <i>Totals</i>                              |                         | <b>19.674</b> | 100%       | <b>18.865</b> | 100%       |

Table 9 presents both the operational definitions established for each pattern and their distribution across each text type. Data show a similar distribution across text types of the two most prevalent patterns, i.e. *pattern 5* with a total of 55% in regulations, and 52% in directives; while *pattern 3* represents the 32% across regulations and 36% across directives. Both patterns are the most prevalent non-finite structures across the entire corpus, thus aligning with the results from table 7.

#### 4.2.2. Passive structures

Example (11) of subsection 3.4.2. described the grammatical patterns distinguished for passive voice structures. Passives have been categorized into several patterns; *be* passives are the most frequent of all passive patterns. Table 10 shows the frequency and distribution of these patterns across regulations and directives.

Table 10. Distribution of passive patterns by operational definition and text type

| Passive structures     |                          |              |            |              |            |
|------------------------|--------------------------|--------------|------------|--------------|------------|
| Operational definition | Pattern No.              | Regulations  |            | Directives   |            |
|                        |                          | No. of hits  | %          | No. of hits  | %          |
| <i>Be</i> passive      | <b><i>Pattern 1P</i></b> | <b>5.649</b> | <b>64%</b> | <b>5.693</b> | <b>66%</b> |
| Modal passive          | <b><i>Pattern 2P</i></b> | <b>2.102</b> | <b>24%</b> | <b>1.969</b> | <b>23%</b> |
| Perfect passive        | <i>Pattern 3P</i>        | 346          | 4%         | 313          | 4%         |
| Long passive           | <i>Pattern 4P</i>        | 592          | 7%         | 515          | 6%         |
| DP long passive        | <i>Pattern 5P</i>        | 205          | 2%         | 164          | 2%         |
|                        | <i>Totals</i>            | 8.894        | 100%       | 8.654        | 100%       |

Table 10 shows that *be* passives are the most frequent pattern across both text types accounting for 64% across regulations and 66% across directives. The data in table 10 also show a consistent and similar distribution of *be* passives in both text types.

The analysis of impersonal grammatical structures across the corpus reveals a consistent and systematic use of non-finite clauses and passive constructions across both regulations and directives. Despite standing for a relatively modest proportion of the overall corpus, these structures show a high degree of regularity in their distribution. Specifically, two patterns dominate each category: *-ing* clauses and *to* clauses account for 88% of all non-finite clauses, while *be* passives and modal passives comprise 88% of all passive structures. This uniformity suggests a standardized linguistic property in EU legal drafting, aimed at keeping objectivity, formality, and general applicability. The findings underscore the functional role of impersonal constructions in legal discourse, reinforcing their importance in achieving the clarity and neutrality characteristic of legislative language.

## 5. Conclusion

This undergraduate dissertation presents a corpus-based research of impersonal structures (i.e. non-finite clauses and passive structures) across two different text types of EU legal acts (i.e. regulations and directives). Legal language seems rather suitable to perform research about grammatical patterns of use, since it exhibits distinctive linguistic features that serve specific communicative functions. Furthermore, it appears that impersonal structures play a pivotal role in shaping the formal, objective, and authoritative tone that characterizes legal discourse.

To conduct this research, a custom corpus of nearly one million tokens was compiled, which exclusively comprised EU regulations and directives. The selection of these two text types was made on the basis of their distinct legal functions. Regulations are binding in their entirety and directly applicable in all member states; in contrast, directives set out goals that member states must achieve, thus allowing for flexibility in national implementation. However, despite these functional differences, both text types are expected to adhere to the conventions of legal drafting, including the use of impersonal language.

The corpus was analyzed using Sketch Engine (2014), a sophisticated corpus analysis tool that enables the identification of grammatical patterns, as well as the extraction of data regarding frequency of use and distribution of impersonal grammatical patterns. Following the compilation of the corpus, a qualitative analysis was conducted to assess the prevalence and use of impersonal structures across the two different types of legal texts. The results of the study showed that both regulations and directives employ impersonal structures with remarkable consistency in both frequency and distribution. This finding serves to underscore their integral role in legal drafting.

The *Longman Grammar of Spoken and Written English* (Biber et al., 2021) serves as the primary reference source for describing impersonal structures. This grammar, based on a large multi-register corpus, provides a robust framework for analyzing grammatical variation across different text types. By adopting a similar corpus-based methodology, this dissertation aimed to uncover the frequency, distribution, and function of impersonal structures in EU legal texts.

The findings of the research confirmed the initial hypothesis: impersonal structures are not only prevalent but also systematically employed across both regulations and directives. Among non-finite clauses, *-ing* clauses and *to* clauses emerged as the most frequently used forms, together accounting for 88% of all non-finite clauses identified in the corpus. These structures often serve to express obligations, conditions, or procedures in a way that avoids specifying an agent, thereby contributing to the generality and neutrality of the legal text.

Similarly, passive voice structures were found to be highly frequent, with *be* passives and modal passives comprising 88% of all passive forms. These structures allow for the depiction of actions and obligations without attributing them to a specific agent, which is particularly useful in legal contexts where the focus is on the action or requirement itself rather than on who performs it. The consistent use of these structures across both text types suggests a shared stylistic and functional approach in EU legal drafting, aimed at ensuring clarity, consistency, and legal precision.

The uniformity in the use of impersonal structures highlights their strategic role in promoting objectivity, neutrality, and general applicability—core principles of legal discourse. By minimizing personal reference and emphasizing actions, conditions, and obligations, these constructions help to create a tone of impartial authority. This is particularly important in legal texts, which must be interpreted and applied consistently across different jurisdictions and contexts.

Additionally, the findings may contribute to a deeper understanding of the linguistic mechanisms that support the clarity and formal tone of legal language. They show how grammatical choices are not merely stylistic but serve important communicative functions. In this sense, impersonal structures can be seen as a key component of the rhetorical strategy of legal discourse, enabling to convey complex information in a precise manner.

Despite the valuable insights gained, this dissertation is not without limitations. The corpus, while substantial, was limited to two types of EU legal acts and did not include other types of legal production such as case law, treaties, or national legislation. Additionally, the analysis focused primarily on frequency and distribution, with less emphasis on contextual or functional variation. Future research could expand the scope

of the corpus to include a wider range of legal texts and explore how impersonal structures function in different types of legal texts or in multilingual legal drafting contexts.

Another possible direction for future research would be a comparative study between EU legal texts and those from other legal systems, such as common law jurisdictions. Such a study could reveal whether the patterns observed in this dissertation are unique to EU legal drafting or reflect broader tendencies in legal language. Furthermore, incorporating a diachronic perspective could shed light on how the use of impersonal structures has evolved over time in response to changes in legal practice or drafting conventions.

In conclusion, this dissertation has proved that impersonal grammatical patterns—particularly non-finite clauses and passive structures—are a defining feature of EU legal language. Their consistent and strategic use across regulations and directives underscores their importance in achieving the communicative goals of legal discourse, enabling to convey complex information promoting objectivity, neutrality, and general applicability. These structures help to ensure that legal texts are clear, authoritative, and adaptable to diverse legal and cultural contexts.



## 6. Bibliographical references

Ainsworth, J. (2014). Lost in translation? Linguistic diversity and the elusive quest for plain meaning in the law. *The Ashgate Handbook of Legal Translation*. Seattle University School of Law.

Biber, et al. (2021). *Grammar of Spoken and Written English*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.232>

Biber, et al. (2007). *Corpus linguistics: investigating language structure and use*. Cambridge University Press.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. 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> John Benjamins Publishing Company

Egbert, et al. (2020). *Doing linguistics with a corpus: methodological considerations for the everyday user*. Cambridge University Press.

EUR-LEX. (n.d.) *Access to European Union law*. Available at: <https://eur-lex.europa.eu/collection/eu-law/legal-acts/recent.html/>

Kilgariff, et al. (2014). *The Sketch Engine: Ten years on Lexicography I*. Available at: <http://www.sketchengine.eu/>

McEnery, T., & Wilson, A. (2001). *Corpus Linguistics: an introduction*. Edinburgh University Press.

Official Journal of the European Union (OJ) (nd). Available at: <https://eur-lex.europa.eu/oj/direct-access.html/>

Paltridge, B., & Starfield, S. (2013). *The handbook of English for specific purposes*. Wiley-Blackwell.

Pearson, J. (1998). *Terms in context*. John Benjamins Publishing Company.

## 7. Appendix

### 7.1. Corpus 0

The Corpus 0 document, which comprises all non-linguistic information regarding each text included within the corpus, is available on the following link.

Link to the Corpus 0 Database:

[TFG Corpus 0 Database Ana M. Franco Calvo](#)

### 7.2. Compiled corpus

The compiled corpus is available online at Sketch Engine (2014) for members of the Universidad de Valladolid via the following link, which requires logging in using institutional access.

Link to Sketch Engine

[Sketch Engine](#)

Once you have logged in, the corpus will be available within the section “Shared with me” under the name:

Corpus TFG - AFC – UVA

For more information about access to the corpus please contact the author of this dissertation or the Sketch Engine Support Team.