



Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS

DEPARTAMENTO de FILOGÍA INGLESA

Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

SPANISH AND ENGLISH TERMINOLOGICAL STUDY
BASED ON COMPARABLE CORPUS OF EMPLOYMENT
CONTRACTS

IDALIA GONZÁLEZ MORAIS

Tutor: Belén López Arroyo

2015/2016

ABSTRACT

The evolution of new technologies has transformed the working method of the study of language in the last few years. The use of electronic corpora has facilitated the task of both experts and learners, especially in the speed of compilation and analysis of data.

For this reason, this final work is based on a corpus linguistic methodology for the analysis of the specialised language of employment contracts in Spanish and English. The analysis includes a glossaries for the Spanish and English languages of the most common twenty terms in this type of document. Each of these words includes a real example of use and our translation proposal.

Keywords: *AntConc, corpus, corpus linguistics, employment contracts, specialised language, terminology.*

RESUMEN

La evolución de las nuevas tecnologías ha transformado el método de trabajo del estudio del lenguaje en las dos últimas décadas. El uso de corpus electrónicos ha facilitado la tarea tanto a expertos como a aprendices, sobre todo en la rapidez de compilación y análisis de los datos.

Por ello, este proyecto se basa en la metodología de la lingüística del corpus para analizar el lenguaje especializado de los contratos de trabajo tanto en español como en inglés. El análisis finaliza con un glosario en español y otro en inglés formados por los veinte términos más comunes en este tipo de documentos. Cada una de estas palabras cuenta con un ejemplo real de su uso y una propuesta traducción propia.

Palabras clave: *AntConc, contratos de trabajo, corpus, lenguaje especializado, lingüística del corpus, terminología.*

INDEX

Abstract	3
Index	4
Introduction	5
Justification of the area chosen and aim	5
Hypothesis and objectives.....	7
Linkage to the competencies of English Studies.....	8
Relevance for the users	9
Parts of the work	9
A. Legal language	11
A.1. Spanish Employment Contracts	11
A.2. English Employment Contracts.....	13
B. Methodology	16
B.1. Corpus linguistics	16
B.2. What is a corpus	17
B.3. Types of corpora.....	18
B.4. Approaches to corpus linguistics.....	23
C. Criteria for compilation of the corpus	26
D. Results and analysis	30
D.1. English Wordlist.....	30
D.2. Spanish Wordlist	45
D.3. Real examples and equivalent translations.....	57
E. Conclusion	63
Bibliography	64

INTRODUCTION

Justification of the area chosen and aim

This final project uses part of the knowledge that has been obtained over the years while studying the degree in Estudios Ingleses, especially in the fields of linguistics and translation. Indirectly, cultural and instrumental knowledge about Spanish and English languages, supplement the development of this work in accordance with the topic. While it is true that the final project focuses on a course from all that encompass the degree of English Studies, we are using some others for the development of the above mentioned work.

The subject of the final project is New Technologies Applied to Estudios Ingleses; it is well known that nowadays there is a vast number of new technologies that are making our work much easier; examples of these tools in the field of linguistics are corpus management software, translation memories or terminology management tools. However, all of them are just tools to help and the output they produce needs to be interpreted.

Taking into account the above considerations, the aim of this project is to make an English Spanish glossary with the most common legal terms used in Employment contracts. Documentation is very important when doing a translation or writing in English, particularly if we are dealing with a specialized language. Terms have an accurate and specific meaning that must be known if we want to produce a text that can be accepted for the discourse community it is addressed to. So, dictionaries seem to be a very important tool. However, some dictionaries lack of examples of language use, and or the information they include is limited. Apart from that, in the world we are living nowadays, new terms are produced on a daily basis and hence, sometimes, dictionaries cannot be updated so fast and hence they do not include relevant terms. Our intention then is to complete or complement the information dictionaries include with a glossary.

This project is based on the use of new technologies to study a specialized language. As we said before, the specialized field of study will be *law* and within this, we will focus on employment contracts. We have chosen this particular part of the law because it is an area we have been working before. Apart from that, we are familiar with this area from

the point of view of the translation from English into Spanish courses, and this is why we have the knowledge of the laborious work involved on it.

Our glossary will be based on a corpus. We believe that a corpus is a great tool to learn and contribute to provide information about language for specific purposes (LSP). Corpora show the real use of specialized languages, since they contain real texts. In addition, they also show the behavior and the real context of the words and the linguistic units. Also, they are test benches to refute or confirm hypotheses. Corpora offer us endless amounts of frequency analysis, such as terminology or usual linguistic and stylistic patterns, which are closely related to the purpose of our project.

The most common specialized terms in Legal language will be extracted from the employment contracts that will form our corpus. According to Paul Baker, Andrew Hardie & Tony McEnery in their work *A Glossary of Corpus Linguistics*, "A lexical syllabus is based on the argument that it makes sense to teach the most frequent words in a language first as these words tend to have a wide variety of uses" (2006 : 107). This can also apply to specialized language, and that is what we will do with the glossary and its most frequent terms. Sinclair and Renouf already mentioned (1988: 140-160) that a study plan for teaching languages can be constructed taking into account the use of frequency-based information given by a corpus.

As we mentioned above, the information contained in dictionaries is limited, especially if we are dealing with specialized language. But not all the difficulties encountered have to do with the choice of one term or another. Other problems we have to face when we deal with Legal English and Spanish are polysemic terms, terms that, depending on the place of the text or the sentence mean one thing or another, or the insufficient information provided in specialized dictionaries that can ruin our work, especially if we do not have all the tools that are required. That is another reason why a corpus is an excellent support tool that enables us to make a good translation, to write technical texts or even improve the translations that are already made.

Hypothesis and objectives

Our hypothesis starts from the fact that the information contained in dictionaries is not enough. For this reason, we believe that the use of other sources of documentation, such as corpora, complements the help of dictionaries.

The objectives we aim at with scientific research are the following:

- I. To prove that the terminology that exists in a specialized field as the language of law has a greater presence in a more specific subfield such as employment contracts.
- II. Many of these terms in Spanish have no terminological equivalent in English and vice versa. So, our second aim will be to provide equivalents by an explanation or fixed expressions.
- III. To use a tool that will provide us a complete terminological study; in other words a comparable corpus in English and Spanish languages, and also specialised in employment contracts collected from the Internet.
- IV. To study the most common terms in employment contracts from our bilingual corpus, so as to contribute to the establishment of equivalences in both languages and analyze legal conceptual problems in the area of translation.

The results of the foregoing tasks will lead us to the ultimate goal of this work. The aim of this paper is to make a glossary that includes the most frequent specialized terms contained in employment contracts using technological tools. This glossary will comprise the English terms and their respective equivalences in Spanish.

Apart from that, we would be delighted if this glossary may contribute, as far as possible, to research on legal translation, due to the contribution of real data on the subject. But above all, the most common and realistic use of this work, is to give an easier access to legal texts in Spanish and English languages.

According to the aims previously established in this section, we will follow the procedure set below:

- I. The compilation of comparable and bilingual corpus of employment contracts in English and Spanish languages, as a basis for subsequent terminological study.
- II. The analysis of the compiled corpus by using software programs to extract a list of most frequent terms in employment contracts.
- III. The study of each of the most common terms to establish their equivalences in both languages.

Linkage to the competencies of English Studies

The competences pursued throughout the work have the function of making a glossary as complete as possible. Without these skills, the performing of the glossary would be impossible. The following skills are essential requirements not only for this work, but also to understand, to analyze and interpret what is understood, and how to search and find what is not understood, to achieve a parallel understanding between Spanish and English languages.

- Learning to read, understand, interpret and translate the grammatical features of the Spanish and English language, and their behavior in the legal field of contracts.
- Knowledge of the different linguistic registers and their adaptation to the legal context.
- Acquiring the theoretical and practical basic knowledge pertaining to legal language.
- Addressing the translational problems and confusion because of morphological similarities between certain words.
- Knowledge of a specific part of the culture of each country and its evolution in history when compared with other synchronic or diachronic corpora.

- Knowledge of the techniques and methods of linguistic, literary, cultural and instrumental analysis of a second language.
- Search and interpretation of relevant data that end with the issuance of judgments that include reflections on specialized topics.

All the competencies mentioned above could be summarized in the main objective of this project. With this project, we will display the information, problems and solutions that we have encountered for users, both specialist and non-specialist. All this information will be contained in the glossary. Every word of the glossary will provide the most complete information we may obtain from our corpus.

Relevance for the users

The final glossary is framed within the field of lexicography, and the glossary is intended for two types of users. The main user is the professional who can be a translator, a technical writer, an English student or a legal professional who need to have access to the equivalents in English. Also, this glossary is intended for ordinary people who have the interest, the curiosity, or the need to clearly understand the meaning of the terms of any contract of employment in English without being an expert on the subject.

Experts may think that there is enough information in bilingual dictionaries, but we know from our own experience that the information is limited and does not provide us adequate accuracy when we use a concept in translation or writing. We can also be criticized for giving more than one fixed solution, but it is understood that not all fixed translations serve equally and we must adapt them to the context.

Parts of the work

This work is divided on four chapters. The first chapter contains the necessary concepts to understand the legal language in Spanish and English. The second chapter is devoted to the methodology we are going to use. Second chapter also defines what a corpus is, the different types of corpora that exist, and our approaches to Corpus Linguistics. The

third chapter explains the criteria for the compilation of the corpus and the methodology we use to create a corpus zero. In the last chapter of the work, we show the analysis and results obtained. This part concludes with two glossaries in Spanish and English with the twenty most common specialized terms in employment contracts. Each term is provided with an equivalent in the other language and examples of translation.

A.LEGAL LANGUAGE

The legal system is used to handle major events. For this reason and for the historical tradition, legal language is characterized by being always very formal. This formality is the way it has to reflect the importance and significance of the legal system (Biel 2010: 6). To achieve this formal language, it has the characteristics we will define below; we want to mention that we will only describe the features found just in those employment contracts that form part of our corpus in English and Spanish, since the legal language is such a broad aspect, that its complete description would be part of a bigger task or essay.

A.1. Spanish employment contracts

A.1.1. Lexical features in Spanish employment contracts:

- Legal archaisms: There is a lexical and syntactic gap between legal language and standard language that makes difficult to understand for the average citizen. This distance is in large part, due to the use of archaisms. While the language we use commonly is developing at a staggering speed and we leave behind certain words and expressions, the language of law is still using archaic language, which is obsolete and also unknown to a person who is unfamiliar with the field of law, as *ventura*, *arbitraje de equidad* or *cláusula compromisoria*.
- Word formation using prefixes and suffixes: legal language could also accept word formation. Among them, we want to highlight the suffixes *-able* (*aplicable*), *-ible* (*atribuible*), *-aje* (*arbitraje*), *-ario* (*societario*, *disciplinario*, *honorarios*), and *-ativo* (*organizativo*), because of their Latin origin.
- Redundancy with the use of binomials and trinomials: Another of the most representative characteristics of the language of law is the use of repetitions instead of synonyms. The need for accuracy in the legal field makes the writings redundant. Undoubtedly, the accuracy they use, make them repeat terms and concepts which results in a decrease of the quality. Consider a writer: He will want to give literary quality to his works by using synonyms. The author gives more importance to style. In our corpus we have found the next sentence as an example of redundancy, which is consisted of a word that accompanies other

word and both have a similar meaning: “(...) *de profesión u oficio auxiliar de bus (...)*”. Also, the binomial *daños y perjuicios* appears in some of our contracts.

A.1.2. Discourse features in Spanish employment contracts:

- Use of sentences with imperative tone: “(...) *la cual deberá ser cumplida (...)*”, “(...) *el empleador pagará mensualmente (...)*”, “*Es obligación del auxiliar (...)*”.
- Use of pronominal impersonal forms with the particle “se”. We have found “*Se deja expresa constancia que (...)*” and “*Se establecen como obligaciones esenciales del Trabajador (...)*”.

A.1.3. Syntactic features in Spanish employment contracts:

- Fixed expressions: Legal writings have an established organization that perform an essential demarcative role for citizens and organizations. In our Spanish subcorpus of employment contracts, we have found the following stereotyped formulas: “*El trabajador se compromete y obliga*” or “*Para constancia, previa lectura*”.
- Tendency to nominalization: instead of using the verb form, Spanish contracts tend to use the nominalized form, with a preceding or a following verb. The next example uses the suffix –miento to create the nominalization: “*En caso de incumplimiento, deberá (...)*”.
- Syntactic constructions: We noticed a particular structure with the use of “que” at the beginning of the sentences which encompass a specific term or agreement.
- Use of non-conjugated forms, as gerund: These non-conjugated forms produce an effect of impersonal writing, but actually, it is addressed to a specific person. “*En anexo de contrato se debe especificar las labores a que se compromete el trabajador como aprendiz y las materias que el empleador se compromete a impartir, debiendo firmarlo ambas partes, quedando con copia el aprendiz.*”.

- To emphasize the precision and avoid confusions, Spanish employment contracts hold long sentences with subordinate sentences, opposing sentences (exceptions within the rules) and disjunctive clauses (repercussion of the rules on two or more possibilities):
- Use of –adverbs ending in *mente*: *legalmente*, *respectivamente*, *mensualmente*, *igualmente*, *personalmente*, *básicamente*, *expresamente*, *individualmente*, *obviamente*, *conjuntamente*, *efectivamente*, *permanentemente*, *debidamente*, *libremente*, *estrictamente*, *precedentemente*, *colectivamente*.
- Excessive use of anaphoric elements to make reference to something aforementioned: “*dicho instrumento*”, “*dicho acuerdo*”, “*lo dicho*”, “*lo anterior*”, “*de acuerdo a lo establecido*”, “*del referido contrato*”.

A.2. English employment contracts

A.2.1. Lexical features in English employment contracts:

- Formal language: employment contracts and law in general tend to use formal words to differentiate it from ordinary language. Most of the formal words come from other languages, above all, Latin and French. In our corpus we have found a lot of very formal words. Due to their frequency, among others, we highlight the use of *commence* and *commencement* instead of *begin* or *start*; the use of *execute* instead of *sign*; and the use of *terminate* instead of *end*, and *make* instead of *do*. One of the words that appear more frequently in English subcorpus and comes from the Latin, is *cause*.
- Legal archaisms: as it occurs in Spanish employment contracts, it can also be found archaic words in Legal English. Most of them are formed by adverbs plus a combination of particles, as *hereto* (adverb *here* plus particle *to*) in reference to the document itself, or the use of *whereof* in reference to *what*. Other archaic forms that appear in our English subcorpus are *hereinafter*, *hereunder* or *hereby*. Archaic language difficult the understanding of the writings but it is necessary to apply precision to them.

- Loanwords: They are used to express legal concepts accurately. An example from our corpus is *force majeure*.
- Redundancy: above all using of two synonyms together to remark the importance of the concepts, as *state and condition*, *duties and responsibilities* or *pay and indemnify*.
- Binominal and multinominal elements: accompanied with the use of coordinators to create lists of two or more concepts: *Accept and undertake*; *legal and binding*; *terms and conditions*; *personal injury, cargo, and public liability*; *liabilities, claims, costs, and expenses*.

A.2.2. Syntactical features in English employment contracts:

- Long sentences: instead of simple and reduced sentences to express a single idea, English contracts “tends to put all such sequences into the form of very complex sentences capable of standing alone, by means of an array of subordinating devices” (Zhu and Wu, 2011: B). The following sentence are taken from our English subcorpus and exemplify this feature:

“ If Lessee remains in possession of the Premises with the consent of Lessor after the natural expiration of this Agreement, a new tenancy from month-to-month shall be created between Lessor and Lessee which shall be subject to all of the terms and conditions hereof except that rent shall then be due and owing at eight hundred and fifty dollars (\$850.00) per month and except that such tenancy shall be terminable upon fifteen (15) days written notice served by either party. “

- Use of passive sentences: Passive constructions prevail over active voice in employment contracts in English. These are used to draw the attention on the action but not the person. Serve as an examples the following ones: “ (...) *references to a year shall be deemed (...)*”, “(...) *said information has been or is threaten to be disclosed.*”, “(...) *substantial performance has been given by the Board or any officer (...)*”.

- Nominalization: we have found different types of nominalization. The most common are the verbs which the addition of suffix as *-ment* (e.g. *engagement*), *-age* (e.g. *storage* and *coverage*) and *-tion* (e.g. *application*, *determination*, *reduction*, *calculation*, *reorganization*, *termination*, and *arbitration*).
- The use of negatives: in English contracts the use of the negative in an endless of variants is common. The most frequent one is through the use of “non” or “not” (*non-appealable. (...) the **non-performing** parties **shall not** constitute a default of breach of this Agreement.*). But it is also usual the negative form by the addition of “*unless*” and “*except*”.
- The use of conditionals and complex conditionals: conditionals are used to explain terms, conditions or penalties. We can see the conditional in the following phrases: “*If any provision of this Agreement or the application thereof shall, for any reason and to any extent, be invalid or unenforceable (...)*”, “*(...) but only if the amendment or waiver is in writing and signed (...)*”.
- Syntactic constructions: As it happens in Spanish contracts with the particle “*que*”, the same is applied to its equivalent “*that*” in English. We found list of conditions or rules that are headed by the particle “*that*”.
- Flexible, general, or vague language: strangely enough, another feature of the language of law is the use of metaphors and ambiguities. While it is true that writers try to be clear, in certain situations also they also try to cheat or not make it clear what they are talking about.

As we have said a few lines above, the Spanish legal system, have the Roman law as a common root. It is therefore not surprising that many of the terms used even today in the language of law remain in Latin. Legal language has that conservative trait that leans in favor of Latin terms. The same happens with legal writings regulated by the Common Law. English contracts also contain Latin words, but its influence in vocabulary is highly canonic.

B.METHODOLOGY

In this chapter we will describe in detail the methodology adopted to conduct this project, whose main objective is to make a glossary of terms in order to establish the possible equivalent translations in employment contracts in English and Spanish languages.

The methodology we will follow is based on corpus linguistics. Corpus linguistics is the perfect tool to explore the linguistic characteristics we are going to study, based on the real use of language in authentic texts. We begin this section by explaining what corpus linguistics is. Since our work has to start by the compilation of a corpus to obtain term candidates, we will explain what a corpus is, and then we will describe and define the different types of corpus we know, according to a classification based on general parameters. Once we have explained these concepts, we will describe the two types of approaches to corpus linguistics and also the reason for our choice from one of these approaches.

These principles allow us to frame our own corpus: a comparable corpus in Spanish and in English languages. This project encompasses lexicography, terminology and translation; and its conclusion should fit perfectly with the rules that govern each of the two languages and also be accepted by the members of the specialized community.

B.1. Corpus linguistics

As I said before, this project is based on the analysis of a specialized language within a specific field. Corpus linguistics is defined by McEnery and Hardie (2012: v) as “the study of language data on a large scale –the computer-aided analysis of very extensible collections of transcribed utterances or written texts”. This definition can be complemented by Bowker and Pearson’s, who define corpus linguistics as “an approach or a methodology for studying language use. It is an empirical approach that involves studying examples of what people have actually said, rather than hypothesizing about what they might or should say” (2002:9). This is related to descriptive grammar. Descriptive grammar studies how speakers and writers use their own language, providing us a systematic study of the current use of a language and its structures.

Unlike this, prescriptive grammar guides us by means of imposing rules to speak properly. In the past, researches who worked with the use of language were based on intuition. Today, with the appearance of corpus-based studies, the researchers are based on objectivity. Researchers are based on what speakers and writers actually say (the language in use) and not on what they believe that speakers and writers say or must say.

But it is necessary that we pay attention to objectivity rather than to intuition, because the use of an specialized language is always very complicated, particularly for people who are not experts; this is really clear in this project if we consider that it is based on employment contracts from the field of law which uses not only a specialized language, but also complicated and repetitive structures. So, we must lay aside intuition to study the language of law in an objectively way. Legal language is arcane and intricate because the law itself is arcane and intricate, and we cannot simplify it using another language because as Mellinkoff (1963: vi) says, “the law is a profession of words”. This is why the purpose of this paper is not to correct people whose field is the law, but to try to study and understand the vocabulary used by them to subsequently find suitable equivalents collected in the glossary.

B.2. What is a corpus

There are many definitions of corpora given by the most prestigious linguistics, but without a doubt, the definition which best suits this project is the one given by Sinclair: “A corpus is a collection of pieces of language text stored in electronic format, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.” (Sinclair, 2005). Sinclair uses the word “pieces” because it is a neutral word and also because “some corpora still use sample methods rather than gather complete texts or transcripts of complete speech events” (Sinclair, 2004). All text must be in the same format. For easiness and speed handling, it is logical that they are in electronic format. One of the criteria that should have a corpus is representativeness (representative of a language or a language variety). There are different degrees of representativeness. The highest degree of representativeness, according to Renouf and Kehoe (2009: 223), includes "the use of the competence of expert corpus linguists in determining the number and relative

importance of genres, the appropriate degree of their sub-classification, and the size of the text sample to be included ".

There are two reasons why we have chosen this definition. The first one is because it allows us to compile texts that are not necessarily excerpts of natural conversations. The second one is because it also allows us to collect texts that may be a variation of the general language, that is to say a specialized languages. This last fits with our purpose of compiling a specialized corpus because the language of law is not the common one which we use in everyday life. This way, we will compile a corpus of employment contracts, which will contain a specialized language that differs from the common one we use daily, but belongs to our own language. There are lots of corpora composed of conversations in real situations and oral and written general corpus, but a specialized corpus is not out of the shell. If this is difficult, the obstacles in finding a specialized corpus on a subfield as restricted as employment contracts, increases much more. For this reason, it is necessary compile our own corpus.

B.3. Types of corpora

There are numerous types of corpora classified according to specific characteristics. The following classification is sorted from the most general to the most specific characteristics. Authors like Sinclair (1996) or Torruella and Llisteri (1999) have proposed different classifications based on various criteria, although in practice it is not always clear and there are other subtypes of corpus. Generally, the main parameters to classify corpus are the following:

- According to the format of language
- According to the specificity of texts
- According to the number of languages that comprises the corpus
- According to the time period covered by the texts
- According to the process that has undergone the corpus
- According to the prescribed limits

- B.3.1 According to the format of language

The first thing we have to do is consider one important matter: choosing the type of texts that we will include in our corpus according to the format, in other words, the mode. According to the mode (Dash, 2008: 55), we can distinguish three types of corpus: written corpus, speech corpus and spoken corpus.

1. A written corpus is formed by data originating from written, publishing and electronic sources.
2. A speech corpus compiles formal and informal talks, conversations, dialogues, lectures, or even monologues. It excludes media texts. Examples of this type of corpus are the *Speech Corpus of English* and the *American Speech Corpus*.
3. Finally, a spoken corpus is an advanced variant of speech corpus. It is a corpus that compiles texts of spoken language, only modified by the phonetic transcription, as manner of annotation. The speech corpus differs from spoken corpus in the formation and mode. In spoken corpus, samples are obtained from spoken texts and later transcribed to be collected in written form (Dash, 2008: 55). An example of spoken corpus is the *Lancaster/IBM Spoken English Corpus*.

- B.3.2 According to the specificity of texts

The second distinction has to do with the language used. We consider this criteria a good way to start concretizing a corpus. This first classification consists of two ranges:

1. The most general would be a general corpus, which is a corpus that contains different types of texts based on the general language.
2. On the contrary, we can reduce the range with a specialized corpus (corpus that contains texts of a specific field and a specialized terminology).

- B.3.3 According to the number of languages that comprises the corpus

After this first classification, we consider we must choose the language of our corpus. Considering all possibilities, we can build monolingual corpus, bilingual corpus and multilingual corpus (McEnery et al., 2006: 8).

1. The former is formed by texts in the same language. Monolingual corpora are useful to study language evolution in a particular period or in multiple periods.
2. If we want to compare one language to another, it is best to compile a bilingual corpus with a balanced volume of texts in both languages.
3. The last type is the multilingual corpus, constituted by corpora in more than two languages.

Within these last two categories, there are three subtypes of common corpus: parallel corpus, comparable corpus and corpus of translations.

1. A comparable corpus is a corpus containing components that are collected using the same sample frame and similar balance and representativeness (McEnery, 2003: 450). This means that the comparable corpus contains similar texts in two or more languages in order to compare them both. These texts have to belong to the same genre and the same register. A comparable corpus has the advantage that the texts in the source language are not influenced by their texts in the target language. This means that their textual forms, collocations or style, are presented in their natural form.
2. Other widely corpus used is a parallel corpus. A parallel corpus contains texts and the translations of those texts, into another language (or languages) with the purpose of comparing both (or them). This comparison allows the “double-checking facilities for translation equivalents” (Dash, 2008: 76). While for a comparable corpus the sampling frame is essential, for parallel corpus is not relevant because the texts have their exactly translations. This type of corpus is very useful to contrastive studies and to be more precise with regard to terminology and phraseology, being hardly recommended for these purposes. A great advantage of the parallel corpus over the comparable corpus is that there is

a fixed semantic link between the text and its translation, therefore, the equivalents are the appropriate ones. An example of parallel corpus is the Canadian Hansard Corpus. It is a bilingual corpus of the discussions that occur in the Parliament of Canada in both French and English languages (some of the official languages of the country).

3. Corpus of translations: it is a corpus which contains translation equivalents into different languages. Normally, it includes books or parts of a book that were translated into several languages, as the Holy Bible. With this type of corpus we can compare the different encryptions in the languages. One of the best examples is the corpus created by the European Commission, which contains around one million sentences and their translations in the official languages belonging to the European Union.

- B.3.4 According to the time period covered by the texts

The latter leads us to a division in accordance with temporality. If we focus on temporality, we can distinguish two types of corpus.

1. The first one is a synchronic corpus, in which all of the texts are collected within the set time frames.
2. The second one is a diachronic corpus, which includes texts of various periods of time.

- B.3.5 According to the process that has the corpus undergone

Before proceeding to the next type of corpus, we must define the term “corpus annotation”. We understand "corpus annotation" in this work as the additional linguistic information, which records the usual analysis in a systematic manner, collected over the years of study. This allows us to distinguish between an unannotated corpus and an annotated corpus.

1. Unannotated corpus are made up of plain texts without additional information. Samples “are in a simple format without any additional information tagged to them from external sources” (Dash, 2008: 73).
2. Annotated corpus includes tags that contain extra information recorded by researchers. This means that, when a corpus is annotated, is because it has already been scanned in any way. The most common additional information that can provide the linguistic annotation of a corpus is called part of speech annotation (POS). Part of speech annotation describes the grammatical category of a lexical item, including tense, number and genre (e.g. plural common noun, possessive pronoun, feminine adjective, past tense verb). In addition, this type of annotation has the feature of being able to differentiate between homophonic words. For example the word "not" and the word "knot" are homophones, this means that they have the same pronunciation but are spelt differently, and their grammatical categories are also distinct (“not” is an adverb and “knot” is a noun). Other types of annotation are:
 - Lemmatization: it removes the reflective endings to expose the base form of a word, known as lemma (Lu, 2014: 55)
 - Syntactical parsing: It is the method by which the elementary morphosyntactic categories are converted into complex syntactic relationships with others.
 - Semantic domain classifications: a code is assigned to every word from a thesaurus-type dictionary, as WordNet.
 - Coreference: is the linking of all individual allusions when they "refer to the same entity" (Hirschman et al., 1997).
 - Zero corpus: it is a type of annotation which, through tags, provides extra-linguistic information. This will be explained in detail in section C.

- B.3.6. According to the prescribed limits

After this classification we must consider one important matter. We need to consider if we want a sample corpora, or a monitor corpora.

1. Sample corpus, also known as static corpora, contains a fixed number of samples of texts.
2. Monitor or dynamic corpora are developed according to researcher needs and they are constantly updated (McEnery and Wilson, 2001). It is “a large corpus that is not static and fixed but that is constantly being updated to reflect the fact that new words and meanings are always being added to”, according to Meyer (2002: 15).

B.4. Approaches to corpus linguistics

There are two types of approaches to corpus linguistics: a corpus-based approach and a corpus-driven approach. Corpus-based approach is primarily used to “expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study” (Tognini-Bonelli, 2001: 65). This means that the corpus-based approach uses data from a corpus to verify, reject, or define more precisely a theory or a hypothesis. The corpus-driven approach works differently. This type of approach creates hypotheses arising from the findings of the corpus, and not the other way round. Linguists that use the corpus-driven approach “come to a corpus with no preconceived theory, with the aim of postulating linguistic categories entirely on the basis of corpus data [...], however, theories should not be discarded altogether, because “intuitions are essential in analysing data” (McEnery et al., 2006: 10).

While it is true that the main goal of corpus-based studies is to collect texts of speakers in their natural states for subsequent interpretations, it is also important to be able to interpret texts which use a specialized language. This is important because specialized languages are natural in their respective field of action/activity. Natural texts seem to cover only general language. For legal experts, including translators and technical writers, specialized texts and vocabulary are natural texts because they are very familiar

with them. And this specialized language is also governed by concrete patterns in terms of syntax and or grammar or in terms of the genres used. Patterns are useful to recognize a text and classify it into a general type or a more detailed subdivision.

Taking into account the previous point of view, we realize that it is related with the first of the characteristics with which Biber et al. (1998:4) carried out the analysis based on corpus. Biber, Conrad and Reppen, (1998:4) described the corpus-based analysis with the four following features:

“

1. It is empirical, analyzing the actual patterns of use in natural texts;
2. It uses a large and principled collection of natural texts, known as a ‘corpus’, as the basis for analysis;
3. It makes extensive use of computers for analysis, using both automatic and interactive techniques;
4. It depends on both quantitative and qualitative analytical techniques.”

According to them, these characteristics have a main advantage nowadays: the use of computers. Data management through computers allows us to save time and to avoid possible human errors. These authors highlight the trustworthy analysis that computers provide and the possibility “to identify and analyze complex patters of language use, allowing the storage and analysis” (1998: 4) of large databases.

The first feature of this corpus-based analysis has been discussed previously.

The second characteristic reveals that a corpus is essential to collect large a number of samples in order to work with them all together at the same time and to study them. This allows us also to work with each of the samples separately.

The next feature is related to the use of computers. Due to the large size of the corpus, the use of computers is crucial. Computers are an inexhaustible tool that enables the programming of different configurations for the various specific studies of corpora. Not only it allows us to analyze the corpus specifically, but it also has the advantage of accelerate the analysis process, thereby obtaining more data, and therefore, the possibilities further interpretations, judgments, and conclusions. The use of corpus management software allows us to work in two different ways. On the one hand we can

work automatically, using the basic functions of the program concerned; and on the other hand we can work through an interactive way, by programming commands of the program we use to study the corpus and wait for the results.

The fourth and final feature refers to analytical techniques, also called corpus analysis. The first one is the qualitative analytical technique, in which a linguistic phenomenon, the behavior of a word, or a group of words is described. And the second one is the quantitative analytical technique, in which the frequency of linguistic phenomena in the corpus is measured. Depending on the frequency with which they appear, more complex statistical models explaining the evidence found can be built.

Both techniques provide information to corpus studies. Quantitative analysis provides statistical data used to know what terms there are in the corpus and which ones are representative of the language under study or a particular study. Qualitative analysis requires the interpretation of the data and provides information on the language in use. The choice of one of these two types of analysis does not imply the exclusion of the other option; in fact, it is very common nowadays the combination of both. Or what is the same, but in words of Bunge (1995: 3), "there can be no opposition between quantitative and qualitative methods, since quantity and quality are mutually complementary rather than exclusive. Indeed, every quantity is either the amount of a collection of items sharing a certain quality, or the intensity of a quality. Hence, in the process of concept formation, quality precedes quantity." Quantitative analysis could be applied to explain why patterns are used through the study of examples, because "a crucial part of the corpus-based approach is going beyond the quantitative patterns to propose functional interpretations explaining why the patterns exist. As a result, a large amount of effort in corpus-based studies is devoted to explaining and exemplifying quantitative patterns" (Biber et al., 1998: 9).

C. CRITERIA FOR COMPILATION OF THE CORPUS

The criteria for designing a corpus depends on the needs of researchers, the aim of the study, the type of corpus, or even the recipient of the corpus. According to Bowker and Pearson (2002: 54), the criteria for compiling a corpus "include size, text extracts vs full texts, number of texts, medium, subject, text type, authorship, language and publication date". Certain aspects such as the languages that we will use, the type of document we want to study, and the genre which the texts belong, are already mentioned in this work, but hereafter, we will describe our corpus using the rest of the criteria mentioned in this paragraph.

We need to take into consideration that there are two types of criteria for linguistic analysis of a corpus: internal criteria (linguistic) and external criteria (non linguistic). "The internal criteria are those which are essentially *linguistic*: for example, to classify a text as formal/informal is to classify it according to its linguistic characteristics (lexis/diction and syntax)." (Atkins et al., 1991: 8). External criteria are basically the genre of the text, the context, and the purpose or the origin of the text. The external criteria of a text must be settled without the requirement to read that text.

In this project we will use a comparable corpus, that is to say, contracts originally written in English and in Spanish. We discard the compilation of a parallel corpus because we want the texts written in the source language and without their own translations. We believe that the translations of these texts minimize the naturalness and authenticity to the terms, fixed phrases, and legal expressions.

Employment contracts belong to the same genre and register in Spanish and English, We are building a small corpus, because the most common terms we can find in a small corpus, will be the same than in a large corpus, since we are talking about proportions (See the concepts of closure and saturation below). Thereby, the first thing we are going to do to carry out this project is the collection of data, more specifically, we are going to search employment contracts in a size deemed representative, about 50,000 words per language. The representativeness of a specialized corpus can be regulated by degree of "closure" (McEnery & Wilson, 2001: 166) or "saturation" (Belica, 1996: 61-74) of it. The degree of closure or saturation can concern to a variety of language or a specific language characteristic. This implies that the characteristic appears to be limited or have

hardly any modification beyond certain parameters. To define the concept of “saturation”, we will use the next quote from Teubert (2000:1.16): “To measure the saturation of a corpus, the corpus is first divided into segments of equal size based on its tokens. The corpus is said to be saturated at the lexical level if each addition of a new segment yields approximately the same number of new lexical items as the previous segment, that is, when the curve of lexical growth is asymptotic, or flattening out. The notion of saturation is claimed to be superior to such concepts as balance for its measurability”. In this work we define token as each instance of a type of word in the text. The frequency, therefore, is the number of repetitions (token) of the same word, or type of word in a corpus.

Representativeness is the main characteristic that has corpus studies to create successful outcomes. In our case, the representativeness plays a fundamental role when deciding between the compilation of full texts or the compilation of text extracts. If we want our corpus of employment contracts to be representative, we must compile full texts although they differ in size. If we give priority to the size of the texts and we shorten them, we risk losing some of the information contained in them and the final results obviously vary, and thus, they will lose credibility. Therefore, it is obvious that we will compile full employment contracts.

The issue of representativeness is somewhat controversial, since the experts fail to agree on what amount of texts is enough to represent a language or part of a language. Experts know that the representativeness consists of quantity and quality, but according to Leech, the theory of representativeness "must be regarded largely as an act of faith" (1991: 2). Dealing with quality, the main problem is to include texts that come from sources that are not reliable (Gelbukh et al., 2002: 10) diminishing the quality of the corpus. To avoid the low quality of the texts, it is indispensable to establish a useful collection of criteria to compile the corpus. Seghiri (2014: 86) proposes a “compilation protocol”, formed by the next four steps:

1. **Finding data:** “will consist in searching relevant documents on the web” (Seghiri, 2014: 86). Seghiri divides the searching into two. The first one is “institutional search”, provided by web pages of “international companies, organizations and institutions” (2014: 86) because of their “high standard of

quality and reliability” (2014: 86). The other type is the “thematic search” (2014: 86), which is accomplished “by using the key word searches on good search engines” (2014: 86). Most of search engines allow limit the search “to a specific domain” (2014:86).

2. **Downloading:** we can do it manually or by means of “programs like BootCaT”.
3. **Formatting:** it is difficult to find the samples in the format used to compile a corpus: plain format (.txt). The most common formats we usually find are PDF (.pdf) and HTML (.html). So, we need to convert them into plain format (.txt) manually or by a program.
4. **Storage:** this is the last step which “consists of saving the documents that have been downloaded, correctly identifying and arranging them”. We will use specific criteria to codify our documents to conform a corpus zero, as we show below.

Once we have this information, we will classify each document by creating a corpus zero. A corpus zero is a type of annotation (vid supra B.3.5) that includes extra-linguistic information. It may contain information about the author, source, type of document, the language in which it is written, the year in which the text was written, etc. This annotation is encoded through slots. Each slot may consist of abbreviations and numbers. A corpus zero can hold as many slots as desired or required, depending on the amount of information we want to encode, because the intention is to identify the texts. Thereby, it makes easier to locate any file containing the original corpus in case of problem, correction, search or improved of documents. The coding of our corpus zero will have 14 slots that provide information as follows:

- The first three slots are numbering characters.
- The fourth and fifth slots refer to the type of document. In this case are employment contracts (CT).
- The next two slots provide supplementary information about sources.
- The next slot indicates the source of the document. In this project we will find web sources (W).
- The next four slots will be occupied by the year in which the contract was made.

- And the last two slots indicate the language of the contract. In this corpus, it could be Spanish (ES) or English (EN).

In addition our corpus will also be a dynamic corpus, in which we can include or throw away the documents we want to, according to our needs.

As there are no comparable corpora of employment contracts available, we will start from scratch, collecting different examples of contracts, which are accessible through the web over the last twenty years, both in Spanish and English. This leads us to define the medium we use for the extraction of the texts and the date of publication of these ones. The medium that we will use to collect the texts will be the Internet because, currently, many authors support the origin of the corpus on The Internet. As Sharoff says "The Internet is a natural source of linguistic data, providing an abundance of texts of various types in a large number of languages. These texts are already in electronic form suitable for corpus studies, either as downloadable pages, or as a resource to be searched using search engines" (2006: 63). The Internet allows us to compile large number of large texts rapidly, which was previously unmanageable for a single researcher. To make a good and useful glossary of specialized legal terms, it is essential to take as model real texts in use today. In other words, real employment contracts of the last twenty years. The Internet allows us to compile large number of large texts rapidly, which was previously unmanageable for a single researcher.

D. RESULTS AND ANALYSIS

After finding employment contracts, downloading them, formatting into plain format and storing them, the next step is acquiring a software program. We have chosen the AntConc software. The reasons why we have chosen this program are its ability to perform different analysis rapidly and its free online access. The tools that we need for our analysis are mainly two: wordlist and a concordance. The Wordlist tool creates word lists from the corpus that we select. The concordance tool generates concordance lines in general or keyword lists with the surrounding text. The combination of both will be very useful for our objectives. Wordlists are useful for quantitative analysis creating a list of the most common words occurred in a corpus. Concordance contributes to qualitative analysis because the researcher has to interpret the data and select the actual use of the words from the wordlist.

D.1. English Wordlist

Taking into account that we are dealing with a bilingual corpus, first of all, we generate a list of words sorted by frequency of occurrence in employment contracts in English. This list gives us roughly information about our subcorpus in English, such as the most common words or the position the terms we seek for our project are. To obtain a wordlist we load each of the text files (.txt) containing an individual employment contract. When all files are loaded, we select the tab WordList and, after checking all the features, we select the Start button. Figure 1 shows the list of the most frequent words in the English subcorpus.

In this screenshot (Figure 1), we can see the position, the number of times they are repeated and the words themselves. Most of these programs generate noise. We define noise as different extracted data that are not useful to us for our research purposes, such articles and pronouns. In our case, we want a terminological word list, however, the first word with semantic content is at position 11 (*employee*). Therefore, all words before that term are noise, and what is between that term and the next term, is also noise (see Figure 1).

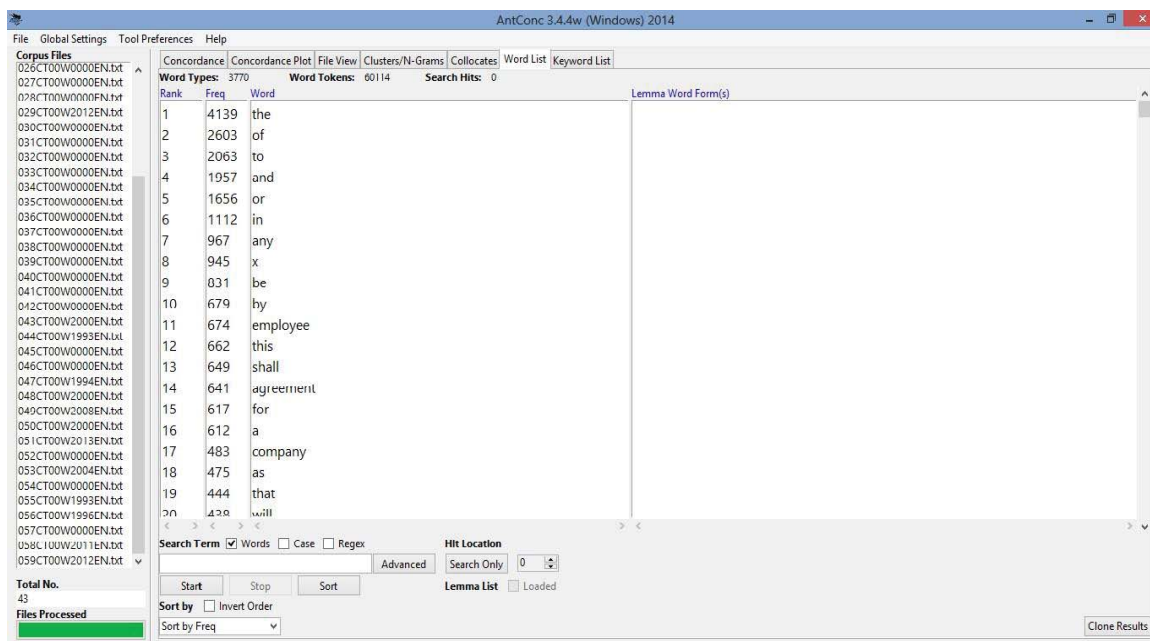


Figure 1. WordList English subcorpus.

One way to remove noise available for AntConc is by *stopword lists*. These are exclusion listings containing different types of grammatical words. Most of these lists are made of definite and indefinite articles, numerals, adverbs, and words of general content. We need to be very careful with the words of general content, because in our corpus, they may have a specialized nature.

Stopword lists may be obtained on various websites, previously created by researchers according to their needs. We have decided to use the website <http://www.unine.ch/info/clef> to download one *stopword list* to remove from the general list the 1000 most common English words. These words would be equally the most common ones in any corpus of English language and do not give us relevant data. We study the *stopword list* by introducing each suspect word of interest (as *state*) in a legal online dictionary (<http://dictionary.law.com/>). The words we find in the legal dictionary, are those that remain on the *stopword list*. In addition, we include all articles, numerals, adverbs, determiners, pronouns, prepositions, and a list of the most common nouns in English, to be more precise and to eliminate more noise.

D.1.1. Analysis of data

N°	WORD	FREQUENCY	N°	WORD	FREQUENCY
1	The	4139	11	Employee	674
2	Of	2603	12	This	662
3	To	2063	13	Shall	649
4	And	1957	14	Agreement	641
5	Or	1656	15	For	617
6	In	1112	16	A	612
7	Any	967	17	Company	483
8	X	945	18	As	475
9	Be	831	19	That	444
10	By	679	20	Will	438

Table 1. 20 most common word in English employment contracts.

The previous table includes the 20 words that most frequently occur in our employment contracts subcorpus in English before using the stopword list. The only words on this list that have lexical categories (that have meaning by themselves) are the nouns *employee*, *agreement* and *company*. The high presence of these three words clearly indicates the subject matter of the corpus, referring to the essential parts of the fact: the contract (*agreement*), the worker (*employee*), and the contractor (*company*). Regarding the category of verbal forms, we find three representations:

- The modal verb *shall*: This is not a surprise to us, since *shall* is the dominant modal verb in legal English, used "in the imperative sense to impose a duty or obligation on the legal subject to whom it refers" (Bhatia, Candlin and Gotti, 2003: 347).
- Infinitive of verb *to be*: In the presence of this verb in its infinitive form, we can see how in the language of employment contracts abound non-conjugated verb

forms of the most used auxiliary verb. Their presence is also high because they usually are used with modal verbs, being *to be* the main verb.

- The modal verb *will*: If we compile a general corpus of law, we could say that *will* is the way that law gives to the word testament. However, in this case, we can not treat it as such, since employment contracts have no direct relationship with the last wishes. Will behaves as a form of obligation, usually descriptive of the work stated in the contract.

The rest of the words that conform this list, are words with functional content, which have no semantic content if they are alone. These words are articles (*a, the*), connectors (*and, or, as*), prepositions (*of, to, in, for, by*), adjectives (*any*), pronouns (*this, that*) and adverbs (*as*).

The result of the list of most frequent words after using our *stopword list*, can be seen in Figure 2.

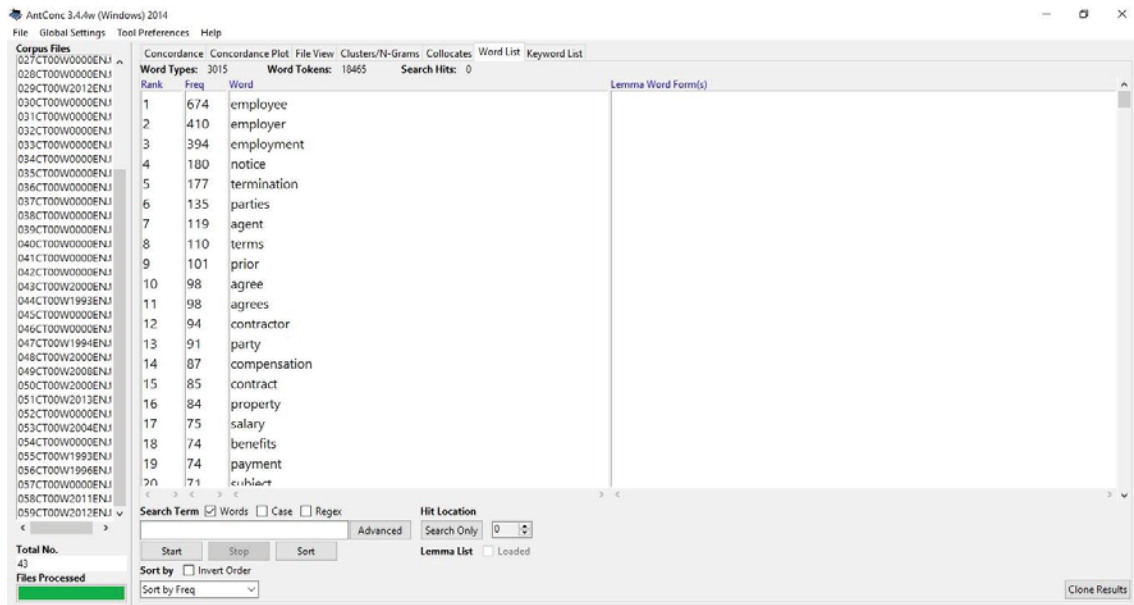


Figure 2. WordList English subcorpus without noise.

The results are closer to our expectations than before, since noise disappeared after using the *Stoplist*. All words that appear on the top twenty words are widely used in the law, some are terms and others are not. To resolve this issue, we analyze each of the words using two tools. The first one is an AntConc option that allows us to see the words from the list in context. Through the use of this tool, we can see the concordance

of a word by clicking on it, showing the words that appear in its usual linguistic context and it also gives us exactly which file these extracts are taken from. The second is using a specialized dictionary, to categorize them as terms or as words used in the language of law. Before analyzing the data, we need to validate and verify that actually the words of the list are in accordance with the definition of *term*.

Once we have our list with the term candidates and we have checked their meaning in context, we proceed to double check the list by using the most reliable online dictionary in English (The Oxford dictionary), and two legal dictionaries online (dictionary.law.com and legal-dictionary.thefreedictionary.com/). If the word we were studying did not appear in the legal dictionaries, it was probably not a term. We study the list of the most frequent words to reach twenty terms.

Our candidate terms are the first forty one words that appear in AntConc results. If we did not obtain 20 terms from the candidates, we continue analyzing the following words in order of appearance. The candidates are:

1: employee	2: employer	3: employment	4: notice	5: termination
6: parties	7: agent	8: terms	9: prior	10: agree
11: agrees	12: contractor	13: party	14: compensation	15: contract
16: property	17: salary	18: benefits	19: payment	20: subject
21: cause	22: confidential	23: provisions	24: act	25: due
26: effect	27: law	28: writing	29: conditions	30: terminate
31: expenses	32: good	33: arbitration	34: hereunder	35: unless
36: breach	37: rights	38: hereof	39: perform	40: owner
41: premises				

Table 2. English candidate terms.

We begin with the first word of our list: *employee*. Specialized dictionary defines this word as *a person who is hired for a wage, salary, fee or payment to perform work for an employer*. Of course, this is a specific and important term in employment contracts, and it is our first term.

The second candidate is *employer*. According to the Oxford Dictionary, employer is *a person or organization that employs people*. A similar definition is provided by the specialized dictionary (*a person or entity which hires the services of another called a principal in the law of agency*). This term is as important as employee, and its or second term.

The third candidate is *employment*, defined in the specialized dictionary as *the hiring of a person for compensation*. This is a specific term of employment contracts and it is also a term.

We continue with *notice*. The Oxford dictionary gives us as the first definition as follows: *The fact of observing or paying attention to something*. This is the first definition. However, within the second definition, we find the following: *A formal declaration of one's intention to end an agreement, typically one concerning employment or tenancy, at a specified time*. This indicates that the second use is as a specialized term. The next step is to search for that word into the legal dictionary, to confirm that it is really a term. Dictionary.law.com defines notice as *information, usually in writing in all legal proceedings, of all documents filed, decisions, requests, motions, petitions, and upcoming dates. Notice is a vital principle of fairness and due process in legal procedure and must be given to both parties*. To check whether this specialized definition fits perfectly with what we are looking for, we will use the Concordance tool of AntConc. The Concordance tool allows us to see the collocates of the word itself and the files where it occurs (see Figure 3). According to the results, we can say that *notice* is a term used in most employment contracts in our subcorpus (it is very important to pay attention to repetitions of a term that occurs only in a single contract, because it could be a compilation problem) and the concordances fit with the specialized definition.

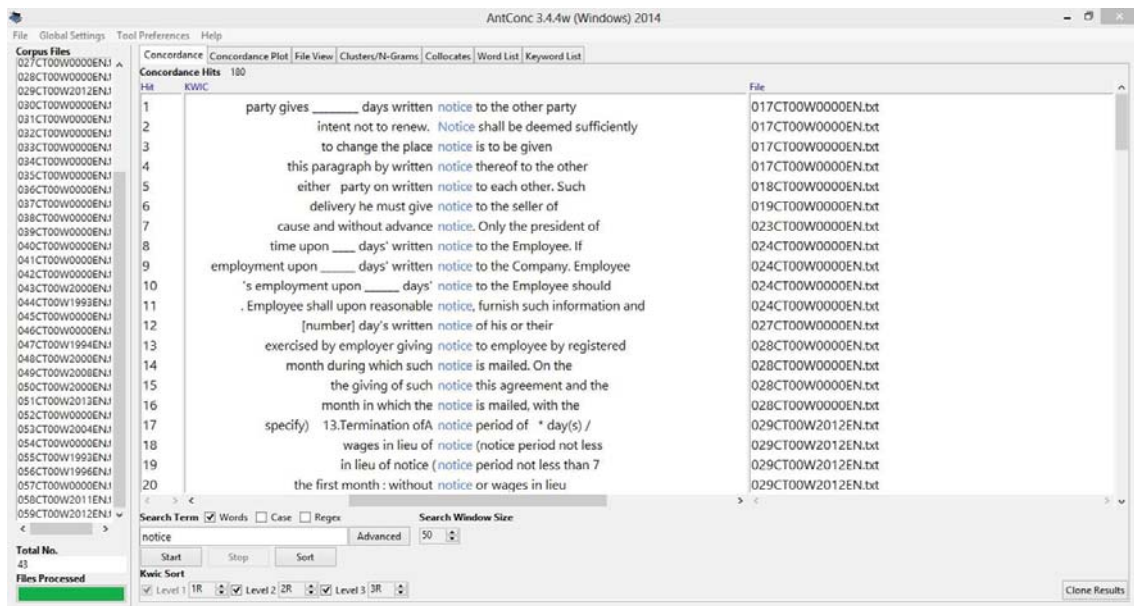


Figure 3: Concordances of *notice*.

Accordingly, we take for valid the definition of term for the word *notice*.

The second word of our list is *termination*, which shares lemma with *terminate*, in position twenty-five, reaching a total of 240 appearances. The Oxford dictionary says that *termination* is "chiefly North American" and defines it as *an act of dismissing someone from employment*. However, when looking for this word in the specialized dictionary, *termination* (of a contract) *signifies the process whereby an end is put to whatever remains to be performed thereunder. It differs from Rescission, which refers to the restoration of the parties to the positions they occupied prior to the contract*. Using the concordance tool, we can see that this word is used to refer to the termination of the agreement or contract (see Figure 4). So this word is a term.

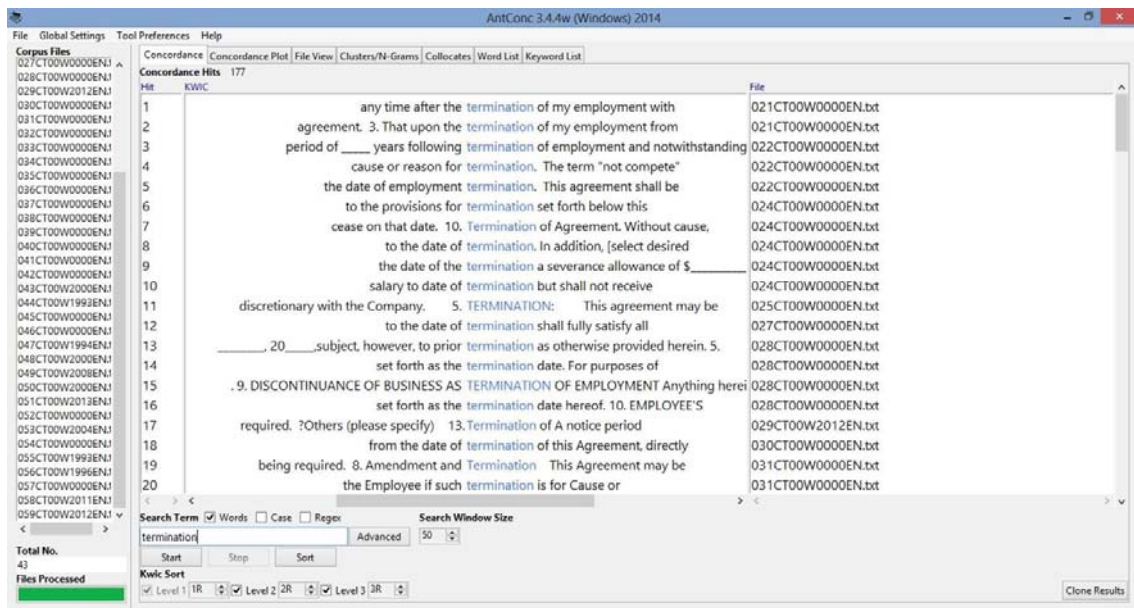


Figure 4: Concordances of *termination*.

The next word is *parties*. *Parties* is the plural form of the word *party* (word that appears at number eight in our wordlist). These words share the same lemma, and we must join them as a unit, reaching a total of 226 appearances. The third definition of this word is *a person or people forming one side in an agreement or dispute* in the Oxford Dictionary. Normally, general dictionaries, as it is the case of the Oxford Dictionary, include concise and abbreviate definitions of words and terms. On the other hand, due to the restrictions terms include, specialized dictionaries offer more accurate, precise, and hence longer, definitions than general ones. The Law Dictionary provides a definition of parties that fits perfectly with its concordances. According to the analysis, we can say that *party* is a term.

The next word we have to validate as term is *agent*. The Oxford dictionary definitions for this word are not very accurate. Fortunately, the specialized dictionary gives us a more precise definition (*a person who is authorized to act for another (the agent's principal) through employment, by contract or apparent authority*) of what an *agent* is in legal terms. This definition is far from what we know as *agent* in a general English. Due to this analysis, we can classify the word *agent* as a term.

The next word to analyzed is *terms*. Its concordance lines makes reference to *terms and conditions*, which is a specific term in employment contracts. According to the specialized dictionary, *term* is specific of the field of contracts and means a specified

condition or provision. And *condition* is defined in the same dictionary *as a term or requirement stated in a contract, which must be met for the other party to have the duty to fulfill his/her obligations. Terms and conditions* is also a term.

The next word on our list is *prior*. In the legal dictionary this word is described as the record of crimes which belongs to an accused person. However, this definition is labeled as slang. If we analyze this word with the concordance tool, we get some results in which this word is used to refer to something previous or something that precedes something. This fits with the definition that provides the general dictionary. Then, this word is not a term.

The next word is *agree*, which shares lemma with agreement, agrees, and agreed making a total of 239 appearances. *Agree* is defined in The Oxford Dictionary as *concur or have the same opinion about something*. This is one of the most frequent used words in English, but we decided not to include it in our stopword list because in legal language, this word is the equivalent of *contract*. *Agree* is the verb of the noun agreement, and it is a synonym of contract. It also appears on our list at number fifty with forty three occurrences, the word *Agreed*.

Contractor is the following word in the list. Commonly, a contractor is *a person or firm that undertakes a contract to provide materials or labour to perform a service or do a job* (The Oxford Dictionary), but the first legal definition says that defines it as *a person or entity that enters into a contract*. The second definition in the specialized dictionary gives us a much more detailed version of the Oxford Dictionary definition. If we look at the concordances of this word (Figure 5), we can observe that it do not always refer to the same definition. For example, in several concordances, they talk about an independent contractor, a term which can also be found in the specialized dictionary.

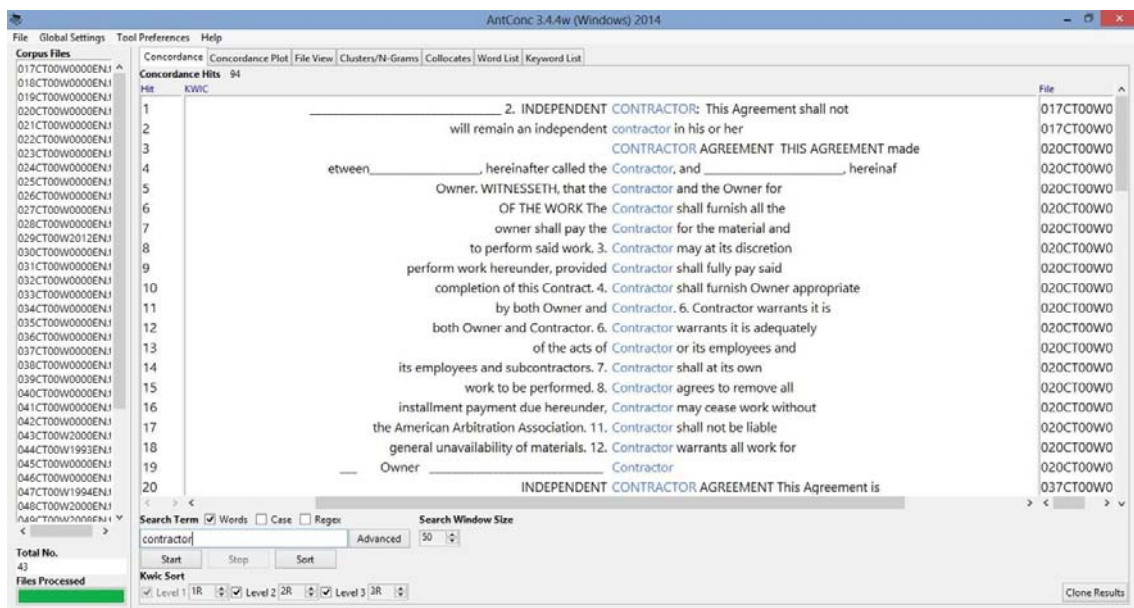


Figure 5: Concordances of *contractor*.

According to our analysis, we accept *contractor* as a legal term frequently used in employment contracts, due to their different definitions, all related to the legal language.

The ninth word of our list is *compensation*, defined in the dictionary of the English language as *something, typically money, awarded to someone in recognition of loss, suffering, or injury*. According to the specialized dictionary, is a *pecuniary remedy [...] such as worker's compensation*. It is also known as *compensation the payment for work performed, by salary, wages, commission or otherwise*. It can include giving goods rather than money. The concordances of this word show that, in almost all the employment contracts compiled, there is a section dedicated to *compensation*. This makes us think that it is an important concept in this type of documents, and thus, we decided to validate the word as a term.

The next word is *contract*. We have previously said that *agreement* is a synonym of *contract*. The Oxford Dictionary definition does not go far enough while specialized dictionary has an extensive explanation. We will classify this word as a term, but we will not include it as such in the list. We will attach it as synonym of *agreement*.

We continue with the analysis of the word *property*. In the non-specialized dictionary, this is defined as belonging or possession (something to someone). The specialized dictionary goes beyond this definition, specifying and complementing it, and also

distinguishing different types of property. To find the type of property that occurs in our subcorpus, we use the concordance tool. We found a wide variety of results. In most of them the word property is used as belonging, but we have also encountered "Law of Property Act", "real property" (*which is any interest inland, real estate, growing plants or the improvements on it*, defined by <http://legal-dictionary.thefreedictionary.com>), and "intellectual property". These specificities of the word *property* that only exist in the legal language make possible the classification of this word as term.

In the position number twelve of our list we have the word *salary*. Certainly it is a common word known by the non-specialized community, but we must analyze it to know if it has more connotations than we already know. The results of the specialized and non-specialized dictionary provide the same definition of the word. Word concordances are based on the definitions of both dictionaries, differing in secondary features as "base salary", "annual salary", and "regular salary", which do not alter the meaning of the main word. Although it is a term used in employment contracts, it has become common, so it is no necessary to include it in our list of terms, since what we want to achieve is to provide new data that supply specific knowledge. It is a so well known term that we decided not to include it in our list of terms.

We continue with *benefits*, plural of the word *benefit*. This word appears in at number thirteen, but his singular also appears on the list at number forty-four. As both words share the same lemma, we decide to join them, reaching a total of 122 occurrences. The Oxford dictionary defines *benefit* as *an advantage or profit gained from something*. In dictionary.law.com this word is related to the word contract and defines it *as any profit or acquired right or privilege, primarily through a contract*. This gives us an idea that we can be facing a very specific term in this type of documents. The concordance lines for benefits and benefit, show specifications of this privilege in different situations which may be the worker. The legal dictionary defines these particularities, which does not have the word in common usage. Therefore, we are in front of another term.

We dismiss the next word as term (*payment*) because its meaning in the legal dictionary does not vary with the Oxford one. Although it is used in employment contracts, we do not include it in our list of terms because its meaning is so well known.

Of the 71 occurrences of the word *subject*, only sixteen are not followed by the preposition *to*. *Subject to* in legal dictionary has a specific meaning (*referring to the acquisition of title to real property upon which there is an existing mortgage or deed of trust when the new owner agrees to take title with the responsibility to continue to make the payments on the promissory note secured by the mortgage or deed of trust*). We must consider *subject to* as a term.

The next word is *cause*, a word that is in the legal dictionary, but it has the same definition in the Oxford dictionary. This word appears in the legal dictionary because it comes from the Latin word *causa* and is widely used in the language of law. It could be a term in other specific languages, but not in the employment contracts.

The word *confidential* does not appear with its own meaning in the specialized dictionary. It appears next to other words that do not match with the collocates found in our corpus.

We cannot classify the next word, *provisions*, as a term, because it happens the same as before with the previous word. It is not defined in the specialized dictionary (neither in its singular *provision*).

The next word, *act*, is presented in two different forms in the corpus because of its collocates. On the one hand, we find a first definition on a specialized dictionary (*in general any action by a person*). On the other hand, we find the second definition of the same dictionary (*a statutory plan passed by congress or any legislature which is a "bill" until enacted and becomes law*). We can distinguish these two uses of the same word because the first letter of the word *act* is capitalized when legal statutes are named. Although this word appears in specialized dictionaries, its definitions do not differ from those found in the Oxford Dictionary, and it could be a term in other specific languages, but not in employment contracts.

The word number twenty of our list is, *due*, which appears in both the specialized dictionary and the non-specialized, is defined as *owed as of a specific date and expected at or planned for at a certain time*. The two definitions are similar and if we study its collocates, we can see that it use refers to cause or something that is payable. We conclude that *due* is not a term.

The next word is *effect*, that after specialized dictionary search without result, and observe its collocates, many belonging to the binomial expression *force and effect*, we can conclude that it is not a term.

The word *law* is obviously defined in the specialized dictionary (*any system of regulations to govern the conduct of the people of a community, society or nation, in response to the need for regularity, consistency and justice based upon collective human experience*) and in the Oxford dictionary as *the system of rules which a particular country or community recognizes as regulating the actions of its members and which it may enforce by the imposition of penalties*. Looking at the collocates, we have found a large number of "bylaws", according to <http://dictionary.law.com>, *a generic term for any body of regulations for conduct, including specialized rules (military law), moral conduct under various religions and for organizations, usually called "bylaws"*. However, it is not something that is typical of employment contracts, and again, it could be a term in other specific languages, but in employment contracts, it is not a term.

With the next word, *writing*, we should not waste much time because it does not appear in the specialized dictionary and its concordances refer to something that should be left in writing.

In the position twenty nine of our list is *conditions*, which is a commonly used word, and in the specialized dictionary is not defined alone. However, if we look at their concordances, we see that its meaning is the same as the common use and its frequency is given by the repetition of the binomial *terms and conditions*. This is a specific term only used in employment contracts and it is a term.

The word *expenses* is not defined in the legal dictionary, however there is an expression that appears twice, which is *out-of-pocket expenses*, and cannot be included in the list of terms for its low appearance in the corpus. In the Oxford dictionary, this word is defined as *the cost incurred in or required for something*, and it is commonly used out of law.

The next word, *good*, usually co-occurs along with different words, which make the definition more specific. We found the word alone only in the Oxford dictionary, and defined as *to be desired or approved of*. If we analyze the concordance lines, we can see that many of good goes with words like *good faith*, *good will* *good reason*, and *good cause*. *Good cause* and *good reason* must be studied together because are synonyms.

Good cause is defined as *a legally sufficient **reason** for a ruling or other action by a judge* in specialized dictionary. *Good faith* must be defined together with *good will* because they bear the same meaning, frequently used in the language of law. *Good faith* is defined separately as a term in specialized dictionary as *honest intent to act without taking an unfair advantage over another person or to fulfill a promise to act, even when some legal technicality is not fulfilled*. By virtue of its specific use, we classify these two expressions as terms, considering that in this way, *good faith* occurred 17 times (5 of them are good will) and *good reason* occurred 21 times.

The next word, *arbitration*, is defined in the legal dictionary as *a mini-trial, which may be for a lawsuit ready to go to trial, held in an attempt to avoid a court trial and conducted by a person or a panel of people who are not judges. The arbitration may be agreed to by the parties, may be required by a provision in a contract for settling disputes, or may be provided for under statute*. The Oxford dictionary gives us a very vague definition of arbitration (*the use of an arbitrator to settle a dispute*) and, after checking the concordances, we can see that it fits perfectly with the definition given by the specialized dictionary and that it is a legal term.

Hereunder is the next word from our list. Word we already know, which is very formal, archaic, and does not appear in the specialized dictionary. This word is not a term.

Unless is another word that does not appear in the specialized dictionary. It is a frequent word used in employment contracts due to its use (introduce the fact in which a previous statement is not genuine) in long formal sentences. *Unless* is not a term.

With fifty-eight occurrences in the English subcorpus, we have the word *breach*. The Oxford dictionary shows us that is related to law and employment contracts: *an act of breaking or failing to observe a **law**, **agreement**, or code of conduct*. Dictionary.law.com defines *breach* as *literally, a break. A breach may be a failure to perform a contract (breaking its terms), failure to do one's duty (breach of duty, or breach of trust), (...) the act of failing to perform one's agreement, (...)*. It is clearly a legal term.

The next word to study is *rights*, plural of *right*. The Oxford dictionary defines *right* as *morally good, justified, or acceptable and true or correct as a fact*. However, legally, it has specific connotations that we must take into account. Legal dictionary defines *right*

as *the collection of entitlements which a person may have and which are protected by the government and the courts or under an agreement (contract)*. This is specific of the employment contracts and we will include *right* in our list of terms.

With the word *hereof*, happens the same as with the word *hereunder*. It is a formal and archaic word used in legal English to refer to the document itself and it is not a legal term.

The word *perform* is commonly known as the act of *carry out, accomplish, or fulfil (an action, task, or function)*, according to The Oxford dictionary. In legal language is *to fulfill one's obligations under a contract*. It may seem the same, but studying the concordances, all of them it makes reference to duties, obligations, responsibilities, in short, the employ. Although it is a word with a similar use in general language, we decided to include *perform* in our list because it has a specific legal meaning.

Owner is the next word from our list. The Oxford dictionary defines it as *a person who owns something*, but in the legal dictionary it is defined as *one who has legal title or right to something*. *Contrary to the cynical adage: "Possession is nine-tenths of the law," possession does not necessarily make one a legal owner*. Because of this, we must discard the definition of the non specialized dictionary and include *owner* as a term, since it has a specific meaning in legal language.

The following word is *premises*. The Oxford dictionary defines this word as *a house or building, together with its land and outbuildings, occupied by a business or considered in an official context*. If we look for the same word in the specialized dictionary, we find two definitions. We reject the first one, which makes reference to *real estate (land and the improvements on it, a building, store, shop, apartment, or other designated structure)* and we observe the second, which refers to *legal pleading*. According to the legal dictionary, *in legal pleading, premises means "all that has hereinabove been stated," as in a prayer (request) at the end of a complaint asking for "any further order deemed proper in the premises" (an order based on what has been stated in the complaint)*. Because of the reasons stated above, we consider *premises* as a term.

We analyze forty-one candidate words to obtain twenty terms. They are ordered by appearances, being the first one the term which has the highest number of appearances taking into account their belonging to the same lemma, if applicable. The list of terms and its appearances are the following:

TERM	APPEARANCES	TERM	APPEARANCES
Employee (lemma)	728	Contractor	94
Employer	410	Compensation	87
Employment	394	Contract	85
Notice	180	Property	84
Termination (lemma)	240	Arbitration	59
Agree (lemma)	239	Breach	58
Party (lemma)	226	Right	58
Benefit	122	perform	57
Agent	199	Owner	56
Terms and conditions	110	premises	56

Table 3. English glossary.

D.2. Spanish Wordlist

To create the wordlist in Spanish, we followed exactly the same procedure as for the English language. The result of the wordlist with noise can be observed in Figure 6.

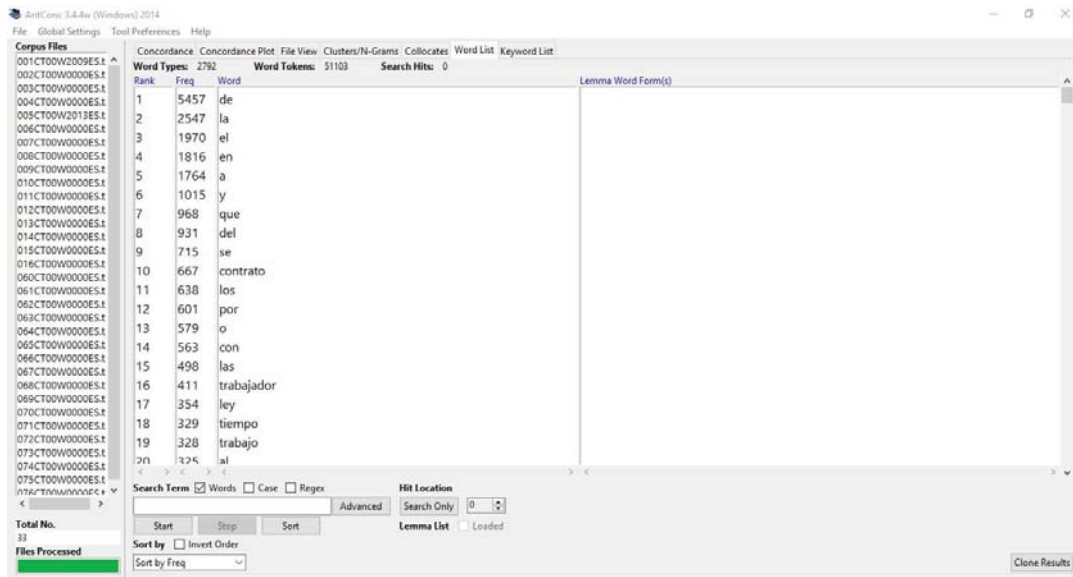


Figure 6. Wordlist Spanish subcorpus.

As with the wordlist in English, the most frequent words in employment contracts in Spanish are grammar/grammatical words. In the position number ten we find the first word with semantic content. We now proceed to study these data.

D.2.1. Analysis of the data

The following table (table 2) shows the twenty words most frequently used from the subcorpus of employment contracts in Spanish.

Nº	WORD	FREQUENCY	Nº	WORD	FREQUENCY
1	De	5457	11	Los	638
2	La	2547	12	Por	601
3	El	1970	13	O	579
4	En	1816	14	Con	563
5	A	1764	15	Las	498
6	Y	1015	16	Trabajador	411
7	Que	968	17	Ley	354
8	Del	931	18	Tiempo	329
9	Se	715	19	Trabajo	328
10	Contrato	667	20	Al	325

Table 4. 20 most common word in Spanish employment contracts.

Of these twenty words, only five have semantic content. The noun that appears most frequently is *contrato*, followed by *trabajador*, *ley*, *tiempo*, and *trabajo*. All these nouns are clearly related to the theme of the corpus. In the top of the list we find the preposition *de*, which has much higher frequency than the next word. Apart from *de*, the prepositions *en*, *a*, *por*, *con* y *al* have been found in the list. The Spanish language uses much more definite articles, both masculine and feminine, than the English language.

This is also reflected in the language of law, where four of the most frequent words are the definite articles *la*, *el*, *los*, and *las*. The particle *del*, number eight in the list, is a contraction formed by the preposition *de* plus the definite article *el*. Following the particle *del* in the list, we find the personal pronoun of the third person *se*. According to the dictionary of the Real Academia Española (hereinafter known as the RAE), it is the atonic form of *él*. This form is very common because it is used for female, male, and neutral forms.

The rest of the words that appear on the list belong to the category of coordinators. The coordinators which most frequently appear are the copulatives *y* and *que*, followed by the disjunctive *o*.

After this analysis, we highlight the absence of verbs among the twenty most frequent words of Spanish subcorpus, while in English subcorpus, we had the verbs *shall*, *will*, and *to be*.

As we know, to carry out an analysis according to our goal (elaborate a glossary with the twenty most common terms of employment contracts), we must eliminate the noise, as we did with English subcorpus. In this case we have made a stopword list formed by the first 2000 most frequent words in Spanish, extracted from the list of the 10,000 most frequent words in Spanish of the RAE corpus. Since it is our mother tongue, we removed from these 2000 words, those that we considered suspicious to be terms. To this list, we have added a common stopword list, extracted too from the webpage <http://members.unine.ch/jacques.savoy/clef/spanishST.txt>. Finally, we have added repeated words in Spanish employment contracts but are not terms, as *DNI*, *NIE*, *dña*, *domicilio*, *euros*, *boe*, or *art*.

The result of the list of most frequent words after using our *stopword list*, can be seen in Figure 7.

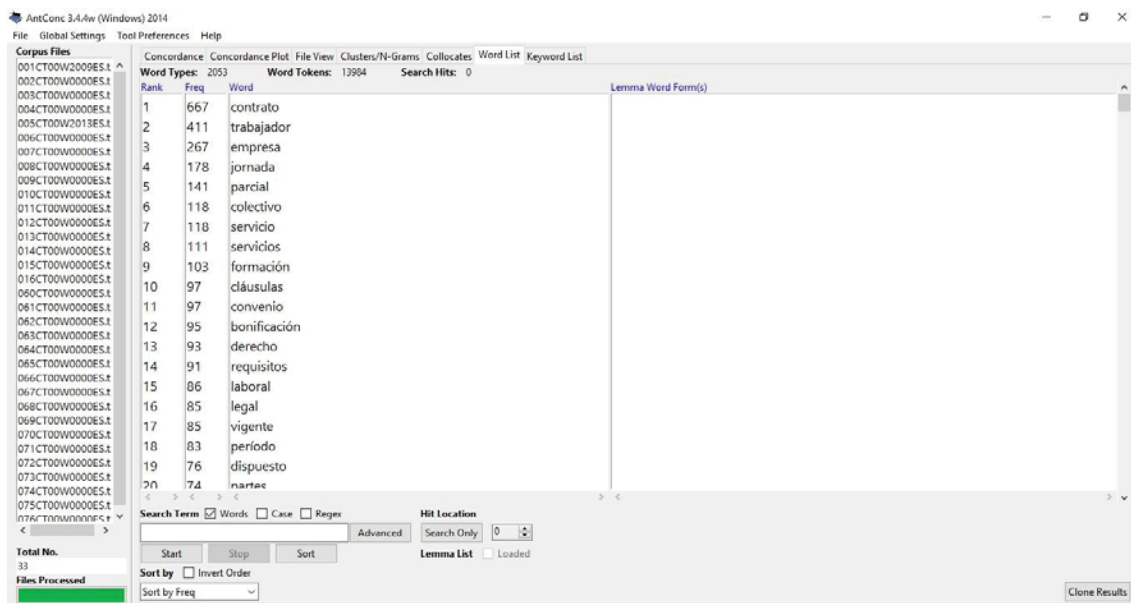


Figure 7. WordList Spanish subcorpus without noise.

Now that we have eliminated a great part of the noise, we proceed to the analysis of the data. To classify a word on the list as a term, we have used two dictionaries. On the one hand, we have used the online RAE dictionary (which labels the semantic field of terms in a distinctive color, implying that these are specific to a specialized language) and on the other hand, the online law encyclopedia <http://www.encyclopedia-juridica.biz14.com/i/index-p.htm>. In addition to the dictionaries, we employ the concordance tool available in the program AntConc. We follow the same parameters as in the English analysis.

We use the same method as in the English analysis. Our Spanish candidates are:

1: trabajador	2: empresa	3: jornada	4: parcial	5: colectivo
6: servicio	7: servicios	8: formación	9: cláusulas	10: convenio
11: bonificación	12: derecho	13: requisitos	14: laboral	15: legal
16: vigente	17: período	18: dispuesto	19: partes	20: salario
21: estatuto	22: específicas	23: corresponda	24: cuotas	25: resolución
26: aprobado	27: reducción	28: representante	29: cliente	30: establecidos
31: semanales	32: certificación	33: proveedor	34: legislativo	35: objeto

Table 5. Spanish candidate terms.

Contrato is the first word of our list. *Contrato* and its plural, *contratos*, appears in a total of 711 times. The RAE defines it as *pacto o convenio, oral o escrito, entre partes que se obligan sobre materia o cosa determinada, y a cuyo cumplimiento pueden ser compelidas*. However, this word has a lot of connotations that are not found in general dictionaries. We cannot focus on the word *contrato* in general, but in the concept of *contrato de trabajo*, which is what our corpus refers to. *Contrato de trabajo* is described in the legal encyclopedia as *negocio jurídico bilateral celebrado entre dos personas por el que una parte, denominada «trabajador», presta de forma voluntaria sus servicios retribuidos por cuenta ajena y dentro del ámbito de organización y dirección de otra persona, física o jurídica, denominada «empleador»* and also, it is a *convención por la cual una persona, el asalariado, pone su actividad profesional a disposición de otra persona, el empleador o patrono, que le paga en cambio un salario y tiene autoridad sobre él*. *Contrato* is our first term.

We continue with the second word, which is *trabajador*. The definitions of the RAE are *que trabaja; muy aplicado al trabajo; and persona que tiene un trabajo retribuido*. In the legal encyclopedia, definitions are explained with more detailed, but not with different meaning (as, for example, *quien trabaja; todo aquel que realiza una labor socialmente útil* or *la parte retribuida en el contrato de trabajo*). This word is also a term because it complies with our methodological criteria.

The next word, *empresa*, does not seem at first sight a term, but if we compare the definition of the legal encyclopedia with that of the RAE, we note that none of them fits with the specialized definition. According to the legal encyclopedia, in *derecho laboral*, a *empresa* is *grupo de trabajadores que cumplen una actividad común bajo la autoridad de un mismo empleador. Varias sociedades jurídicamente distintas pueden, respecto del derecho del trabajo, constituir una sola empresa*. The only definitions of the RAE that could fit in this context are two: *lugar en que una empresa realiza sus actividades* (referring to a building) and *unidad de organización dedicada a actividades industriales, mercantiles o de prestación de servicios con fines lucrativos* (very unespecific). The definition of the encyclopedia is specific to the *derecho laboral español*, and obviously with the meaning it has in the concordance lines. For this reason, *empresa* must be the third word of our list of terms.

The next word is *jornada*, with 178 appearances. It occurs with the collocates, *jornada de trabajo* appearing 69 times, but if we look at all the concordance lines, they all refer to *jornada* as *jornada de trabajo*. According to the legal encyclopedia, *jornada* belongs to *derecho laboral* and is defined as *una condición de ejecución de la prestación relativa al tiempo durante el cual se trabaja, medido en el marco de un contrato indefinido o temporal*, or stated differently by the RAE (second definition), *tiempo de duración del trabajo diario*. The encyclopedia gives us data and fundamental rules about how a *jornada laboral* should be in Spain. This word has several different meanings, however, in legal language, it is specific. Therefore, we should include it in our list of terms.

The next word is *parcial*. It occurs in concordance lines with the phraseological expression *a tiempo parcial*. This expression makes reference to a type of contract. The definition of the RAE which more suits our needs is *no cabal o completo*. Legal encyclopedia says that *el contrato de trabajo se entiende celebrado a tiempo parcial cuando se haya acordado la prestación de servicios durante un número de horas al día, a la semana, al mes o al año inferior al 77% de la jornada a tiempo completo establecida en el Convenio Colectivo de aplicación o, en su defecto, de la jornada ordinaria máxima legal*. We believe that *a tiempo parcial* is a specific expression used in *derecho laboral* and has specific connotations, which makes it a term.

Of the 118 occurrences of *colectivo*, our next word to study, 105 refer to collective agreement. However it is defined in the dictionary of the RAE as *acuerdo vinculante entre los representantes de los trabajadores y los empresarios de un sector o empresa determinados, que regula las condiciones laborales*. This is a very basic definition when compared with that provided by the legal dictionary. They include it into *derecho laboral* and says that it is *la norma especial típica del Derecho del Trabajo (...)*. *Éste constituye la materialización del ejercicio del derecho a la negociación colectiva laboral entre los representantes de los trabajadores y de los empresarios (...)*. *El Tribunal Constitucional (...) ha afirmado que la negociación colectiva es la expresión abreviada de un proceso de diálogo, acercamiento y, eventualmente, acuerdo entre los trabajadores organizados y el empresario o una o varias organizaciones de empresarios. La composición dinámica del conflicto que*

genera el trabajo por cuenta ajena es, de este modo, el fundamento de la negociación colectiva. Clearly, *convenio colectivo* is a term.

In positions six and seven we found *servicio* and its plural, *servicios*, respectively, sharing both words the same lemma. We must count them as one word, with a total of 229 appearances. These are divided into *servicio público* (49 + 8 singular plural concordances), *servicio determinado* (17 sing), *obra o servicio* (28 sing), and the rest of the occurrences makes reference to the provision of different professional services. We can not classify the word *servicio* as a term, nor the terms as *public service*, because their frequency is low. We have joined these words with the same lemma and when we analyzed them, we found different meanings. *Obra y servicio* is part of a longer term, which is *contrato por obra y servicio*. In this case, it is a term.

The word *formación*, does not appear in the specialized dictionary and its concordances make reference, as describes the RAE, as *la acción o efecto de formarse*. That is, *preparar intelectual, moral o profesionalmente a una persona o a un grupo de personas*. The word *formación* is not a term.

The next word is *cláusulas*, with 97 appearances, plural of *cláusula*, with 18 appearances. Both words share the same lemma. We have observed that this word also appears with errata (*clausulas* appears 17 times and *cláusulas* 5 times). Observing the concordance lines, all these words refer to the same thing and we should take them as a unit, with a total of 137 appearances. The RAE dictionary defines this Word as *cada una de las disposiciones de un contrato, tratado, testamento o cualquier otro documento análogo, público o privado*. Legal encyclopedia, which contains the same definition, says that it is *en sentido general*. Also notes that *en sentido particular, disposición que se inserta en un acto jurídico con el objeto de modificar sus efectos normales (...)*. It is a legal term, since it is not used outside the legal context.

The word *convenio* appears 97 times, of which 88 goes along with the word *colectivo*. We already classified *convenio colectivo* as a term.

The next word is *bonificación*. The RAE defines it as *acción y efecto de bonificar*, being the second entry of *bonificar* the following: *conceder a alguien, por algún concepto, un aumento, generalmente proporcional y reducido, en una cantidad que*

ha de cobrar, o un descuento en la que ha de pagar. This word does not appear in the encyclopedia. But in the Spanish contracts, especially those of SEPE (Servicio Público de Empleo Estatal), workers have *bonificaciones* depending on the type of contract (permanent, temporary) and the worker (autonomous, disabled). There are also *bonificaciones* for the good quality of the work done. These are called *bonificaciones tributarias*, and are defined by the legal encyclopedia as *tipo de beneficio fiscal que se caracteriza, por lo general, por aplicarse sobre la cuota tributaria, provocando una disminución en su cuantía.* This is another term.

The word *derecho* is defined by the legal encyclopedia with two clear senses: *como Derecho objetivo se refiere al conjunto de normas que rigen la vida del hombre en sociedad y sus relaciones con los demás miembros de la misma, y como derecho subjetivo hace alusión a las facultades concretas que el ordenamiento reconoce a los individuos dentro del marco del Derecho objetivo.* However, in the definitions provided by the RAE, we have to go to the tenth definition to find something slightly similar (*Facultad de hacer o exigir todo aquello que la ley o la autoridad establece en nuestro favor, o que el dueño de una cosa nos permite en ella*). At first glance, this word looked like it might not be a term but it is.

The next item is *requisito*. The RAE defines it as *circunstancia o condición necesaria para algo.* The legal encyclopedia adapts the RAE definition, and defines it as *circunstancia o condición necesaria para la existencia o ejercicio de un derecho, la validez y eficacia de un acto jurídico o la exigencia de una obligación.* So, according to our parametres, we must consider the word *requisito* as a term.

The word *laboral*, is included in our list of terms because although the general (*perteneciente o relativo al trabajo, en su aspecto económico, jurídico y social*) and legal (*relativo a los aspectos económicos, sociales y jurídicos del trabajo*) definitions are almost identical, the word *laboral* forms part of the term *derecho laboral* and also *pertenece al trabajo* as it is shown in our concordance lines, and then, is a term. As was the case of the word *requisito*.

With the word *legal*, it happens the same as with the previous two words. The definition of the RAE is *prescrito por ley y conforme a ella* and the legal encyclopedia definition is *que resulta de la ley* or *conforme a la ley*. This word is a

specific term of legal documents that contain *disposiciones* or *estatutos legales*, and we can not exclude it from our list.

The word number seventeen, *vigente*, has the characteristics of the last three previous words. In the RAE it is defined as *dicho de una ley, de una ordenanza, de un estilo o de una costumbre: Que está en vigor y observancia*. The legal encyclopedia defines it as *palabra derivada del latín vigere: tener vigor. Se aplica a las leyes, decretos, ordenanzas, edictos y costumbres que están en vigor y observancia*. Although the definition is short, it gives us two important data. The first is that the word derives from Latin, and we know that many legal terms are Latin words. The second is the applications, that shows us the specific field in which this word is used. To possess vigour is a term, and therefore, so are its derivatives. *Vigente* is the next term entering in our list.

The word *período* appears 83 times. It collocates with *prueba*. The word *período* itself is not a term, however, *período de prueba* it is. We should note that this term appears only 33 times. And the last word of our list of terms appears 52 times. Therefore, *período de prueba* does not form part of our list.

The following words, *dispuesto* and *partes* does not even appear in the specialized encyclopedia. Although the English equivalent of *partes*, *party*, are terms, in Spanish it is not.

The word *salario*, is the next candidate, and it is also the singular of *salarios*, which appears also two times in the Spanish subcorpus. The RAE dictionary defines *salario* as *paga o remuneración regular* and *cantidad de dinero con que se retribuye a los trabajadores por cuenta ajena*. This is the general definition known for Spanish speakers, but the legal encyclopedia provides much more information. *Salario* is a specific term of *derecho laboral*, and is defined as *retribución económica total, en dinero o en especie, que percibe un trabajador por su prestación de servicios por cuenta ajena, ya retribuyan el trabajo efectivo o los períodos de descanso computables como de trabajo*. It is also *a prestación pagada por el empleador al asalariado en retribución de su trabajo* and it could be *salario de base o mínimo* (24 and 12 appearances respectively), *salario diferido*, *salario indirecto*, *salario al rendimiento*, *salario al tiempo*, *salario bruto* o *salario básico*.

To these occurrences we must add 44 more that belong to the word *salariales*, because they share the same lemma. The word *salario* should be considered a term and be part of our glossary.

Estatuto and its concordance lines are the next items to be analyzed: *trabajadores*, being *Estatuto de los Trabajadores* a legal term based on a law. *Estatuto* is a *documento de constitución de una sociedad o asociación redactado por escrito y que contiene cierto número de cláusulas obligatorias que fijan los objetivos así como las normas de funcionamiento de la sociedad o asociación*. According to <http://www.estatuto-trabajadores.com>, *el Estatuto de los Trabajadores es la norma principal que rige los derechos para las personas que voluntariamente prestan sus servicios a un empleador a cambio de una retribución, en España*. This law is one of the basic principles of *derecho laboral* in Spain, and should be known by all speakers of a language, and therefore, included in our glossary.

The next word is *específicas*, and in all its concordance lines it appears with the word *cláusulas*, being the concept of *cláusulas específicas*. Unfortunately, this concept does not appear in legal dictionaries. This concept is adjusted to contract specifications and it is not a term

The next word is *corresponda*, which do not appear in the legal encyclopedia and it is commonly used as the definition provided by the RAE (*pagar con igualdad, relativa o proporcionalmente, afectos, beneficios o agasajos*). This word is not a term because it has a common use with its common meaning, is not related to law, and does not appear in specialized dictionaries.

The next word is *cuotas*, plural of *cuota*, both belonging to the same lemma, making a total of 91 appearances. The RAE dictionary defines *cuota* as *parte o porción fija y proporcional*. This definition is too limited to deal with such an important legal document. So we must provide the information contained in the legal encyclopedia, which defines *cuota* as *parte determinada y fija que corresponde dar o percibir a cada uno de los interesados en un negocio, suscripción, empréstito, etc. / Lo señalado de antemano; como una obligación, contribución, derecho, etc., en forma periódica, temporal o por una sola vez*. We need to consider *cuota* as a term because we can not take its meaning for granted.

The next word is *resolución*, and the definition that best fits the concept we are looking for in the dictionary of the RAE is *decreto, providencia, auto o fallo de autoridad gubernamental o judicial*. As usual, the definitions of the RAE for this project are poor and inaccurate in terms of legal language. The legal encyclopedia defines *resolución* as *acto jurídico comunitario en el que se expresan las intenciones comunes o la voluntad de las instituciones europeas para el desarrollo de una determinada materia. Tienen relevancia política pues perfilan el camino a seguir en el futuro*. This word is a legal term, and according to our methodology must be included in our list of terms.

Aprobado and *reducción* are common words that do not change their meanings in employment contracts, they do not appear in the specialized dictionary, and not form part of our glossary of terms.

The next word is *representante*. In legal terms, according to the legal encyclopedia, a *representante* (*representante legal*) is *la persona interpuesta entre las dos partes de un contrato y en el cual el representante ha actuado en nombre e interés de una de ellas, llamada el representado, quedando éste obligado con quien contrató con el representante. Este no se limita a transmitir una manifestación de voluntad contractual del representado, sino que expresa su propia voluntad de contratar ligando así al representado, aunque se supone que lo haga siguiendo las instrucciones recibidas del representado*. The RAE dictionary does not provides us a definition for this concept. But it is considered a term.

We discard the next words, *cliente*, *establecidos*, and *semanales*, because we did not find them in the legal encyclopedia.

The next word, *certificación*, in the general use, means *acción y efecto de certificar* or *documento en que se asegura la verdad de un hecho*, according to the RAE. Our legal tool describes it as *copia total o parcial, autenticada por el Secretario Judicial de un órgano judicial (...), documento justificativo de la verdad de algún escrito, acto o hecho* or *acto por medio del cual una persona da fe de algo que le consta*. *Certificación* enters into our list of terms.

The word *proveedor* does not appear in the legal encyclopedia and we have realized that all occurrences are in the same contract, so it might be a feature specific of just one contract. For these reasons, we do not consider it as a term.

The word *legislativo* is defined by the RAE, *dicho especialmente de un organismo: Que legisla / Perteneciente o relativo a la legislación* or *Autorizado por una ley*. The legal encyclopedia defines this word in a very similar ways (*dícese del código, cuerpo o texto de leyes / Se aplica al derecho o potestad de hacerlas o darlas / Lo autorizado por una ley*). The concordance lines of this word make reference to the Decretos Reales, and the decretos of employment contracts in Spanish are apecific and related to these type of documents. So, we must consider this word as a term.

We analyzed thirty four candidate terms to obtain twenty terms. The list of terms and its appearances are the following:

TERM	APPEARANCES	TERM	APPEARANCES
Contrato	667	Requisito	92
Trabajador	411	Cuota (lemma)	91
Empresa	267	Laboral	86
Jornada	178	Legal	85
A tiempo parcial	141	Vigente	85
Cláusula (lemma)	137	Estatuto de los trabajadores	66
Bonificación (lemma)	130	Resolución	59
Salario	114	Representante	57
Convenio colectivo	97	Certificación	54
Derecho	93	Legislativo	52

Table 6. Spanish glossary.

D.3. Real examples and equivalent translations

Now we are going to propose the equivalences of the different words that form both glossaries in the opposite language. That means, we create a list of Spanish equivalences for the English glossary and the English equivalences for the Spanish glossary. We have reached these equivalencies after the knowledge we have acquired doing this project.

D.3.1. From English to Spanish:

TERM AND EXAMPLE	EQUIVALENT TRANSLATION
<p>Employee</p> <p>The Employee's work schedule will vary according to the needs of the Employer.</p>	<p>Empleado. Trabajador.</p> <p>El horario de trabajo del trabajador variará en función de las necesidades de la empresa.</p>
<p>Employer</p> <p>Both parties agree that the foregoing list of duties and responsibilities are not exhaustive and may be changed, modified, or increased at the discretion of the Employer.</p>	<p>Empresa (contratadora). Compañía.</p> <p>Ambas partes acuerdan que la lista anterior de deberes y responsabilidades no son exhaustivas y pueden ser cambiadas, modificadas o aumentadas a voluntad de la Empresa.</p>
<p>Employment</p> <p>This employment agreement hereby entered into between [name of employer], the Employer, and [name of employee], the Employee.</p>	<p>Empleo. Trabajo.</p> <p>Por la presente, este acuerdo de trabajo celebrado entre [nombre de la empresa], la empresa y [nombre del trabajador], el trabajador.</p>
<p>Notice</p> <p>This employment is an at-will employment that may be terminated without cause and without advance notice.</p>	<p>Notificación o aviso.</p> <p>Este empleo es un empleo a voluntad que puede ser rescindido sin causa y sin previo aviso.</p>
<p>Termination</p> <p>That upon the termination of my employment from the Company.</p>	<p>Resolución (de un contrato).</p> <p>Que, a la resolución de mi empleo de la empresa.</p>
<p>Agree/contract</p> <p>Employer and Employee agree as follows.</p>	<p>Contrato, acuerdo.</p> <p>El empleador y el empleado acuerdan lo siguiente.</p>
<p>Party</p>	<p>Cada una de las partes que acuerdan el contrato.</p>

<p>The parties are executing this Agreement voluntarily and without any duress or undue influence.</p>	<p>Parte contratante oparte contratada.</p> <p>Ambas partes están ejecutando este acuerdo de forma voluntaria sin coacción ni condicionamiento indebido.</p>
<p>Benefit</p> <p>During your employment, you will be provided the additional benefits stated in your Schedule.</p>	<p>Beneficio o privilegio.</p> <p>Durante su empleo, se le proporcionarán los beneficios adicionales indicados en su horario.</p>
<p>Agent</p> <p>For the Agent's services, the Company shall pay the Agent the following commission percentage.</p>	<p>Representante legal.</p> <p>Por los servicios del representante legal, la empresa pagará al representante legal el siguiente porcentaje de la comisión.</p>
<p>Terms and conditions</p> <p>Employee is willing to be employed by employer, and employer is willing to employ employee, on the terms and conditions hereinafter set forth.</p>	<p>Términos y condiciones.</p> <p>El trabajador está dispuesto a ser contratado por la empresa, y la empresa está dispuesta a contratar al trabajador, de acuerdo con los términos y condiciones establecidos de aquí en adelante.</p>
<p>Contractor</p> <p>The Company shall not be responsible for withholding taxes with respect to the Contractor's compensation hereunder.</p>	<p>Contratista, parte contratante.</p> <p>La empresa no será responsable de los impuestos retenidos con respecto al salario base de la parte contratante a continuación.</p>
<p>Compensation</p> <p>The Employee's full compensation will be reinstated upon return to work.</p>	<p>Salario base. En algunos casos, indemnización.</p> <p>La total indemnización del empleado le será reintegrada a su regreso al trabajo.</p>
<p>Contract</p> <p>This contract shall become effective on [date] and remain in effect until it is terminated by either party.</p>	<p>Contrato. Acuerdo.</p> <p>El presente contrato será efectivo desde [fecha] y permanecerá en vigor hasta su resolución por cualquiera de las partes.</p>
<p>Property</p> <p>The records will be available to and remain the sole property of the Company at all times.</p>	<p>Propiedad.</p> <p>Los registros estarán disponibles y continuarán siendo propiedad exclusiva de la empresa en todo momento.</p>
<p>Arbitration</p>	<p>Mediación. No traducir como arbitraje, pues este término es usado en español para política internacional o asuntos comerciales. Mediación se usa cuando se habla de un asunto laboral.</p> <p>Cualquier disputa que surja a partir de este</p>

Any dispute under this contract shall be required to be resolved by binding arbitration of the parties hereto.	contrato deberá ser resuelta por una mediación vinculante de las partes del presente documento.
<p>Breach</p> <p>You shall promptly inform the Employer of any circumstances which could cause you to be in breach of this clause.</p>	<p>Infracción o incumplimiento de una parte del contrato o de su totalidad.</p> <p>Usted deberá comunicar a la empresa acerca de cualquier circunstancia que podría ocasionar el incumplimiento de esta cláusula.</p>
<p>Right</p> <p>No modification of or amendment to this Agreement, nor any waiver of any rights under this Agreement, will be effective unless in writing signed by the party to be charged.</p>	<p>Derecho a algo. Correcto, adecuado o apropiado.</p> <p>Ninguna modificación o enmienda de este contrato, ni ninguna renuncia de cualquier derecho en virtud del presente contrato, será efectiva a menos que se de por escrito y firmado por la parte afectada.</p>
<p>Perform</p> <p>To the extent required by law all work shall be performed by individuals duly licensed and authorized by law to perform said work.</p>	<p>Cumplir con las obligaciones del contrato. Desempeñar, realizar, llevar a cabo.</p> <p>Según lo requiera la ley, todo el trabajo deberá desempeñarse por personas debidamente autorizadas por la ley para el desempeño de dicho trabajo.</p>
<p>Owner</p> <p>Owner agrees to compensate the broker for the services rendered in accordance with this contract as follows.</p>	<p>Propietario.</p> <p>El propietario acuerda remunerar al representante legal por los servicios prestados de acuerdo con el presente contrato de la siguiente manera.</p>
<p>Premises</p> <p>Employer hereby employs employee as ___ at the above-mentioned premises, and employee hereby accepts and agrees to such employment.</p>	<p>Lo dicho anteriormente. Los supuestos o premisas anteriores. Los considerandos. Instalaciones.</p> <p>Por la presente, la empresa contrata al trabajador como ___ en las instalaciones mencionadas anteriormente, y por la presente, el trabajador acepta y se compromete a dicho empleo.</p>

D.3.2. From Spanish to English:

TERM AND EXAMPLE	EQUIVALENT TRANSLATIONS
<p>Contrato</p> <p>Con independencia de lo anterior, las partes se obligan a registrar el presente contrato en la asociación empresarial y profesional a la que en su caso pertenezcan.</p>	<p>Contract. Agreement.</p> <p>Regardless of the foregoing, the parties undertake to register this contract in the business and professional association which they belong.</p>
<p>Trabajador</p> <p>Para la aplicación de las reducciones se deberá mantener en el empleo al trabajador contratado al menos 18 meses desde la contratación.</p>	<p>Employer, worker.</p> <p>For the application of the reductions, employer must maintain the employee on the employment at least 18 months since the hiring.</p>
<p>Empresa</p> <p>La empresa se compromete a mantener la estabilidad en el empleo del/de la trabajador/a, que contrata en los términos establecidos en el art. 10 del R.D. 1451/1983, de 11 de mayo.</p>	<p>Company.</p> <p>The company is comitted to maintain the employment stability of the employer, which contracts under the terms established in the art 10 of R.D. 1451/1982, on the 11th of May.</p>
<p>Jornada</p> <p>En caso de contratos a jornada completa, indiquese si el contrato celebrado es de bajo rendimiento.</p>	<p>Working hours. Working day.</p> <p>In full working day contracts, indicate whether the contract is low performance.</p>
<p>A tiempo parcial</p> <p>Si el contrato se celebra a tiempo parcial esta subvención se reducirá proporcionalmente.</p>	<p>Part-time (job).</p> <p>If the contract is part-time type, this grant will be reduced proportionally.</p>
<p>Cláusulas</p> <p>Cláusulas específicas de obra o servicio determinado.</p>	<p>Clause.</p> <p>Specific clauses for work or specific project.</p>
<p>Bonificación</p> <p>La bonificación se aplicará a todas aquellas contrataciones que se efectuen hasta el 30 de junio de 2016.</p>	<p>Bonus. Reward.</p> <p>The bonus will be applied to all those hirings accomplished until 30th of June, 2016.</p>
<p>Salario</p> <p>Salario base y complementos salariales.</p>	<p>Salary. Wage.</p> <p>Base salary and salary supplements.</p>

<p>Convenio colectivo</p> <p>La jornada a tiempo completo prevista en el Convenio Colectivo de aplicación.</p>	<p>Collective bargaining agreement.</p> <p>Full working day contracts planned in collective bargaining agreement.</p>
<p>Derecho</p> <p>El trabajador autónomo económicamente dependiente tendrá derecho a una interrupción anual de la actividad de _____ días.</p>	<p>Right.</p> <p>The economically dependent self-employed person shall have the right to an annual activity disruption of ___ days.</p>
<p>Requisito</p> <p>En el caso de reunir los requisitos establecidos en el Artículo 8 de la Ley 25/2015, de 28 de julio, la empresa aplicará la bonificación correspondiente.</p>	<p>Requirement, requisite.</p> <p>In the event of meet the requirements of the Article 8 of the Law 25/2015, of July 28, the company will apply the pertinent bonus.</p>
<p>Cuota</p> <p>Sustituir a trabajadoras/es por maternidad, sin bonificación de cuotas.</p>	<p>Fee. Quota.</p> <p>Substitute employers due to maternity without fee bonus.</p>
<p>Laboral</p> <p>No tener experiencia laboral o que esta sea inferior a tres meses.</p>	<p>Labour in British English. Labor in American English.</p> <p>Not have work experience or that latter is a minor than three months.</p>
<p>Legal</p> <p>Ambas partes se reconocen mutuamente la capacidad legal necesaria para contratar y a tal efecto exponen.</p>	<p>Legal.</p> <p>The parties mutually recognize their sufficient legal capacity to contract, and to this effect, attest.</p>
<p>Vigente</p> <p>Señalar el nivel profesional del tutor, según el sistema de clasificación profesional vigente en la empresa.</p>	<p>Current. Currently in force. Prevailing.</p> <p>Point the professional level of the tutor, according to current job classification system in the company.</p>
<p>Estatuto de los trabajadores</p> <p>Habrà de respetarse, en todo caso, lo dispuesto en el art. 14.1 del Texto Refundido de la Ley del Estatuto de los Trabajadores (...).</p>	<p>Statute of workers' rights. Employee rights. Workers regulations.</p> <p>Shall be respected in any case, the provisions of the article 14.1 of the consolidated text, belonging to the statute of workers' rights (...).</p>

<p>Resolución</p> <p>Que su capacidad de trabajo se encuentra disminuida en un ___ por ciento, según resolución del correspondiente Equipo Multiprofesional.</p>	<p>Decision, resolution.</p> <p>That your work capacity is decreased by ___ per cent, in agreement with the resolution of the corresponding multi-professional team.</p>
<p>Representante</p> <p>El representante de la menor, si procede.</p>	<p>Agent.</p> <p>The agent of the minor, if applicable.</p>
<p>Certificación</p> <p>El trabajador deberá entregar al empresario fotocopia compulsada del título, certificación de su solicitud o certificación acreditativa de la terminación de los estudios.</p>	<p>Certification.</p> <p>Employee shall provide the employer a certified copy of the qualification, certification of the request or document certifying the studies completion.</p>
<p>Legislativo</p> <p>Respetando lo establecido en el art. 14.1 del Texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por R.D. Legislativo 2/2015, de 23 de octubre (BOE de 24 de octubre).</p>	<p>Legislative.</p> <p>According with the principles established in article 14.1 of the article 14.1 of the consolidated text, belonging to the statute of workers' rights, approved by Royal Legislative Decree 2/2015 of October 23 (BOE of October 24).</p>

E. CONCLUSION

In this last chapter we will collect the main conclusions accomplished all along this project. After the extraction, analysis, and definition of the terms, the conclusions we reached are the following.

Comparing the terms obtained in both languages, we have studied 34 Spanish terms to achieve the list of the most frequency words in Spanish, while in the English corpus, we have had to analyze 51 words to reach the same number of terms. There are only two words in the English list which their equivalents appears also in the Spanish list. These words are *agree/contract-contrato* and *right-derecho*. Some terms are formed by more than one word, as the case of *good faith* and *good reason* in English, and *convenio colectivo* and *estatuto de los trabajadores* in Spanish. From that lists, our attention is drawn to the fact that the English word *party* forms part of the specialized legal language, and its equivalent in Spanish, *parte*, does not belong to the same category in its language. This explains that employment contracts in English uses Latin words that do not have other meanings out the specialized language.

This is a project based on different branches within linguistics and social sciences, as they are corpus linguistics, translation and legal language. Thus, the results obtained have consequences for all these fields. This project looked for increasing the knowledge we have about the different frequency words in the comparison of Spanish and English employment contract documents. This could provide the basis for a major study.

We have confirmed that corpora are good tools for the study of a specialized language. The different tools as wordlist, concordance, and collocations, are really usefull not only to obtain the frequency of words, but to put that words in context to define properly their meanings, avoiding confusions.

BIBLIOGRAPHY

- Baker, P., Andrew Hardie, and Tony McEnery. *A Glossary Of Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2006. Print.
- Belica, C. Analysis of temporal change in corpora. *International Journal of Corpus Linguistics*, 1(1), 61–74. doi: [10.1075/ijcl.1.1.05bel](https://doi.org/10.1075/ijcl.1.1.05bel) .1996.
- Bhatia, V. K., Christopher Candlin, and Maurizio Gotti. *Legal Discourse In Multilingual And Multicultural Contexts*. Bern: P. Lang, 2003. Print.
- Biber, D., Susan Conrad, and Randi Reppen. *Corpus Linguistics*. Cambridge: Cambridge University Press, 1998. Print.
- Biel, Ł “Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential.” *Reconceptualizing LSP*. Online proceedings of the XVII European LSP Symposium 2009 Aarhus. Eds. Carmen Heine and Jan Engberg. Aarhus: n.p., 2010. Web.
- Bowker, L., and Jennifer Pearson. *Working With Specialized Language*. London: Routledge, 2002. Print.
- Dash, N. S. *Corpus Linguistics*. Delhi: Pearson Longman, 2008. Print.
- Gelbukh, A., Grigori Sidorov, and Liliana Chanona-Hernández. Compilation of a Spanish representative corpus. *CICLing-2002, Intelligent Text Processing and Computational Linguistics*. Lecture Notes in Computer Science N 2276, Springer-Verlag, 2002, pp. 285–288. 2002.
- Hirschman, L. et al. Automating coreference: The role of annotated training data. In *Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*. 1997.
- [Http://dictionary.law.com/](http://dictionary.law.com/) February-May, 2016.
- [Http://dle.rae.es/?id=DgIqVCc](http://dle.rae.es/?id=DgIqVCc) March-May, 2016.
- [Http://www.encyclopedia-juridica.biz14.com/i/index-p.htm](http://www.encyclopedia-juridica.biz14.com/i/index-p.htm) February-May, 2016.

[Http://www.oxforddictionaries.com/](http://www.oxforddictionaries.com/) February-May, 2016.

Leech, G. 1991. "The state of the art in corpus linguistics". *English Corpus Linguistics*, K. Aijmer and B. Altenberg (eds.). London: Longman. 8-29

[Legal-dictionary.thefreedictionary.com/](http://legal-dictionary.thefreedictionary.com/) February-May, 2016.

Lu, X. "Computational Methods for Corpus Annotation and Analysis". Springer, 2014. Web. 1 March, 2016.

McEnery T., Richard Xiao And Yukio Tono. *Corpus-Based Language Studies*. An Advanced Resource Book. London and New York: Routledge. 2006.

McEnery, T. and Andrew Wilson. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2001. Print.

McEnery, T. *Swearing In English*. London: Routledge, 2003. Print.

McEnery, T., and Andrew Hardie. *Corpus Linguistics*. Cambridge: Cambridge University Press, 2012. Print.

Mellinkoff, D. *The Language Of The Law*. Boston: Little, Brown, 1963. Print.

Meyer, Charles F. *English Corpus Linguistics*. Cambridge, UK: Cambridge University Press, 2002. Print.

Renouf, A., and Andrew Kehoe. *Corpus Linguistics*. Kenilworth: Rodopi, 2009. Print.

Seghiri, M. "Too Big Or Not Too Big: Establishing The Minimum Size For A Legal Ad Hoc Corpus". *Hermes – Journal of Language and Communication in Business* n° 53 2014. Print.

Sharoff, S. "Creating General-Purpose Corpora Using Automated Search Engine Queries". *Wacky! Working papers on the Web as Corpus*. N.p., 2006. Web. 9 Feb. 2016.

Sinclair, J. "Corpus and Text - Basic Principles" in *Developing Linguistic Corpora: a Guide to Good Practice*, ed. M. Wynne. Oxford: Oxbow Books: 1-16. 2005.

- Sinclair, J. (1996) *Preliminary Recommendations on Corpus Typology*. EAGLES Document EAG-TCWG-CTYP/P, May 1996.
- Sinclair, J. *Corpus And Text - Basic Principles*. Oxford: M. Wynne, 2005. Print.
- Sinclair, J., and Antoinette Renouf. "A Lexical Syllabus for Language Learning." *Vocabulary and Language Teaching*. Eds. Carter, R. and McCarthy, M. London: Longman, 1988. 140-160. Print.
- Teubert, W. Corpus linguistics-A partisan view. *International Journal of Corpus Linguistics* 4(1):1-16. 2000.
- Tognini-Bonelli, E. *Corpus Linguistics At Work*. Amsterdam: J. Benjamins, 2001. Print.
- Torruella, J., and Joaquim Llisterri. Diseño de corpus textuales y orales. En J. M. Blecua, Gloria Clavería, Carlos Sánchez y J. Torruella (eds.), *Filología e Informática: Nuevas tecnologías en los estudios filológicos*. Barcelona: Universidad Autónoma de Barcelona. 1999.
- Zhu, J., and Quian Wu. "A Brief Analysis on the Stylistic Features of English for Business Contracts." *Nintendo r4 3ds*. 21 September, 2015. Web. 19 December, 2015.