



---

# **Universidad de Valladolid**

FACULTAD de FILOSOFÍA Y LETRAS  
DEPARTAMENTO de FILOLOGÍA INGLESA  
Grado en Estudios Ingleses

## TRABAJO DE FIN DE GRADO

A Corpus Analysis of English Phraseology in Court  
Judgments of the European Union

Eva Luis Álvarez

03/07/2019

Tutor: Belén López Arroyo

2018/2019



## ABSTRACT

Specialized phraseology is a key field for the study of LSPs, as it has been demonstrated that meaning is achieved from word combinations and not from words in isolation. Phraseology has been widely studied by several authors—in particular, Roberts (1994), Gläser (1994) and Timmis (2015)—and only one author has focused on the specialized phraseology of legal language: Biel (2014). However, there are not many studies about the different subfields of legal language, specially about the phraseology of the court of justice. Thus, we aim at extending the knowledge about the legal phraseology of judgments from a functional approach. To this effect, we will compile a corpus of English court judgments of the EU, and we will analyze and classify their lexical chunks according to Biel's work (2014). The results will prove that there is a specific phraseology for the genre of judgments, and that corpus linguistics is an effective tool for its study. Finally, we will conclude that the analysis of the specialized phraseology of judgments is useful for their adequate understanding, writing and translation.

Keywords: corpus analysis, judgment, legal language, lexical chunk, LSP, specialized phraseology.

## RESUMEN

La fraseología especializada es un campo clave para el estudio de las LFE, ya que se ha demostrado que el significado se obtiene a partir de combinaciones de palabras y no de palabras aisladas. La fraseología ha sido ampliamente estudiada por varios autores, en particular, Roberts (1994), Gläser (1994) y Timmis (2015), y sólo un autor se ha centrado en la fraseología especializada del lenguaje jurídico: Biel (2014). Sin embargo, no hay muchos estudios sobre los diferentes subcampos del lenguaje jurídico, especialmente sobre la fraseología del tribunal de justicia. Por lo tanto, nuestro objetivo es ampliar el conocimiento sobre la fraseología jurídica de las sentencias desde un enfoque funcional. Para ello, compilaremos un corpus de sentencias judiciales inglesas de la UE, y analizaremos y clasificaremos sus *lexical chunks* según el trabajo de Biel (2014). Los resultados demostrarán que existe una fraseología específica para el género de las sentencias y que la lingüística de corpus es una herramienta eficaz para su estudio. Finalmente, concluiremos que el análisis de la fraseología especializada de las sentencias es útil para su adecuada comprensión, redacción y traducción.

Palabras clave: análisis de corpus, sentencia, lexical chunk, LFE, fraseología especializada.

## INDEX

ABSTRACT .....	1
1. Introduction .....	3
1.1. Objectives .....	4
1.2. Sections outline.....	5
2. Languages for Specific Purposes .....	6
2.1. Legal language.....	7
2.1.1. Legal language in English judgments .....	8
3. Phraseology .....	9
4. Methodology .....	12
4.1. Corpus Linguistics.....	12
4.2. Corpus description.....	13
4.3. Criteria used for the compilation of the corpus .....	14
4.3.1. Analysis of the corpus: representativity.....	15
4.3.1.1. Qualitative representativeness .....	15
4.3.2. Quantitative representativeness .....	16
4.4. Type of study .....	17
4.5. Selection of candidate terms.....	18
5. Results and classification .....	21
6. Conclusion.....	29
7. Bibliography.....	31

## 1. Introduction

This final project aims to provide an analysis of the legal phraseology contained in English court judgments of the European Union from a functional approach because, as characterized by Allen (2007: 254), “functionalism holds that linguistic structures can only be understood and explained with reference to the semantic and communicative functions of language, whose primary function is to be a vehicle for social interaction among human beings”. Moreover, one of the reasons why this work is focused on phraseology is the growing interest in the study of word combinations that is arising nowadays because, in fact, “words do not exist in isolation but cluster together in particular ways to make larger meaningful units of language” (Jones and Waller 2015: 84).

This study has been developed thanks to the knowledge acquired in the subjects of linguistics and specialized translation—more specifically, legal translation—of the degree. Within linguistics, this paper combines two fields that are commonly interconnected and that are increasingly acquiring relevance in research studies within the field of Languages for Specific Purposes (LSP): corpus linguistics as a tool for the analysis of specialized languages, and specialized phraseology as the way to study the data contained in the said corpus. Besides, the combination of these fields can serve as the basis for translation tasks, or even for the creation of translation tools such as term bases, terminological records, glossaries and dictionaries.

As we are going to make use of corpus linguistics for the analysis of the phraseology of English judgments, we can justify the choice of the subject “New Technologies Applied to English Studies”, since the use of technology allows us to compile and manage corpora, and to create translation memories or glossaries, among many other uses. Technology helps to gather and produce data, but when it comes to analyze all these data, it must be done through the use of other disciplines: in this case, corpus linguistics.

Therefore, this project aims to analyze the phraseology of English judgments and include it within the different classifications of specialized phraseology that have been developed throughout the years—particularly within the classification of phraseology in legal language developed by Biel (2014). This analysis can be relevant for the writing,

understanding and translation of judgments. It can serve as a tool for learner judges to know the specific phraseology of judgments and write their decisions properly because, as noted by López and Moreno, “one also needs to know the contextualized use of the terms in order to produce technically appropriate texts” (López and Moreno 2019: 32-59). Besides, even though according to Pearson’s classification of specialized texts (1998) judgments belong to an expert-to-expert communicative setting, this type of text is addressed to laymen and, thus, they may not understand the content due to their lower level of expertise. Indeed, related to the previous idea, it may also facilitate the task of intralinguistic translators, since they will get to know some specific word combinations that they will have to remodel in order to suit a non-expert or semi-expert target audience (Gotti 2016: 16).

### **1.1. Objectives**

As previously mentioned, the aim of this project is to carry out a functional analysis of the legal phraseology of English court judgments of the European Union in order to facilitate the understanding, composition and translation of this type of texts. Thus, our objectives are the following:

- To prove that there is a specific phraseology for the different subfields within legal language, especially for judgments.
- To classify the phraseology extracted from the corpus according to a classification of the phraseology in legal language.
- To demonstrate the usefulness of the compilation and exploitation of monolingual corpora as a way to approach the peculiarities of specialized languages and to create tools such as glossaries and parallel texts, among others.

If the aforementioned objectives are accomplished through this project and our hypotheses are proved, this study would serve, in future researches, as an approach to the creation of a monolingual glossary or a legal writing guide including the most frequent lexical chunks within English judgments and their phraseological classification. As a result, we will contribute to the facilitation of the aforementioned tasks, and to make the particular phraseology of these texts and, more generally, their language more accessible and easier to understand.

## 1.2. Sections outline

*Section two* presents some theoretical background about Languages for Specific Purposes. First, it gives a definition of LSPs, noting that they are used in various communicative settings and, thus, supporting the necessity of studying specialized languages in order to facilitate their understanding. This section also tries to clarify the boundaries between LSPs and LGP: it gives a definition of LGP and supports the theory of LSPs as “LGP-based” languages (Di Prisco 2018). Within this section there are two subsections: on the one hand, there is an introduction to legal language and the preciseness of its nature (section 2.1.), and, on the other hand, a classification of judgments in terms of genre and discourse, and a description of their distinctive features according to different authors.

*Section three* defines the concept of phraseology and presents the different classifications of phraseology that have been made throughout the years. Then, it presents a classification that is specific for legal language, which is the one we are going to use for our analysis of the phraseology of English court judgments.

*Section four* is dedicated to the methodology that we have followed for the compilation and exploitation of our corpus. This section has different subsections: the first one (section 4.1.) provides the theoretical framework about Corpus Linguistics, to later describe what is a corpus and the characteristics of our particular corpus (4.2.), and the criteria used for its compilation (4.3.). At the same time, section 4.3. is divided in one more subsection that deals with corpus representativity, defining it and showing the representative analysis that we have carried out in order to prove that our corpus was representative of the field of judgments. Then, there is another subsection for the description of the type of study (4.4.) and another one focused on the process of selection of candidate terms (4.5.).

*Section five* provides the results of the classification of the phraseology contained in the English court judgments that constitute our corpus. Finally, *section six* aims to present the conclusions that we have reached through the development of this study.

## 2. Languages for Specific Purposes

As the aim of this work is to describe the phraseology of English judgments, we deem convenient to provide its theoretical framework. The language of English judgments is part of the legal language and, therefore, it is a Language for Specific Purposes, so it is necessary to give a definition of LSPs and its peculiarities. Bowker and Pearson describe LSP as “the language that is used to discuss specialized fields of knowledge” (2002: 25). Nevertheless, taking into account this definition, it may be thought that LSPs are only used by specialists in different fields, however, they are also used by semi-experts and non-experts. Therefore, as there are different communicative situations and registers in which LSPs are used, it is very important to study specialized languages, since they can be used by any individual regardless of their level of expertise.

On the other hand, there have been many theories trying to explain the relationship between LSPs and Language for General Purposes (LGP): LSPs as elements of general language, LGP as a subset of LSP, LGP and LSP as equal, yet distinct, etc. Thus, even though LSPs and LGP may seem, by their definitions, opposite languages, the reality is that the limit between them is not clearly delimited. Indeed, this situation is caused by the fact that specialized languages “are based in the phonetic, grammatical and lexical structures of natural language and they are characterized by the creation of a unique terminology, as well as the orientation to certain kinds of grammatical and discursive structures” (Calvi 2015: 15). For that reason, we must not fall into the error of considering LSPs as the opposite of LGP, which is the everyday language used to talk about usual, yet informal, matters, because LSP “uses part of general language, such as grammatical constructions and some general-language words” and, therefore, it is LGP-based (Di Prisco 2018).

Therefore, the main difference between LSPs and LGP relies in the fact that, among other distinguishing aspects, LSPs have a specific, unique terminology and phraseology, and it is necessary to have a high command of them to create and understand any specialized text. In this work, we are going to focus on the specialized language of law, which, as many others LSPs, requires much research in order to facilitate the writing of legal documents by lawyers, their translation by specialized translators, and their understanding by any citizen.



## 2.1. Legal language

Broadly speaking, legal language is the tool used for the creation of law. As it is the vehicle for the dissemination of legal documents, it is a very formal language that tends to include technical and sub-technical terms, always ensuring reliability and precision. Borja (2000) states that legal language is the one used in the relations in which the public power intervenes, whether in the manifestations of this power (legislative, executive or judicial) made by the citizen, or in the communications of the citizens addressed to any type of institution. Therefore, as legal documents can be written by both legal professionals and citizens, it is necessary to study this specialized language in order to facilitate both the creation and comprehension of these texts without taking into account, as mentioned in the previous section, the level of expertise of the sender or the receiver.

However, in matters of the creation of legal documents, the writer must be more precise than in other types of documents, since the reading of the text cannot lead to misinterpretation. Borja (2007), in her description of the legal writing style, makes this clear by noting that the lawyer must “take pains to say precisely what he means, no more and no less—not only so that a person reading in good faith can understand, but so that a person reading in bad faith cannot misunderstand” (Borja 2007: 123). For that reason, before starting to write any kind of legal document, it is important to study or analyze the nature of legal texts in general, and later getting familiar with their vocabulary, starting always from the smallest units of meaning—which in most of the cases constitute bigger word combinations or multiword units. Thus, following these steps, it is less probable to fall into the error of using ambiguous terms and expressions that cause, as previously mentioned, misinterpretations.

Furthermore, given all that, there are some groups who believe that the language of the law is obscure and suggest the creation of legal documents written in plain language—that is, clarifying and simplifying legal language so that it can be understood by ordinary people who are part of the legal system. However, there are those who oppose to this approach, basing their position on the fact “that technical accuracy is an essential prerequisite of good justice, and that if linguistic precision is watered down to suit the demands of an uncomprehending majority, legal certainty will all but disappear” (Alcaraz

and Huges 2014: 5). In this paper, we try to support the approach of the second group, since we consider that the language of the law constitutes a specialized type of discourse that is needed for the communication of legal affairs in an expert-to-expert communicative setting (*vid supra*: section 1) and that deserves its analysis and study.

### **2.1.1. Legal language in English judgments**

In order to classify judgments in terms of genre, we are going to follow Borja's classification of English and Spanish legal texts (Borja 2000: 133-134). She differentiates six categories: *textos normativos*, *textos judiciales*, *jurisprudencia*, *obras de referencia*, *textos doctrinales* and *textos de aplicación del derecho (públicos y privados)*. Judgments (as well as writs of summons, pleadings or appeals, among others) fall into two categories: *textos judiciales* and *jurisprudencia*. Thus, their discourse situation if we consider them *textos judiciales* is the following:

- Sender: Administration of justice / citizens
- Receiver: Citizens / administration of justice
- Tone: Very formal / formal
- Mood: Written to be read
- Purpose: Any kind of communication between the administration of justice and the citizens

On the other hand, if we take their classification as *jurisprudencia*, the tone and the mood remain the same, but the situation changes in the rest of aspects:

- Sender: *Órganos superiores de justicia*
- Receiver: Citizens
- Purpose: As source of law, to serve as a basis for judges in their decisions

It is important to consider that, as we are analyzing court judgments of the European Union, the administration of justice or *órganos superiores de justicia* in this case is the Court of Justice of the European Union. Besides, as regards the text typology in terms of communicative intention, the main text type in both cases is the instructive, even though they may also be argumentative and expository.

The genre of judgments has peculiarities in its language that differentiate it from any other subfields of law. In fact, there are already works which deal with judgments as a genre: it is the case of Alcaraz and Huges (2014: 112-113), who describe judgments in terms of form and note that “they may be produced *extempore* (in an improvised way), but modern practice is for judges to write their opinions and read them out in open court”. They also make a description of their structure, identifying different parts: the introduction—making reference to the nature of the case and to the parties involved—a section called “the facts”—which presents the main matters of fact assumed, admitted or proved—and a series of separate sections aimed at commenting and describing the parties’ disagreements based on the facts and the law.

On the other hand, taking into account the results of Ruiz’s study (Ruiz 2013: 86) which are obtained from a comparable corpus of Spanish and English judgments, regarding lexical and semantic characteristics, English judgments almost lack Latinisms and terms of Greek origin, contrasting with the majority of legal texts. However, there is presence of archaisms such as *therein*, *thereof*, *forthwith*, etc. Moreover, when judges express their opinions, sometimes using personal and impersonal references, they show a respectful language, and use the first person singular (except in the conclusion, where they use impersonal constructions) or the first person plural in those cases in which the court is formed by more than one judge.

In sum, there is no doubt that judgments constitute a genre within the field of law and, as such, have their own peculiarities in terms of form and expression. Yet, while we have found several works that defend this position, we consider that it is necessary to find more evidence supporting the idea that the language of judgments have some distinctive features within the legal language and, more specifically, a particular phraseology. Thus, the next section addresses the definition of phraseology and the different classifications that we have considered until we have adhered to Biel’s (2014).

### **3. Phraseology**

Since this study is focused on the description of the specialized phraseology of the said LSP, the discipline, its characteristics, and the different classifications that have been proposed throughout the years are equally important and, thus, this section is dedicated to it. Broadly speaking, phraseology is the discipline that covers the study of phrases.

However, along the history of Linguistics, there have been many attempts to categorize the word “phrase” and, if we take into account the different appellations proposed by Thomas (1993)—who is just one of many authors who have tried to delimit it—LSP phrases can be “terminological phrases, LSP phrases, phrasemes, phraseological units or phraseological terms” (Thomas 1993: 56), among others. Therefore, in the next paragraph, the definition of phraseology will be narrowed down.

On the one hand, Roberts (1994) proposes a definition of phraseology by indicating that it is the study of “all habitual word combinations which do not belong to a specific grammatical category” (Roberts 1994: 63). However, to this definition must be added the one provided by Gläser (1994) referring to “the linguistic discipline which investigates the properties of idioms and phrases from a theoretical angle, classifies them according to their constituent structure and codifies them in dictionaries” (Gläser 1994: 45). Thus, Gläser approaches phraseology both from a theoretical and a practical perspective, since she also considers it as “the inventory or stock of phraseological units” (Gläser 1994: 45).

On the other hand, López and Moreno (2019), who have studied word combinations or multiword units, have reached the conclusion that the term used to refer to them is “lexical chunk”, since it “is a generic term covering a range of subtypes which have been classified according to their degree of semantic fixedness, syntactic fixedness, lexical restrictions and institutionalization” (2019: 32-59). Among these subtypes there are idioms, compounds, collocations, lexical bundles, etc. For that reason, in this paper we are going to adhere to their choice of lexical chunks as the term used to analyze the phraseology of judgments.

Timmis (2015) uses the term “multi-word unit” (MWU) as a synonym of “phraseological unit”, providing a classification of MWUs and giving three definitions: “collocations”, “lexical chunks” and “lexical bundles”. Collocations are combinations of two lexical words that appear together or in close proximity (e.g. *make sense*), although they may also appear separated by an article or made up of a verb followed by a preposition—which is the case of phrasal verbs. Lexical chunks are frequent sequences that have meaning and may include both lexical and grammatical words and, for that reason, they encompass a wider amount of sequences of words (e.g. *to a certain extent*).

Finally, lexical bundles are sequences of words found together without a clear semantic or pragmatic meaning (e.g. *it was in the*).

Nevertheless, as discussed by Roberts (1994), “phraseology includes ... all habitual word combinations which do not belong to a specific grammatical category” but she also argues that “not all types of phraseology are equally important for LSPs” (Roberts 1994: 63). For that reason, she considers formulaic expressions and lexical collocations the major types of phraseology found in LSP texts, and, thus, ignores grammatical collocations. Therefore, this choice made by Roberts leads us to Gläser (1994), who provides a grammatical classification of phraseology and shows the relations between phraseology and terminology in English texts. She starts differentiating between simple and complex words or lexemes, these last being the result of word formation processes (i.e. “compounds”, “derivatives”, “acronyms”, “blends” or “clippings”). Then, she defines phrases as “word group lexemes which may range from string compounds to terminological word groups” (Gläser 1994: 51).

Finally, Biel (2014) argues that phraseology is based on “fixed recurrent patterns at the textual, grammatical, and collocational level” (Biel 2014: 177) and makes a formal classification that can be seen as a phraseological pattern with not clear boundaries between categories—both at the macrostructural and microstructural level. Therefore, she classifies phraseology as follows: “text-organizing patterns, grammatical patterns, term-forming patterns, term-embedding collocations, and lexical collocations” (Biel 2014: 178-181):

- Text-organizing patterns, which are repetitive structures sometimes prescribed for drafting guidelines (e.g. *whereas, having consulted*).
- Grammatical patterns, genre-specific recurrent grammatical structures, e.g. conditional clauses (*provided that*) or the passive voice.
- Term-forming patterns (or multi-word terms): “collocates of a generic term” that add a higher degree of specificity (e.g. *European public limited-liability company*).
- Term-embedding collocations, which are verb-based structures that denote “what one can typically do with (or to) the object denoted by the base noun” (Martin, quoted in Heid 1994: 238), e.g. *judgment declaring a merger void*.

- Lexical collocations: routine formulae which are not constructed around terms identified through recurrence (e.g. *subject to this regulation*).

In sum, even though, Timmis' Roberts' and Gläser's classifications of phraseology have to be clearly understood as they are the basis of phraseological studies, it is necessary to note that for the purpose of this work a more specialized classification is needed. Hence, as the aim of this paper is to classify a series of legal lexical chunks extracted from a corpus of English and Spanish judgments contained within the European Union's case law, it is necessary to carry out a phraseological study of this particular field, and, in order to develop this categorization, we are going to base our findings on Biel's classification of phraseology in legal translation.

#### **4. Methodology**

This section is aimed at providing a detailed description of the steps followed for the compilation of the corpus. Thus, it is necessary to give a definition of Corpus Linguistics, since it is the tool that we are going to use in order to analyze the phraseology of the language of judgments. Then, we will describe the criteria used for the compilation of our corpus to later continue with its description, as well as the description of the type of study. The next step will be the analysis of the corpus in terms of qualitative and quantitative representativity. Once these steps are explained in detail, we will explain the program and tools used for the exploitation of the corpus and the criteria used for the selection of the keywords that will serve us as the basis for the phraseological analysis.

##### **4.1. Corpus Linguistics**

Corpus linguistics is a new wave of research with great potential for making progress in pure and applied linguistics as well as in Translation Studies (Laviosa 2002). In general terms, it is the study of language in use through corpora: it is a branch of applied linguistics that uses a large collection of texts to study language. Its main objectives are the documentation and generalization of patterns of use by assessing the extent to which a pattern is found. Therefore, it is an empirical approach that analyzes, through an extensive use of computers, patterns of language, using as the basis for the study a large collection of natural texts, and depending on both quantitative and qualitative analytical techniques (Biber et al 1998: 4). As Laviosa (2002) indicates, as a new wave of research

it has a great potential, and this is because it allows for a consistent and reliable linguistic study, which is carried out with a computer that allows for the storing of large amounts of data and the subsequent analysis by linguists. Therefore, we consider that it is entirely appropriate to make use of corpus linguistics as the tool that allows us to analyze the linguistic characteristics and peculiarities of the particular genre of judgments.

#### **4.2. Corpus description**

By definition, a corpus is “a collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis” (Francis 1992: 17). Thus, this corpus is representative of the field of law and, more specifically, the language of a particular genre within this field: judgments. It is monolingual because it only contains texts originally written in English. As this corpus cannot be taken as representative of the English language as a whole and, thus, it is not focused on languages for general purposes, the classification as general reference corpus must be dismissed. Therefore, as the main purpose of this corpus is to analyze some of the most frequent lexical chunks in the EU English judgments, it is a special purpose corpus restricted to the LSP of law. Besides, because of the formality of this particular genre, judgments must be presented in written form and, for that reason, this is a written corpus containing data from written documents.

Moreover, as the aim of the corpus is not related to the evolution of language over a long period of time but, rather, to the study of language used in a limited time frame—in this case from 2010 to 2018—the corpus is synchronic. On the other hand, this is a closed or static corpus because it has a fixed size: it has not been designed to be constantly expanded because it has been compiled just for the development of this study.

Apart from the type of corpus, the size is important as well: in this case, the corpus is made up of 264,704 words distributed in 40 documents. As “size is a question of fitness for purpose” (Timmis 2015: 2), this number of words—which could be considered small for written language—does not pose any problem because it is used for an LSP study and, therefore, it represents the language of a restricted discourse community. However, to make sure the corpus size was suitable for this linguistic study, we carried out the analysis of its representativeness (*vid infra*: section 4.3.1.).



### 4.3. Criteria used for the compilation of the corpus

For the compilation of the corpus, we restricted the material following different criteria. First, all the documents had to be written, even though the source of extraction—which was the webpage *EUR-Lex*—also contained spoken data. Second, all the material must be originally written in English and, for that reason, translated documents were discarded. In order to select original English texts, we made use of the advanced search option of *EUR-Lex* and we selected only the original texts, which are translated into many different languages corresponding to the constituent countries of the European Union. Finally, we took into account the publication date of the documents, and those texts issued before 2010 were dismissed.

Therefore, taking into account the above criteria, we followed four steps. First, we searched for judgments in the web, opting for extracting all the texts from the same source because it was an institutional webpage and, thus, it was more reliable. The webpage in question was, as previously mentioned, *EUR-Lex*. Once we have decided which was going to be the source of our data, we limited the search to the specific domain of case-law, to later select judgments in the “type of document” option. Besides, as it allowed to sort the data by date of publication, we decided to choose the option that showed documents from 2010. Finally, before downloading the documents, we also filtered the documents by topic, so that we have intellectual, industrial and commercial property; foreign affairs, environment, and freedom of movement for persons’ judgments. In addition, it guaranteed us having texts written by different authors in the same database, which allowed us to have a wider representativity of the features of the genre (since, if we had compiled texts of one single author, the results would have shown features of an idiolect).

The next step was downloading these documents, which was performed manually and taking into account that the documents were not translated versions but texts originally written in English (since this webpage provides not only the source language texts, but also translations into other languages of the European Union constituent countries). Another advantage of this webpage was that it gave the option of downloading the text in PDF (.pdf) or HTML (.html) format, so that it was not necessary to copy the text into a Word document (.doc) and then save it as plain text (.txt). Therefore,



formatting, which was the next step, did not take much time because the PDF document was easily converted into plain format (.txt) through the use of an online file converter.

Finally, we saved the documents and identified them through the design and creation of a corpus zero. A corpus zero consists on the creation of labels for each of the constituent texts of the corpus including all the extralinguistic information regarding each of them. It contains information about the language in which the text is written, the type of law, the country, the topic, the genre and the year of publication. For example, the first document in our corpus is labelled: EN001\_CASELAW\_UK\_IICP\_JUDG\_2018. This coding provides the following information:

- The language in which the text is written (EN).
- The numbering characters (001).
- The jurisdiction, which in this case is the Case Law (CASELAW).
- The country where the document is written (UK).
- The topic within judgments (IICP, which stands for intellectual, industrial and commercial property). The rest of acronyms are FA (foreign affairs), ENV (environment), and FMP (freedom of movement for persons).
- The type of document, which is judgment (JUDG).
- The year of publication of the text, 2018 in this case (2018).

#### **4.3.1. Analysis of the corpus: representativity**

Representativeness is a key feature in corpus linguistics. As a corpus is compiled to represent a particular language, it is important to consider if its size—quantitative representativeness—is wide enough to serve as a sample of a language variety or specialized language, as it is our case, and to prove that the quality of the corpus is achieved. Biber (1993: 243) defines representativeness from the viewpoint of how this quality is achieved and notes that “representativeness refers to the extent to which a sample includes the full range of variability in a population.”

##### **4.3.1.1. Qualitative representativeness**

Qualitative representativeness is a crucial point in the creation of the corpus, since it has into account the quality of the texts and their reliability in terms of their source.

First, the data must be extracted from reliable sources such as international companies, organizations or institutions. In this case, the source of the texts is *EUR-Lex*, a European Union official website which offers access to EU law, case-law by the Court of Justice of the European Union and other public EU documents. Within this website, there is a section called “case law” where one can find judgments and orders including “cases brought by EU institutions, Member States, corporate bodies or individuals against an EU institution or the European Central Bank, cases brought against EU Member States for failing to fulfil their obligations under the EU treaties, national courts' requests for preliminary rulings concerning the validity or interpretation of EU law, and disputes between the EU and its staff.”

#### **4.3.2. Quantitative representativeness**

As the corpus has to represent the language, its size matters, and, for that reason, it is necessary to carry out a reliable and objective analysis of its quantitative representativeness. For this examination, ReCor is an effective tool that provides a posteriori analysis of the minimum size of the corpus or the level of representativeness in a simple graph form.

Apart from providing a graphical representation, this tool is also useful because, by simply introducing the files in the program, it creates three different files containing a statistical analysis, an alphabetically-arranged list of words and a list of words ordered by frequency. Besides, it presents the option of specifying the n-grams, allowing to check if the corpus is representative from 1 to 10 grams. This last option is very useful in order to carry out the phraseological study of Biel's term-forming patterns and term-embedding collocations.

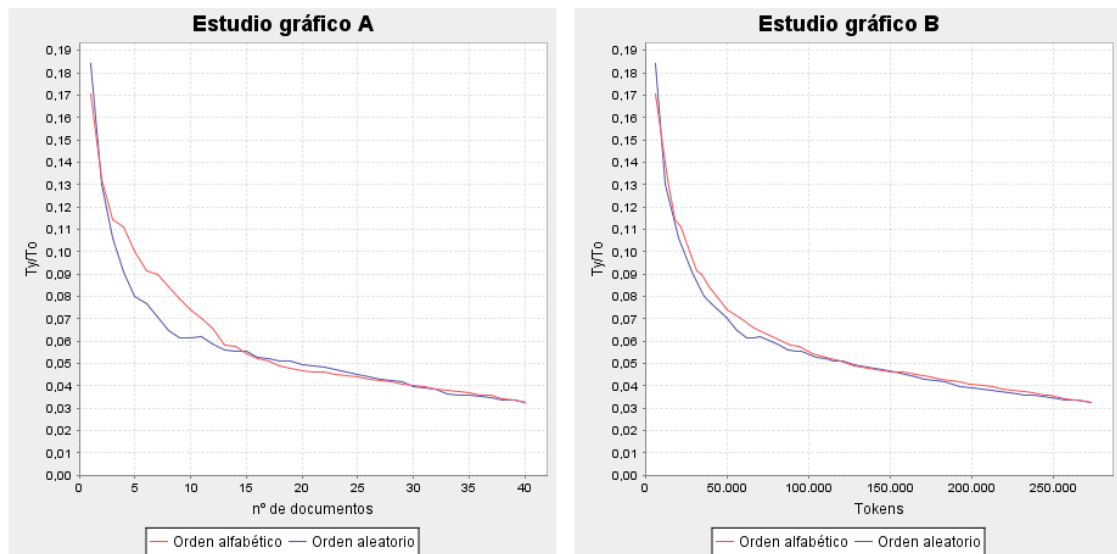


Figure 1. Corpus Representativity.

As can be seen, within graphical representation A, the number of files selected is included in the horizontal axis, whereas the types/tokens ratio is shown on the vertical axis. From that, one can conclude that the corpus begins to be representative with 1-gram from the point of inclusion of approximately 2 documents, since it is at that point where both the red line (files ordered alphabetically) and the blue line (files introduced at random) stabilize and the curve hardly varies either before or after this point. On the other hand, graphical representation B can be used to determine the total number of words that should be set for the minimum size of the collection and, for that reason, the horizontal axis shows the number of tokens and the vertical axis shows the types/tokens ratio. Therefore, graph B shows that the minimum total number of words necessary for the corpus to be considered representative (with 1-gram) is approximately 10,000 words.

#### 4.4. Type of study

As the aim of this corpus is not to explore a theory or hypothesis with the purpose of validate or refute it, the study cannot be described as a corpus-based one. On the contrary, the fact that “corpus-driven linguistics rejects the characterization of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language” (Tognini-Bonelli 2001:84-85), makes clear the type of study that will be applied in this work: the corpus-driven approach. Therefore, it is after the analysis of the corpus when the linguist can make a generalization and

categorization of the data and, thus, obtain results—which shows the inductive nature of this approach, in which the starting point is to make few a priori assumptions and check whether they are right in the corpus.

On the other hand, continuing with the characterization of the study, it can also be asserted that it is a semaseological study, since the first step is based on the elaboration of a list with the most frequent terms in the selected English judgments in order to be able to find out if they are part of larger lexical chunks and, if so, classify them according to the classification that Biel proposes for legal phraseology. Even though the semasiological approach usually goes from a term to find a definition, in this case we will not get that definition, but information to complement it. Thus, this approach is chosen by lexicographers and dictionary makers who know which words have the higher frequency of use and have to decide whether they include them in a dictionary.

#### **4.5. Selection of candidate terms**

For the selection of candidate terms, we have used the Word List tool of AntConc, which automatically generates a list with all the terms contained in the corpus. The list can be sorted by frequency, by word or by word end, so, for the purpose of this work we have selected the frequency option because one of our requisites for a term to be selected was that it has to appear more than one hundred times. However, as in any other document, judgments include what is called noise or empty words which are not relevant for our study, it is the case of articles or pronouns. Thus, the Tool Preferences option of the program gives the possibility of uploading a stoplist—a list containing all the words we do not want to have in the word list—and, thus, eliminate this noise. This stoplist can be downloaded from the internet or be created for one specific corpus, which is our case: we have created a TXT file with all the words we do not want to appear in the word list created by AntConc and we have imported it to the program.

Another requisite was the appearance of the term in more than five different files, since we are looking for patterns based on meaning that characterize the legal genres and, for that, we need the term to be used by different authors. The Concordance tool of AntConc is the one which allowed us to see if the term appeared in different files and, if so, to see the context in which it appeared. Once we applied these two requirements, we

selected fifty terms in order to see their context and if they could be of interest for the study. These candidates were:

1. access	11. being	21. effect	31. main	41. refer
2. accordance	12. case	22. ground	32. make	42. regard
3. action	13. commission	23. having	33. meaning	43. regulation
4. agreement	14. condition	24. host	34. measure	44. respect
5. annex	15. convention	25. interpretation	35. order	45. right
6. appeal	16. cost	26. issue	36. procedure	46. rule
7. appellant	17. council	27. judgment	37. proceeding	47. shall
8. application	18. court	28. law	38. protection	48. should
9. article	19. decision	29. leave	39. provide	49. subject
10. basis	20. directive	30. legal	40. provision	50. would

*Table 1. Term candidates in the EU English Judgments.*

On the other hand, since the Word List tool creates a list of words, but not of lemmas, the Tool Preferences option also gives the option of uploading a lemma list. Once this list is uploaded, the Word List tool generates a list of lemmas with all the lemma word forms associated to them, which appear in a column on its right. In our case, we created a specific lemma list for the above candidate terms, since they were the ones we were interested in (See Figure 2 below).

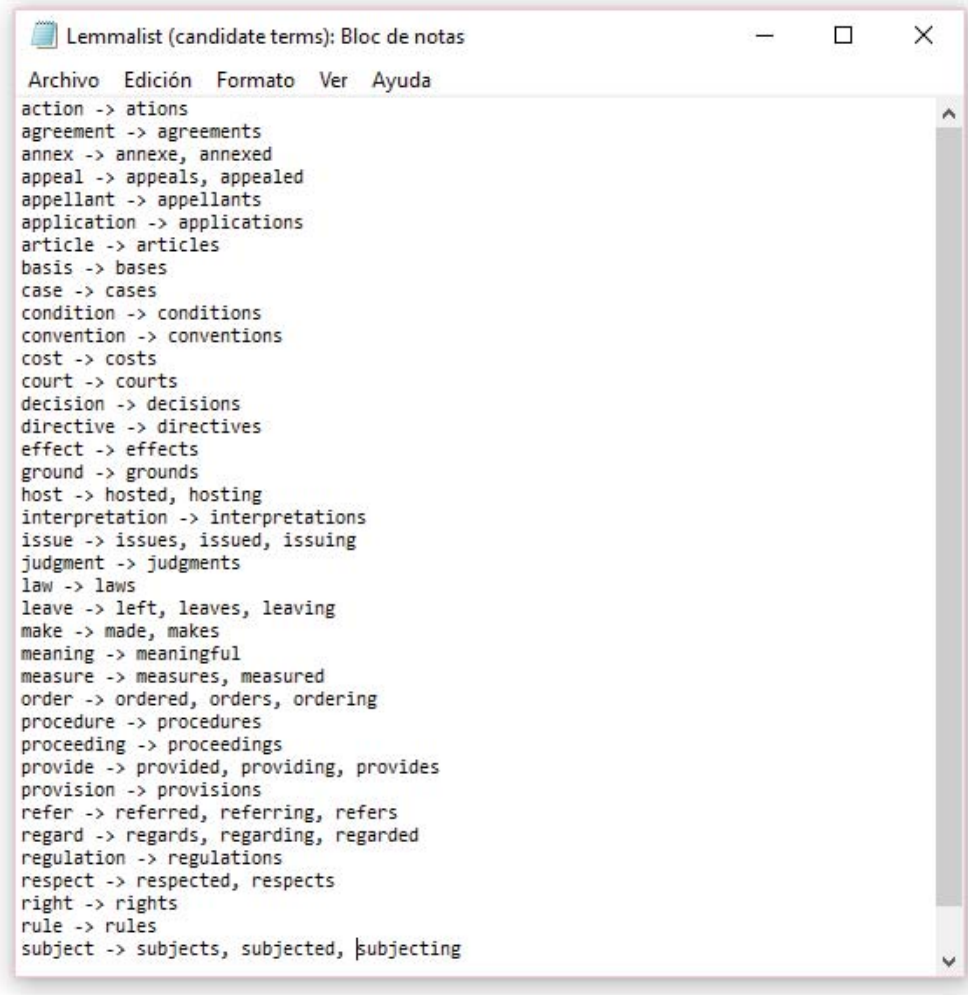


Figure 2: Lemma List for the Candidate Terms.

Then, from these fifty candidate terms, we discarded those which did not appear in specialized dictionaries of legal English or which, after taking a glance at the Collocates tool of AntConc, did not provide enough results for the later classification. As a result, we have a list with eighteen of the most frequent terms in judgments, ranging from 1883 appearances in the case of “court” to 144 in the case of “provide”, considering all the lemma word forms that can be seen in Figure 2 above. This list will be further analyzed in terms of context through the use of the Concordance tool of AntConc, which will allow us to see the context in which each term is used. The following table contains these eighteen terms with their respective appearances in brackets:

article (2389)	case (865)	proceeding (572)
court (1883)	commission (771)	provision (560)

regulation (1368)	appeal (715)	application (452)
directive (1242)	decision (699)	regard (371)
law (909)	council (613)	right (279)
judgment (899)	refer (613)	provide (144)

*Table 2: Selected Terms for the Classification.*

## 5. Results and classification

In this section we are going to classify the results obtained from our corpus of English judgments of the European Union according to Biel’s (2014) classification of legal phraseology. She bases this classification on meaning and proposes five categories that we have previously detailed (vid supra: section 3): text-organizing patterns, grammatical patterns, term-forming patterns, term-embedding collocations, and lexical collocations.

As mentioned in the previous section, in order to obtain the results, we have made use of the Concordance tool of AntConc. This format is what is known as “KWIC”: Key Word in Context. Thus, by entering a word in the search box, the program generates a list with all the occurrences that the term has and indicates the file in which they appear. Besides, it has the Kwic Sort option, which gives the possibility of rearranging the concordance lines at three different levels, being 0 the search word, and 1L, 2L... or 1R, 2R... the words that appear to the left or right of the target word, respectively. Therefore, by using this option, we can find lexical chunks of the search word quickly and in an organized way. Even though the program automatically generates the list with all the concordances of one term, the task of selecting the lexical chunks is carried out manually: we created a table with the selected terms and their respective KWIC. Moreover, in the same way as in the selection of candidate terms, for the selection of the lexical chunks we have considered one requisite: the pattern has to appear at least in two different documents in order to ensure that it is used by different authors and, thus, is a possible category within the classification of legal language proposed by Biel (2014). Once we applied this requisite to our eighteen terms and created the abovementioned table with the terms and their KWIC, we carried out the classification and the results are presented in the following lines. In order to make the classification clearer, we have created tables for each of the five categories (i.e. text organizing patterns, grammatical patterns, term-forming patterns,

term-embedding collocations, and lexical collocations) where we have included the eighteen terms and examples of concordance lines showing that they fall within the category under analysis.

<b>Text-organizing patterns</b>	
article	<p><i><u>Article 2(a) of that directive states:</u></i></p> <p><i><u>Article 28 of Directive 2001/83 provides as follows:</u></i></p> <p><i><u>Having regard to Article...</u></i></p> <p><i><u>In accordance with Article 1(2)(a)(i) and (b)(i) of that regulation, ...</u></i></p>
court	<p><i><u>In the light of the foregoing, the Court finds that...</u></i></p> <p><i><u>On those grounds, the Court (Grand Chamber) hereby rules:</u></i></p> <p><i><u>THE COURT (Grand Chamber), composed of...</u></i></p>
regulation	<p><i><u>For the purposes of this Regulation, ...</u></i></p> <p><i><u>In accordance with regulation 19(3)(a) of those regulations, ...</u></i></p> <p><i><u>In the light of Regulations No 1049/2001 and No 1367/2006</u></i></p> <p><i><u>On the basis of regulation 16 of those regulations, ...</u></i></p>
directive	<p><i><u>In accordance with Directive 2001/83...</u></i></p>
judgment	<p><i><u>By judgment of 29 September 2010, ....</u></i></p> <p><i><u>In accordance with the judgment...</u></i></p> <p><i><u>JUDGMENT OF THE COURT</u></i></p> <p><i><u>(see, by analogy, judgment of 12 March 2015...)</u></i></p> <p><i><u>(see, inter alia, judgment of 7 January 2004...)</u></i></p>
case	<p><i><u>In such a case, ...</u></i></p> <p><i><u>That being the case, ...</u></i></p> <p><i><u>(see, to this effect, inter alia, CASE C-358/01 Commission v Spain...)</u></i></p>
commission	<p><i><u>According to the Commission, ...</u></i></p>
appeal	<p><i><u>(Appeal — EU trade mark — Three-dimensional mark representing the shape of a four-fingered chocolate bar — Appeal directed against the grounds — Inadmissibility — Regulation (EC) No 207/2009 — Article 7(3) — Evidence of distinctive character acquired through use)</u></i></p> <p><i><u>By its appeal, ...</u></i></p> <p><i><u>In support of its appeal...</u></i></p>



decision	<i>By decision of 26 September 2017, ...</i>
council	<i>As amended by Council,</i>
proceeding	<i>For the purposes of this proceeding, ...</i>
provision	<i>According to that provision, ...</i> <i>In accordance with that provision, ...</i> <i>On the basis of that provision, ...</i>
application	<i>By its application, ...</i> <i>By application lodged at...</i>
regard	<i>Having regard in particular to...</i> <i>Having regard to...</i> <i>In that regard...</i> <i>With regard to...</i>
provide	<i>Article 1 of Regulation No 469/2009, headed 'Definitions', provides as follows:</i>

Table 2. Text-organizing patterns.

As can be seen in Table 3 above, most of the selected terms are part or appear within text-organizing patterns. These patterns “have a high frequency in the corpus because they are consistently repeated in legislative instruments with very low variation”, (Biel 2014: 178), which is the case of judgments. Some typical patterns are the titles of documents or sections within documents (e.g. *JUDGMENT OF THE COURT*), and the transitions between sections (e.g. *in accordance with...*, *having regard to...*).

<b>Grammatical patterns</b>	
article	<i>Article 21(1) TFEU must be interpreted as meaning that...</i>
court	<i>If the Court were to...</i> <i>In the event that the Court finds that...</i> <i>The Court may refuse to...</i> <i>The Court should set aside...</i>
regulation	<i>For the purposes of this regulation...</i> <i>That regulation was adopted...</i> <i>This regulation shall apply to...</i>

directive	<i><u>This Directive shall be without prejudice to provisions</u></i>
judgment	<i><u>The judgment under appeal must be set aside.</u></i> <i>...do not permit <u>the final judgment to be given...</u></i>
appeal	<i><u>The appeal must be dismissed in its entirety.</u></i>
decision	<i><u>Where a decision, adopted in accordance with...</u></i>
council	<i><u>That being the case, the Council was in...</u></i> <i><u>The Council shall state individual and specific reasons for...</u></i> <i><u>The Council shall, acting in accordance with...</u></i>
proceeding	<i><u>For the purposes of this proceeding...</u></i>
provision	<i><u>If that provision is to be interpreted literally, ...</u></i> <i><u>It is only if certain provisions of Directive 2004/38 may not be dissociated...</u></i> <i><u>That provision must be interpreted as meaning that it...</u></i>
application	<i><u>For the purposes of the application of Article 3(a)...</u></i> <i><u>The application was submitted...</u></i>
right	<i><u>This right shall not be subject to conditions...</u></i>
provide	<i><u>In those circumstances, provided that the Court has...</u></i>

Table 3. Grammatical Patterns.

Regarding grammatical patterns, there are examples which express deontic modality, as in the case of *Article 21(1) TFEU must be interpreted as meaning that...* or *The Court should set aside...* There are also conditional clauses such as *If the Court were to...* or *In the event that the Court finds that...*, and purpose clauses (e.g. *for the purposes of this regulation/proceeding*). In the case of “application” and “regulation”, there are examples of passive structures, which can be considered grammatical patterns, it is the case of *The application was submitted* or *That regulation was adopted*. These examples of concordance lines (Table 4) recurrently appear in our corpus, which tell us that they are genre-specific grammatical patterns.

<b>Term-forming patterns</b>	
article	<i><u>Article 16 of Directive 2003/87</u></i> <i><u>Article of regulation</u></i>

court	<i>Competent Court</i> <i>Court of law</i> <i>Criminal Court</i> <i>European Court of Human Rights</i> <i>General Court</i> <i>National Court</i> <i>Open Court</i> <i>Supreme Court</i> <i>The referring court</i>
regulation	<i>Basic regulation</i> <i>Commission Regulation</i> <i>Council Regulation</i> <i>Implementing regulation</i> <i>Provisional regulation</i>
directive	<i>Council Directive</i>
law	<i>Applicable law</i> <i>Environmental law</i>
judgment	<i>Final judgment</i> <i>Judgment of the Court</i> <i>The present judgment</i>
case	<i>The present case</i>
decision	<i>Draft decision</i> <i>Final decision</i> <i>Formal decision</i>
proceeding	<i>Criminal proceeding</i> <i>Divorce proceeding</i> <i>Infringement proceedings</i> <i>Invalidity proceedings</i> <i>Judicial proceedings</i>
provision	<i>National provisions</i> <i>The corresponding provisions</i>
application	<i>European patent application</i>

	<i>SPC application</i>
right	<i>Independent right</i> <i>Reproduction right</i> <i>Right of citizens of the Union</i> <i>Right of communication</i> <i>Right of free movement and residence</i> <i>Right of permanent residence</i>

Table 4. Term-forming patterns.

As can be seen in Table 5, most of the term-forming patterns indicate subtypes, they are subtype-denoting collocations that are often subject to terminologization and form distinct terms (Biel 2014: 180). In our corpus, the most frequent pattern is Adj + N, even though we have also found N + N, and a few complex multi-word terms (e.g. *right of free movement and residence*; *European Court of Human Rights*). Term-forming patterns are, according to Biel, multi-word terms, and they can be seen as frozen collocations because of their great structural stability. Some examples of the pattern made up of a noun preceded by an adjective are: *final judgment*, *criminal proceeding* or *basic regulation*. As we have mentioned, there are also patterns formed by two nouns (N + N), such as *Council Regulation*, *divorce proceeding* or *SPC application*, but there are also structures with two nouns joined by the preposition “of” (N + Prep + N): *Article of Regulation*, *judgment of the Court* or *right of communication*.

<b>Term-embedding collocations</b>	
article	<i>Article 7 of that regulation provides for:</i> <i>Article 25 of the Convention reads as follows:</i>
court	<i>The court erred in law</i> <i>The court finds that</i> <i>The court ruled that</i> <i>The General Court failed to ascertain</i>
directive	<i>This Directive lays down</i>
case	<i>The case gave rise to the judgment</i>
commission	<i>The Commission asserts</i> <i>The Commission brought the present action</i>

	<i>The Commission points out</i> <i>The Commission seeks a declaration</i> <i>The Commission states</i>
appeal	<i>To bring an appeal</i> <i>To dismiss the appeal</i> <i>To lodge an appeal</i>
decision	<i>The decision is adopted</i> <i>To adopt a decision</i> <i>To challenge a decision</i> <i>To dispute a decision</i>
council	<i>The Council considers</i> <i>The Council contends</i> <i>The Council notes</i> <i>The Council submits</i>
proceeding	<i>The main proceedings concern</i>
provision	<i>That provision allows</i> <i>That provision confers</i> <i>That provision requires</i>
application	<i>To file an application</i> <i>To make an application</i>
right	<i>The right grants</i> <i>To acquire a right</i> <i>To claim a right</i> <i>To constitute a right</i> <i>To guarantee the right</i> <i>To qualify for a right</i>

*Table 5. Term-embedding collocations.*

Term-embedding collocations are verb-based structures that express actions related to terms and, for that reason, in our corpus we have found the N +V pattern to have a high frequency. In this case, we are focusing only on the nouns, and we can see that they are usually accompanied by specific verbs that express “what one can typically

do wit (or to) the object denoted by the base noun” (Martin, quoted in Heid 1994: 238). There are plenty of examples that demonstrate that the N + V pattern is almost prototypical, some of them are: *the Court erred in law*, *the Commission points out or the right grants*. These collocations “form the skeleton of legal rules by providing action and enabling terms to enter into relations” (Biel 2014: 180).

<b>Lexical collocations</b>	
article	<i><u>In accordance with Article 52(2) of Regulation No 40/94, ...</u></i>
court	<i><u>According to the referring court, there are also divergent decisions in a number of Member States concerning the issue...</u></i>
regulation	<i><u>The fifth recital in the preamble to Regulation No 1612/68 reads as follows:</u></i>
directive	<i><u>...so that the authority establishes the plan required by the Directive in accordance with the conditions laid down by the latter.</u></i>
law	<i><u>The appellants contend that the General Court erred in law in...</u></i>
judgment	<i><u>Notwithstanding the judgments delivered by the Court on...</u></i>
case	<i><u>In cases falling under Article 3(2)(b), ...</u></i>
commission	<i><u>By its application, the European Commission asks the Court to declare that, ...</u></i>
appeal	<i><u>To order to pay the costs of the appeal.</u></i>
decision	<i><u>The Court is to make a decision as to the costs.</u></i>
refer	<i><u>The Commission requests the Court to set aside the judgment under appeal, refer the case back to the General Court and reserve the costs of the present proceedings.</u></i>
proceeding	<i><u>The request has been made in proceedings between [...], on the one hand and, on the other, [...] concerning...</u></i>
provision	<i><u>Within the meaning of that provision</u></i>
application	<i><u>“basic patent” means a patent which protects a product as such, a process to obtain a product or an application of a product, and which is designated by its holder for the purpose of the procedure for grant of a certificate</u></i>
regard	<i><u>...having regard to the written procedure and further to the hearing on...</u></i>

right	<i>...where the Union citizen, having exercised his right of freedom of movement to work in a second Member State, <u>in accordance with the conditions laid down in...</u></i>
provide	<i><u>Article 1 of Directive 2008/50, entitled 'Subject matter', provides:</u></i>

Table 6. Lexical Collocations.

Lexical collocations differ from the previous categories in the sense that they are not built around terms, however, as we wanted to obtain examples of this category, we have taken into account our selected terms and we have glanced at their context in order to see routine formulae that tend to appear near them. The criterion we have followed for the selection of the examples is that of recurrence, since it is the repetition of these structures which let us classify them in the category of lexical collocations. “In fact, it appears quite common for longer sequences of words to pattern together. Some of these recur frequently enough to be treated as units in their own right.” (Schmitt, quoted in López and Merino 2019: 32-59). Thus, we have taken into account only those lines that were repeated in more than two different files and that have exactly the same structure (except for information regarding the number of an article, the name of the appellants or the dates when an instrument was written, among others), and we have extracted one example for each of the selected terms, as can be seen in Table 7.

## 6. Conclusion

After having analyzed and classified the phraseology of the judgments contained in our corpus, we can finally know that there is a specific phraseology in this type of texts that differentiate them from other LSPs and, thus, reassert Biel’s idea that “collocations of editing units used for textual mapping are genre specific features of the language of the law due to their high frequency and salience” (Biel 2014: 190). In fact, we have reached the conclusion that these texts are composed of patterns that are recurrently used and, for that reason, we have proved that their analysis and classification can facilitate tasks such as the understanding, composition and translation of this type of texts. This recurrence in phraseology is due to the fact that legal language is characterized by being extremely formulaic—which has sometimes made the language of law be subject of some criticism. However, we consider that this formulaicity is not a disadvantage but a way to make easier the aforementioned tasks, since “recurrent patterns save cognitive effort

during processing and facilitate concentrating on new information” (Biel 2014: 190). Therefore, when the audience—who are usually non-expert or semi-expert readers—receives a judgment for the first time, they may not understand some pieces of information; however, after reading several judgments and realizing their formulaic nature, they will be able to differentiate between these repetitive patterns that we have classified and that appear in almost every judgment, and the new information—which is the one they have to focus on. The same will happen in those cases in which the audience prefer to hire a intralingual translator: the availability of parallel texts and studies of phraseology in legal language will facilitate and accelerate the process of transmission of information from a specialized language to a general language.

On the other hand, as noted by López and Moreno (2019) “the study of phraseology has become an essential part of the study of Language for Specific Purposes (LSP), especially since corpus linguistics revealed their key role is to study language use and function.” (López and Moreno 2019: 32-59). Thanks to this study, we can also prove this idea and support corpus analysis as an undoubtedly effective tool for the analysis of phraseology and, thus, for the creation of glossaries or terminological records, among others. In fact, the specialized classification we have made of the phraseology of judgments of the European Union can serve as the basis for the creation of terminological records—which present the term in its immediate context and give information about their usage or structure (López and Moreno 2019: 32-59)—and, therefore, contribute to the characterization of the language of the European Union.

Finally, we hope this work serves as a way to support the idea that it is important to study specialized languages and, more specifically, their phraseology, because, in the case of judgments, it has been proved to be one of their distinguishing features. For that reason, we consider necessary the existence of specialized classifications, such as the one proposed by Biel (2014), which has served as the basis for our analysis and classification of the specialized phraseology of English judgments. Actually, through this classification we have reached the conclusion that the best way to adequately understand and write specialized texts—judgments in particular—is to contextualize the use of their terms and, for this task, the most suitable approach is the use of corpus linguistics to obtain recurrent patterns that will later be analyzed through the use of this specialized classifications.



## 7. Bibliography

- Alcaraz, E., & Hughes, B. (2002). *Legal Translation Explained*. New York: Routledge.
- Allan, K. (2007). *The Western Classical Tradition in Linguistics*. London: Equinox.
- Biber, D. (1993). "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8(4), 243-257.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biel, Ł. (2014). "Phraseology in Legal Translation: A Corpus-Based Analysis of Textual mapping in EU Law." In *The Ashgate Handbook of Legal Translation* ed. by L. Cheng, K. K. Sin & A. Wagner, 177–193. Surrey/Burlington: Ashgate.
- Borja, A. (2000). *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel Lenguas Modernas.
- Borja, A. (2007). *Estrategias, materiales y recursos para la traducción jurídica: inglés-español*. Castelló de la Plana: Universitat Jaume I.
- Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.
- Calvi, M. V. (2009). *Las lenguas de especialidad en español*. Roma: Carocci.
- Di Prisco, M. (2018). LGP vs LSP - The Language of Medicine: Vocabulary and Terminology. *TranslationDirectory.com*. Retrieved May 8, 2019, from <https://www.translationdirectory.com/articles/article2545.php>
- Francis, W. N. (1982). "Problems of Assembling and Computerizing Large Corpora." In *Computer Corpora in English Language Research*, ed. S. Johansson, 7-24. Bergen: Norwegian Computing Centre for the Humanities.
- Gläser, R. (1994) "Relations between Phraseology and Terminology with Special Reference to English." *ALFA* 7/8, 41-60.

- Heid, U. (2001). "Collocations in Sublanguage Texts." In *Handbook of Terminology Management*, ed. G. Budin & S. E. Wright, 788–808. <https://doi.org/10.1075/z.htm2.31hei>
- Jones, C., & Waller, D. (2015). *Corpus Linguistics for Grammar: A Guide for Research*. London: Routledge.
- Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, Findings, Applications*. New York: Rodopi.
- López, B., & Moreno, L. (2019). "Lexical Chunks in English and Spanish Sales Contracts." *Terminology* 25(1), 32-59.
- Roberts, R. P. (1994). "Identifying the Phraseology of Languages for Special Purposes." *ALFA* 7/8, 61-73.
- Ruiz, M. A. (2013). "Cognition and Context of Legal Texts: Spanish and English Judgments Compared." *Revista de Lingüística y Lenguas Aplicadas* 8, 76-92. <https://doi.org/10.4995/rlyla.2013.1245>
- Thomas, P. (1993). "Choosing Headwords from Language-for-Special-Purposes (LSP) Collocations for Entry into Terminology Data Bank (Term Bank)." In *Terminology* 8, ed. H. B. Sonneveld & K. L. Loening, 43-68. <https://doi.org/10.1075/z.70.05tho>
- Timmis, I. (2015). *Corpus Linguistics for ELT: Research and Practice*. London: Routledge.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: J. Benjamins.