Automatic term recognition and legal language: a shorter path to the lexical profiling of legal texts?

María José Marín

Universidad de Murcia

Natural Language Processing (NLP) tools offer language scholars a wide array of possibilities to examine, amongst other, the lexicon in any text collection. This research was designed as an attempt to try to measure the degree of precision of three of these methods (Chung 2003; Drouin 2003; Scott 2008a) through their implementation on two corpora of Spanish and British judicial decisions which revolve around the topic of immigration. In addition, the last section of this chapter explores the lexical inventories extracted by each method (the top 500 candidate terms (CTs) in each case) by grouping them into *ad hoc* thematic categories, the most numerous being, as was to be expected, *legal terms*, followed by *territory*, *evaluative items*, *crime* and *family*.

Keywords: ATR methods; Legal corpora; Terminology; Immigration; Judicial decisions

1. Introduction

Specialized terminology plays a pivotal role in languages which are used for specific purposes (LSP) and its definition has been envisaged from numerous perspectives. It has often been conceptualized as a vehicle of communication amongst specialists which conveys "domain-specific key concepts in a subject field that crystallize our expert knowledge in that subject" (Kit, and Liu 2008, 204), in other words, terms are regarded as "textual realisation[s] of a specialized concept" (Spasic et al. 2005, 240).

As Cabré (1999) also acknowledges, the concept of term is a multifaceted construct, since terms display specific semantic and pragmatic traits which are shared by the general and specialized fields of language. Nevertheless, they can be distinguished owing to their capacity to "designate concepts pertaining to special disciplines and activities" (1999, 81). In that vein, Chung (2003) introduces a complementary perspective on the study of terms which articulates itself around the dichotomy between the qualitative and quantitative character of these lexical units, emphasizing the saliency of the statistical data which terms are associated with.

Terms might also be employed for the identification of thematic areas in specialized corpora by, for instance, using them as a point of departure to obtain the collocate networks that revolve around them or simply by classifying them into thematic groups and basing their examination on their statistical relevance, hence their significance in lexical analysis. However, handling and manually processing large corpora in search of specialized terms might become an unattainable task which would necessarily require the systematization if not automatization of the process. ATR methods such as Chung's (2003), Drouin's (2003) or Scott's (2008a) allow the user to retrieve a list of CTs from a specialized corpus when contrasted with a collection of general language texts relatively easily (Chung's method must be applied manually). Yet, the validation of CT inventories becomes essential in order for the methods' precision to be tested, being often performed by comparison with a gold standard, that is, a specialized term glossary which facilitates the assessment of the lists of terms extracted. This process should ideally be supported by human validation as long as the lists of terms are not excessively numerous.

Given the scarcity of research devoted to the study and assessment of ATR methods in the legal field, this chapter seeks to explore in detail three of these methods (Chung 2003; Drouin 2003; Scott 2008a) after their implementation on two corpora of Spanish and British judicial decisions on immigration with the aim of establishing their degree of precision in term retrieval, as presented in Sections 3.1, 3.3 and 4.1.

Along these lines, ATR techniques may also signal major thematic areas that corpora revolve around other than legal terminology, as already stated. Scrutinizing the term inventories which are produced by ATR methods to identify the most representative topics in a corpus might also be another advantage of using ATR techniques for the lexical profiling of legal texts. Section 4.2 was thus designed to that end by introducing an analysis of the thematic areas which the terms retrieved from both corpora could be classified into. On the one hand, the top 500 terms extracted by each ATR method were divided into four *ad hoc* categories, namely, *legal terms, territory, evaluative items, family* and *crime*. Then, the percentage of terms identified by each method which fell into each category was calculated and compared across methods. On the other hand, an automatic text classification software, *UMUTextStats* (García-Díaz et al. 2018; García-Díaz, Cánovas-García, and Valencia-García 2020), was implemented on both corpora and the proportion of items belonging in each of the morphosemantic categories included in the software was determined as a way of comparison with the procedure described above, which resorts to ATR as the basis for thematic classification.

2. ATR and legal language

The literature on ATR methods and software tools has been profusely reviewed (Cabré et al. 2001; Chung 2003a, 2003b; Drouin 2003; Lemay et al. 2005; Maynard and Ananiadou 2000; Kit, and Liu 2008; Pazienza et al. 2005; Vivaldi et al. 2012, to name but a few) often classifying them according to the type of information used to extract CTs automatically. Some of the reviewed methods resort to statistical information, amongst them: Church and Hanks (1990), Ahmad et al. (1994), Drouin (2003), Chung (2003), Fahmi et al. (2007), Scott (2008) or Kit and Liu (2008). Other authors like Ananiadou (1988), David and Plante (1990), Bourigault (1992) or Dagan and Church (1994) focus primarily on linguistic aspects. The so-called hybrid methods rely on both. The work of Justeson and Katz (1995), Daille (1996), Frantzi and Ananiadou (1996; 2000), Jaquemin (2001), Drouin (2003), Barrón Cedeño et al. (2009) or Loginova et al. (2012) illustrate this trend. As stated by Vivaldi et al. (2012), not many of these methods resort to semantic knowledge, namely, TRUCKS (Maynard, and Ananiadou 2000), YATE (Vivaldi, 2001), MetaMap (Aronson, and Lang, 2010) or Meijer et al. (2014). In the recent years, a greater tendency has been shown towards the implementation of machine learning techniques on term/phrase extraction, the work by Arora et al. (2016) or Shang et al. (2018) illustrate this trend.

However, the literature on the evaluation of these methods is not so abundant. There are initiatives for the evaluation of ATR methods like the one organized by the Quaero program (Mondary et al. 2012), which aims at studying the influence of corpus size and type on the results obtained by these methods as well as the way different versions of the same ATR methods have evolved. Some authors also show their concern about the lack of a standard for ATR evaluation which is often carried out manually or employing a list of terms, a gold standard, which is not systematically described (Bernier-

Colborne 2012, 1). For instance, some researchers like Sauron, Vivaldi and Rodríguez, or Nazarenko and Zargayouna (in Bernier-Colborne, 2012), who have worked on this area although there is still much to be done in this respect. Along these lines, Heylen and De Hertog (2015) reflect upon automatic term extraction from specialized corpora by focusing on the subtasks implied in such processes such as corpus compilation or the concepts of unithood or termhood, amongst other. Finally, the research work by Astrakhantsev (2018) could be regarded as a hybrid between the proposal of a novel state-of-the-art ATR method, *ATR4S* (based on the assessment of 13 different ATR methods), which evaluates the degree of precision achieved by each of these methods and their processing time.

The number of studies concerned with the implementation and validation of ATR methods within the legal field is scarcer as opposed to other specialized areas such as biology, anatomy or engineering, to name but a few. The peculiar statistical behaviour of legal terminology might justify this fact. The degree of integration of certain legal terms within the general lexicon can easily be observed. As proved in Marín (2016), 45.41% of the terms identified in a legal corpus also displayed high frequency values in the list of the 3,000 most frequent words of the British National Corpus, a general language text collection. Such statistical behaviour is labelled as semitechnical by authors like Coxhead (2000). Consequently, the automatic extraction of legal terms, which is commonly achieved through corpus comparison, might become unwieldly as opposed to other language areas, where terms are almost exclusively used in specialized texts.

3. Methodology

As stated above, the work by Marín (2014) demonstrates the effectiveness of four ATR methods focused on single-word legal term retrieval as implemented on a corpus of judicial decisions. Some of these methods, which will also be assessed in the present research, performed quite efficiently, finding that Drouin's (2003) appeared to be the most effective one in automatically identifying legal terms. It reached 73,2% average precision for the top 2,000 CTs.

Regarding the selection of the methods described herein, it was made on the basis of their efficiency as evidenced in Marín (2014) and as demonstrated by the authors themselves. In addition, it was also conceived as a procedure to establish a comparison between automatized v. non-automatized methods. As justified below, the first two ATR methods, Scott's (2008) and Drouin's (2003) are fully automatic whereas Chung's (2003) requires the manual application of the algorithm proposed. The results would serve not only as a way to suggest an efficient method in legal term extraction but also to illustrate the advantages and disadvantages of having to implement one of these methods in a manual way.

In this respect, the term *precision* could be defined as the degree of accuracy in automatic term retrieval, which can be measured both automatically and manually. For the automatic calculation of precision there needs to exist an electronic glossary of terms used as the gold standard which CT lists are compared against. Finding a reliable electronic legal glossary to be used as reference in more than one language is not always an attainable task, and manual validation becomes the method implemented to confirm CTs as true terms (TTs). This is the case of the research at issue, where two specialists, one of them a corpus linguist and terminologist, the other one a legal language instructor specialized in corpus linguistics, acted as referees by manually supervising the CT lists and confirming whether the terms extracted could be ratified as TTs.

One of the limitations of manual supervision is the smaller size of the CT inventories, which were limited to 500 in each language for the present study due to practical reasons.

In order to compensate for the degree of subjectivity implied in the manual validation of the TTs found amongst the 3,000 CTs obtained in English and Spanish (2 lists of 500 items per language and method), an inter-rater reliability test was employed whereby the referees had to classify the terms found in the inventories into four main categories, namely, highly specialized terms (occurring in the legal context almost exclusively), semitechnical terms (those shared by the general and specialized fields), undefined (it was not clear whether an item was a term or a general word) and non-terms. Only those items falling within the first two categories were considered as TTs so as to determine the average precision attained by each ATR method. If any of the items included the last two categories (undefined or non-terms) by any of the referees was identified as a member of the first two (technical or semitechnical terms) by the other one, it was also discarded. Nevertheless, before doing so, they were given the chance to discuss and come to an agreement, whenever possible, on some of the items which there was no initial consensus about.

3.1 Method description

3.1.1 Keywords

Scott's (2008a) application, *Keywords* (included in the software package *WordSmith*), could not be deemed an ATR method in itself, at least it is not presented as such by the authors. Nonetheless, given the results examined below and as evidenced in Marín (2014), its degree of efficiency in legal term mining is noticeably higher than other ATR methods specifically designed to that end. However, it is not included exclusively in Scott's software package, other authors like Anthony (2020) or Kilgarriff et al. (2014) also offer the possibility of implementing it automatically employing different parameters to measure the statistical significance of a term in a specialized corpus.

In this case, Scott's version (2008a) was singled out owing to its user-friendly character. Being part of a software package, this tool facilitates greatly the automatic comparison and processing of two large corpora through the implementation of different statistical measures that the user configures.

The automatization of the extraction process saves a considerable amount of time and effort, not requiring advanced mathematical knowledge for the manual implementation of the algorithms underlying these methods, for instance, Dunning's (1993) *log-likelihood*. Scott's software retrieves terms automatically through the identification of those lexical items in a specialized corpus which are "unusually frequent (or unusually infrequent) in comparison with what one would expect on the basis of the larger word-lists" (Scott 2008b, 184), which might signal, in Biber's words, a word's "importance as a content descriptor" (in Gabrielatos 2011, 5).

For the present analysis, Dunning's log-likelihood (1993) was implemented for automatic keyword extraction. As already stated, the identification of the keywords in both legal corpora (which will be described in greater detail in Section 3.4) was achieved by comparing them against two sets of general language texts. The reference corpus in English was a section of LACELL, a 14.8-million-word collection of general English texts which excluded those not coming from British sources. The entire corpus was compiled by the LACELL (*Lingüística Aplicada Computacional, Enseñanza de Lenguas y Lexicografía*[1]*)* research group at the University of Murcia. It is a 21 million-word (118,105 KB) balanced and synchronic corpus which includes both written texts from diverse sources such as newspapers, books (academic, fiction, etc.), magazines, brochures, letters and so forth, and also oral language samples from conversation at different social levels and registers, debates and group discussions, TV and radio recordings, phone conversations, everyday life situations, classroom talk, etc. Its geographical scope ranges from USA, to Canada, UK and Ireland, however, those texts not coming from the United Kingdom were removed to avoid skewedness in the results as well as the transcriptions of the oral samples, given the nature of the study corpus, solely made up of written texts.

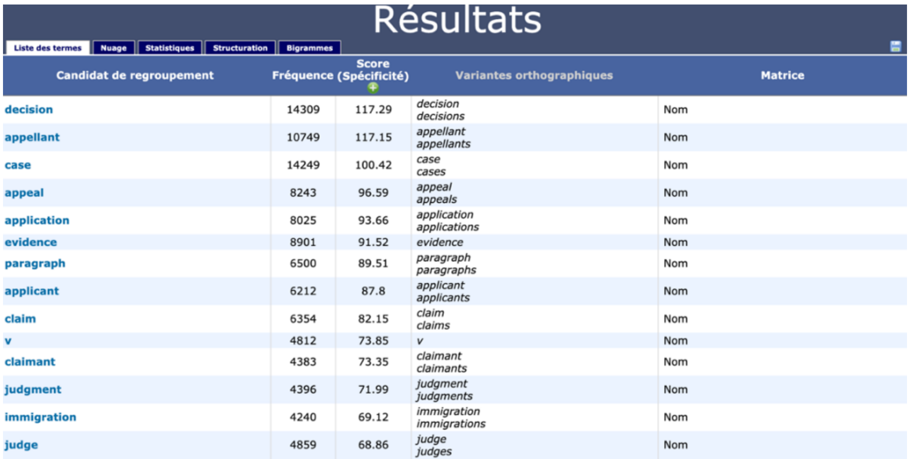
As regards its Spanish counterpart, it was extracted from a larger collection of Wikipedia articles compiled by Reese et al. (2010) in Spanish. The Spanish sample used in the present study comprises 94 texts which roughly reach the 100-million-word target. The range of topics covered by the Spanish reference corpus is wide, touching upon areas such as history, science, medicine or literature as well as other general language areas other than the legal one. This corpus was downloaded from the authors' website[2], which allows users to obtain the texts easily and store them in raw text format for later processing at no cost. The format of these texts does not coincide with the length of the original Wikipedia articles, as each of the sections of the original corpus resulted from merging together different sets of the articles, hence the length of the texts. The texts were rearranged and the word target reduced to facilitate the processing stage, as the software could not cope with the entire corpus as downloaded from the authors' website.

3.1.2 TermoStat

As well as *Keywords, TermoStat*[3] (Drouin 2003) is a user-friendly online tool which manages to extract the specialized terms in a corpus in several languages like French, English, Spanish, Italian or Portuguese. Drouin's technique could be regarded as a hybrid one as it relies both on grammatical and statistical information for term identification, using corpus comparison to that end.

The output term lists are ranked according to their level of specificity, in conjunction with their classification into morphological categories (nouns, verbs, adjectives and adverbs). The system computes lemma frequency (the frequency of the root word including all of its possible realizations) instead of type frequency. The lemmatization of the corpus is implemented with Schmidt's *Tree Tagger* (1999), which also allows for the POS (part of speech) tagging of the corpus texts. The software offers the possibility of retrieving multi-word terms although its degree of accuracy decreases if compared with single-word term identification. Figure 1 displays the results obtained online after processing the legal corpus at hand with *TermoStat*, where the lemma of the selected terms, their frequency, specificity coefficient, variants and POS (part of speech) tag are shown.

Drouin's software does not require the upload of a reference corpus to the online database, as it contains its own general reference corpora in various languages to perform the comparison with the specialized text collection uploaded by the users and the subsequent recognition of specialized terms.



Candidat de regroupement	Score Fréquence (Spécificité)		Variantes orthographiques	Matrice
decision	14309	117.29	*decision* *decisions*	Nom
appellant	10749	117.15	*appellant* *appellants*	Nom
case	14249	100.42	*case* *cases*	Nom
appeal	8243	96.59	*appeal* *appeals*	Nom
application	8025	93.66	*application* *applications*	Nom
evidence	8901	91.52	*evidence*	Nom
paragraph	6500	89.51	*paragraph* *paragraphs*	Nom
applicant	6212	87.8	*applicant* *applicants*	Nom
claim	6354	82.15	*claim* *claims*	Nom
v	4812	73.85	*v*	Nom
claimant	4383	73.35	*claimant* *claimants*	Nom
judgment	4396	71.99	*judgment* *judgments*	Nom
immigration	4240	69.12	*immigration* *immigrations*	Nom
judge	4859	68.86	*judge* *judges*	Nom

Figure 1. Screenshot of Drouin's output CT list: English corpus

As already stated, Drouin's (2003) technique relies on corpus comparison by focusing on the statistical behaviour displayed by the CTs in the specialized context as opposed to the general one. In Drouin's own terms:

> This technique, which relies on standard normal distribution, gives us access to two criteria to quantify the specificity of the items in the set: (1) the test-value, which is a standardized view of the frequency of the lexical units, and (2) the probability of observing an item with a frequency equal to or higher than the one observed in the AC. Because the probability values decline rapidly, we decided to use the test-value since it permits much more granularity in the results. (Drouin 2003, 101).

In order to determine the degree of precision of the software in the identification of specialized terms within the field of telecommunications, Drouin resorts to specialized referees, who manually evaluate the validity of the CT lists provided, insisting on the subjective character of such validation methods. Automatic validation complements the evaluation process by comparison with a telecommunications terminological database, which yields 86% precision in single-word term retrieval.

As a final point, Drouin puts great emphasis on the need to explore the context of usage of those terms which activate a specialized meaning when in contact with the technical environment, the so-called semitechnical terms, often prone to displaying a peculiar statistical behaviour and to trick automatic systems solely based on corpus comparison.

3.1.3 Chung

Similarly to *Keywords* (Scott 2008a) and *TermoStat* (Drouin 2003), Chung's (2003) ATR method implements the corpus comparison technique based on the observation of term frequency both in the general and the specialized areas. The author sets a ratio threshold to tell apart terms from non-terms, using the value > 50 as the cut-off point for a specialized term to be reckoned as such. As the validation method required manual supervision on the part of the referees, a cut-off point was established for the top 500 CTs once the list of CTs was filtered[4]. The frequency ratio for the top 500 CTs ranged from 3652.24 to 87.33 in Spanish and from 4461.11 to 181.45 in English.

Chung's method is not part of a software package or an application, yet, its calculation is quite straightforward. It solely requires obtaining two frequency lists by processing a specialized corpus and a general language one with any software application like *WordSmith* (Scott 2008a) or *AntConc* (Anthony 2020). Then, the frequency scores are normalized by dividing a word's raw frequency by the number of tokens or running words in the corpus[5], this normalization procedure allows for the comparison between two datasets of different size. Once we calculate the normalized frequency of the items in both word lists, the ratio of occurrence of every word in the lists is determined. A word's ratio of occurrence can be obtained by dividing its normalized frequency in the specialized corpus by the same parameter in the general one. Those words standing above the >50 ratio threshold would be regarded as specialized terms, given their higher frequency values in the technical corpus.

As well as Drouin, Chung assesses the efficiency of her method through automatic and manual validation. She asks two referees, who were experts in the field of anatomy, to classify the terms in a sample text taken from her anatomy corpus into four categories depending on their level of specialization. She classifies all the words in the corpus after calculating their ratio of occurrence and also produces four groups based on the results. After comparing the specialists lists with her own, she finds 86% overlap between the most specialized group of terms found by the referees and the ones included in her lists, automatically determined on the basis of their ratio of occurrence.

3.2 Corpus description

Terminological extraction commonly requires comparing a specialized corpus against a general one. A vast majority of ATR methods resort to corpus comparison as a pivotal procedure for automatic term extraction, as already pinpointed in the literature review section. This is why the four corpora included in Table 1 were necessary for the present research so as to facilitate term retrieval in both languages.

Table 1. Corpora

Corpus/language	**# Tokens**	**# Types**	**# Texts**
Legal English corpus	2,396,985 **(2.4m)**	20,236	600
Legal Spanish corpus	3,723,587 **(3.7m)**	25,268	600
LACELL (general Spanish corpus)	14,830,302	264,609	8
Wikicorpus (Reese et al., 2010)	101,322,383	732,795	94

Table 1 comprises the four corpora used in this study, two of which are made up of 600 legal texts each, all of them judicial decisions issued by Spanish and British courts between 2016 and 2017. Since they differ in size, there was a need to normalize frequency scores for comparison[6]. The British text collection comprises roughly 3.7 million words while its Spanish counterpart has 2.4 million tokens (or running words). The texts in these two corpora were obtained from two major databases: the *CENDOJ*[7] (the Spanish legal documentation centre) and the BAILII[8] (the British and Irish Legal Information Institute). Both text collections were compiled so as to be used in different contexts such as corpus-based discourse analysis on migration[9] as well as for the validation of ATR methods in the legal field, which is the objective of the present study. For the texts to be equivalent in generic terms, in spite of their intrinsic differences, the search configuration was set for the engine on the BAILII website to only retrieve British judgments within the case law section, as it offers access to a plethora of different legal texts in English from various sources. In a similar fashion, the Spanish search engine was configured to extract solely those texts under the category *sentencias,* without excluding any court or tribunal regardless of its position within the judicial hierarchy.

The general corpora acting as reference in this research, LACELL and the *Wikicorpus* were not required for the implementation of Drouin's (2003) ATR method, as the software already includes general corpora in different languages to implement the comparison between the general and the specialized fields, which facilitates the task greatly.

Concerning the English reference corpus, LACELL, it is composed of 14.8 million tokens and 264,609 types, that is, every different wordform in a corpus regardless of the number of times it occurs in it. The section of the corpus employed herein excludes those texts not coming from British sources. On the other hand, the *Wikicorpus* (Reese et al., 2010) is made up of roughly

101 million words obtained from Wikipedia articles on many different topics such as history, science or literature, amongst many others, as stated above.

3.3 Method implementation

The degree of complexity involved in the implementation of the three methods selected for validation varies greatly depending on their degree of automatization. Given that both *Keywords* (Scott 2008a) and *TermoStat* (Drouin 2003) are integrated into software applications, it was relatively easy to process both legal corpora as well as the reference ones.

In the first place, LACELL and the legal English corpus were analysed with *WordSmith* (Scott 2008a) to obtain the frequency lists necessary for the software to extract the legal keywords. A similar process was followed to obtain the Spanish set. Then, the system was configured to implement the log-likelihood test (Dunning 1993), which delves into the frequency lists extracted from general and specialized corpora by comparing the frequency scores of the terms in both contexts as well as other statistical data such as distribution. Rayson and Garside (2000) provide a clear description of how keyness is calculated by implementing Dunning's log-likelihood test[10].

A frequency threshold of >3 was established for the system to identify the keywords in both languages with the purpose of discarding those lexical items occurring rarely whose significance would be almost null. In fact, 32.76% of the words in the legal Spanish corpus and 35.77% in the English set were *hapax legomena*, that is, lexical items which can only be found once in a corpus. The amount of *dis legomena*, those terms occurring solely twice, was not so high although 13.27% were identified in the Spanish text collection as opposed to 15.35% in the British corpus. The system produced a set of 4,550 positive keywords in English and 4,028 in Spanish out of which the top 500 were singled out for manual validation.

Table 2 displays the top 20 English keywords obtained prior to the validation process. This sample, in spite of its limited size, illustrates how the terms which are pushed towards the top of the term inventory based on their statistical behaviour, when contrasted with the general corpus, inform on the generic features of the texts themselves. Let us remind the readers about the major features of the legal corpora at hand, which solely comprise judicial decisions, where terms like *decision* (K[11]=32,991) itself, *appeal* (K=29,484), *tribunal* (K=25,938) or *court* (K=24,042) are extremely common. The list of keywords also informs about the actors in judicial proceedings, finding terms like *appellant* (K=24,113), *judge* (K=20,659) or *claimant* (K=15,010) at the top of the term inventory. It can also be observed that, in spite of their lack of terminological value, some function words (*that, the*) entered the top 20 term list, unlike the other three ATR methods tested. However, except for the term *immigration* (K=25,938), unlike Drouin's and Chung's methods, the rest of the keywords below do not throw any light on the major topics the corpus might be articulated around, these elements tend to be pushed to lower positions in the list based on their keyness value.

Table 2. TOP 20 English Keywords as extracted by *WordSmith* (Scott 2018a)

N	**Key Word**	**Freq Leg Corpus**	Keyness	**P**
1	decision	13879	32991.5547	3.1321E-23
2	appeal	11214	29484.0508	4.6024E-23
3	immigration	8512	25938.0547	5.714E-23
4	tribunal	7824	24135.7734	5.7154E-23
5	appellant	7574	24133.7871	5.7807E-23
6	court	11428	24042.8301	6.3717E-23
7	that	81725	23276.2383	6.6486E-23
8	the	305801	21986.9082	9.118E-23
9	judge	7860	20659.1582	9.2531E-23
10	article	7966	20558.25	1.3641E-22
11	case	11742	18067.6367	1.4952E-22
12	evidence	8861	17524.1777	1.8831E-22
13	paragraph	6111	16229.8799	1.9861E-22
14	application	7144	15945.0762	2.3816E-22
15	claimant	4930	15010.2969	2.5946E-22
16	state	9264	14588.8438	3.5848E-22
17	Mr	10959	13102.0137	3.8158E-22
18	respondent	4015	12832.7441	3.846E-22
19	V	6244	12799.2266	3.9256E-22
20	secretary	6258	12712.3477	4.1035E-22

Secondly, Drouin's online software, *TermoStat* (2003), was tested. It facilitates the processing task greatly. Drouin's software is lodged online and only requires the user to register and to upload the corpus to the server, being capable of processing single text files in raw text format (up to 30 Mb) relatively quickly. After analysing both texts collections, two lists of legal terms were obtained in Spanish (4,519 CTs) and English (2,233 CTs). The validation procedure was similar to that applied to Scott's (2008a) software, whereby the top 500 CTs were selected.

In this case, the set of Spanish terms was taken as sample of the top 20 CTs produced by the software before it was actually validated. As presented below, in Table 3, although the list of terms contains some items which point at the generic character of the texts in the corpus (similarly to Table 2) like *sentencia (sentence,* S[12]=347.41*)*, *recurso* (*appeal,* S=208.87), *apelación* (*appeal,* S=202.07), *juzgado* (*court,* S=148.34) or *tribunal* (*court/tribunal,* S=141.22), it also points in other directions, since other terms are pushed to the top of the list which relate to procedural legal lexicon, for instance, *multar* (*to fine*, S=205.56) or *sanción* (*penalty*, S=200.09). Nevertheless, unlike the previous method, *TermoStat* reveals some lexical items which are fundamental for the analysis of the texts in the corpus, which, as stated above, revolve around the topic of immigration. Words like *expulsión* (*deportation*, S= 296.24), *territorio* (territory, S=145.85), *permanencia* (permanence, S=142.08) or *retorno* (*return*, S=139.34) are highly representative of the legal trouble migrants might go through when they are subject to legal proceedings in a foreign country. The statistical significance assigned to such terms might also be indicative of their thematic relevance within this text collection.

Together with the statistical data associated to each term (columns 2 and 3), which are lemmatized, that is, their frequency is computed with regard to the root word (shown in the first column), we are offered the frequency score, in column 4, as well as its POS tag (fifth column).

Table 3. Top 20 Spanish terms as extracted by TermoStat (Drouin 2003)

Candidate (Grouping Variant)	Frequency	Specificity	Variants	Pattern
sentencia	12505	347.41	sentencia___sentencias	Commo Noun
expulsión	8634	296.24	expulsión___expulsiones	Common_Noun
recurso	9906	208.87	recurso___recursos	Common_Noun
multar	4274	205.56	multar___multas	Verb
apelación	4110	202.07	apelación___apelaciones	Common_Noun
sanción	5937	200.09	sanción___sanciones	Common_Noun
artículo	9796	189.55	artículo___artÌculos	Common_Noun
recurrente	3335	185.82	recurrente___recurrentes	Adjective
art	3193	184.68	art	Commo Noun
jurisprudencia	3088	168.87	jurisprudencia	Common Noun
irregular	2789	167.38	irregular___irregulares	Adjective
apartado	5193	155.49	apartado___apartados	Common Noun
contencioso	2226	150.16	contencioso	Adjective
auto	2220	149.83	auto___autos	Common Noun
juzgado	2052	148.34	juzgado	Common Noun
territorio	4442	145.85	territorio___territorios	Common Noun
administrativo	3931	144.4	administrativo___administrativos___administrativas	Adjective
permanencia	2011	142.08	permanencia	Common Noun
tribunal	4799	141.22	tribunal___tribunales	Common Noun
retorno	2224	139.34	retorno	Common Noun

Finally, Chung's ATR method required more steps until both frequency lists (the specialized and the general one) were ready to be processed and the ratio of occurrence could be calculated. Chung's method relies on frequency as the sole parameter for term identification and two wordlists are needed to calculate it. They must be obtained using software like Scott's (2008a) or Anthony's (2020) and then, a comparison must be established. This is done by dividing the normalized frequency of each term in the legal corpus by the same parameter in the general one. By applying the appropriate formulas, the calculation process can become semi-automatic if an excel spreadsheet is used.

Once the CT list was arranged according to the ratio value, misspelled words had to be removed in the first place. Chung's method requires the manual filtering of these elements as they do not occur in the general corpus and would be automatically classified as terms, although their value for terminological analysis is void. As a matter of fact, the group of terms not found in the reference corpus might comprise not only misspelled words but also proper names, whose statistical relevance in judicial decisions is considerable, as Marín (2014) acknowledges, but their thematic content is null.

Having also removed *hapax* and *dis legomena*, and having applied the >3 frequency threshold, two lists of 12,393 Spanish and 16,260 English CTs were ranked according to their ratio value. Following a similar validation procedure to the one applied to *Keywords* (Scott 2008a) and *TermoStat*

(Drouin 2003), the top 500 CTs both in Spanish and English were selected. The major thematic category which the top 20 CTs belong in (see in Table 4), as ranked by Chung's ratio method and similarly to *Keywords*, evidences the corpus texts genre, judicial decisions. In Table 4 we find words such as *respondent* (FR[13]=4461), *appellants* (FR=4356), *petitioner* (FR=1272), or *tribunal* (FR=704), as well as acronyms like *CPR* (Civil Procedure Rules, FR=577), *FCO* (Foreign and Commonwealth Office, FR=640) or *UT* (Upper Tribunal, FR=583), which point in a similar direction.

On the other hand, the acronyms *IA* (Immigration Act) or *UNHCR* (United Nations High Commissioner for Refugees) as well as the term *deportation* (FR=955) are indicative of the major topic that the corpus texts are based on, that is, immigration. Along these lines, the verb *erred* (FR=677) or the noun *proportionality* (FR=555), which might potentially convey attitudinal meanings, could also be of interest in connection to the study of the legal circumstances that surround migration processes.

Table 4. Top 20 English terms as extracted by Chung's method

Term	Freq Leg Corpus	Normed Freq Leg Corpus	Freq Leg Corpus	Normed Freq Leg Corpus	Chung's ratio
respondent	4015	10.9462	9	0.0025	4461.1111
appellants	1307	3.5633	3	0.0008	4356.6667
IA	1170	3.1898	3	0.0008	3900.0000
appellant	7574	20.6492	24	0.0065	3155.8333
EU	1272	3.4679	7	0.0019	1817.1429
petitioner	636	1.7339	5	0.0014	1272.0000
appellant's	311	0.8479	3	0.0008	1036.6667

deportation	2483	6.7695	26	0.0071	955.0000
tribunal	7824	21.3308	111	0.0303	704.8649
UNHCR	412	1.1232	6	0.0016	686.6667
erred	474	1.2923	7	0.0019	677.1429
submits	798	2.1756	12	0.0033	665.0000
FCO	192	0.5235	3	0.0008	640.0000
UT	467	1.2732	8	0.0022	583.7500
CPR	231	0.6298	4	0.0011	577.5000
proportionality	777	2.1184	14	0.0038	555.0000
subsection	441	1.2023	8	0.0022	551.2500
paras	385	1.0496	7	0.0019	550.0000
Sudanese	220	0.5998	4	0.0011	550.0000
WLR	474	1.2923	9	0.0025	526.6667

4. Results and discussion

4.1 Method validation

The implementation procedure followed to identify the legal terms in the Spanish and English corpora led to obtaining two CT inventories per method. The lists were ranked according to the parameters set by each author. Nevertheless, adopting a quantitative perspective, it became necessary to determine the degree of efficiency achieved by each method in order to decide which of them was the most precise in recognizing terms automatically. To that end, as stated above, two specialists were requested to manually supervise the top 500 CTs in each of the six term lists extracted. After classifying the terms in the categories described above and merging together those which both specialists deemed either highly specialized or semi-technical, average precision was calculated for each method by finding the percentage of TTs extracted out of the 500 CTs selected, as shown in Figure 2, for the English corpus and in Figure 3 for the Spanish one.

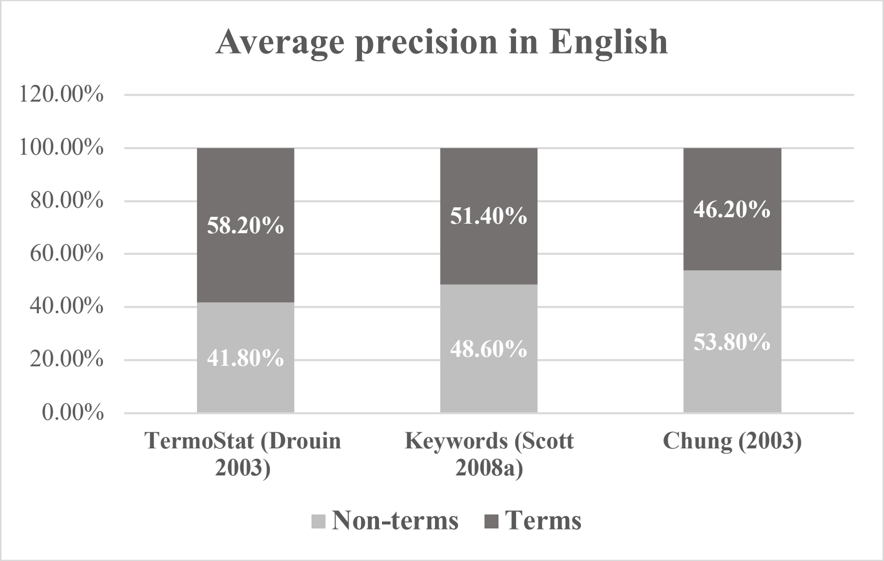


Figure 2. Average precision achieved for the top 500 CTs: English corpus

Figure 2 illustrates the differences found among the three methods assessed in this study when applied to an English legal corpus, finding that Drouin’s software, the most precise ATR method, managed to identify 58.2% TTs on average. Although its degree of precision might seem slightly low, particularly if compared with the figures provided by the authors themselves, it must be emphasized that the process of validation was not automatic. It was performed through the implementation of an inter-rater reliability test which implied discarding some items from the lists when there was no consensus between the referees. This lack of agreement was often caused by semitechnical terms, whose significant presence both in the general and the specialized context made the referees doubt and often invalidate the inclusion of words in the lists like *razonamiento* (*reasoning*), *causa* (*cause*), *judge* or *trial*[14], thus diminishing the proportion of TTs confirmed as such.

Nevertheless, the fact that a CT was excluded from the final list of terms after filtering does not imply that it may not be of interest for the researcher willing to examine the major *topoi* or themes in the corpus. It simply points at a lack of specificity of a considerable number of terms, whose presence and statistical relevance in the general field, something which is probably one of the most distinctive lexical features of legal terms, prevents them from being automatically deemed specific. Similarly, *Keywords,* which stands in second position, managed to extract 51.4% TTs on average successfully, in spite of it not being conceived as an ATR method proper. Yet, the function it performs by retrieving the most statistically significant lexical items in a specialized corpus is very similar to those which were designed specifically to that end.

The third position is occupied by Chung's method which, after manually filtering typos and other meaningless units such as proper names, reaches 47,4% average precision. If the discarded elements had been included in the validation lists, the degree of efficiency of this method would have probably dropped dramatically, since a great proportion of such items are not terms or have no terminological value.

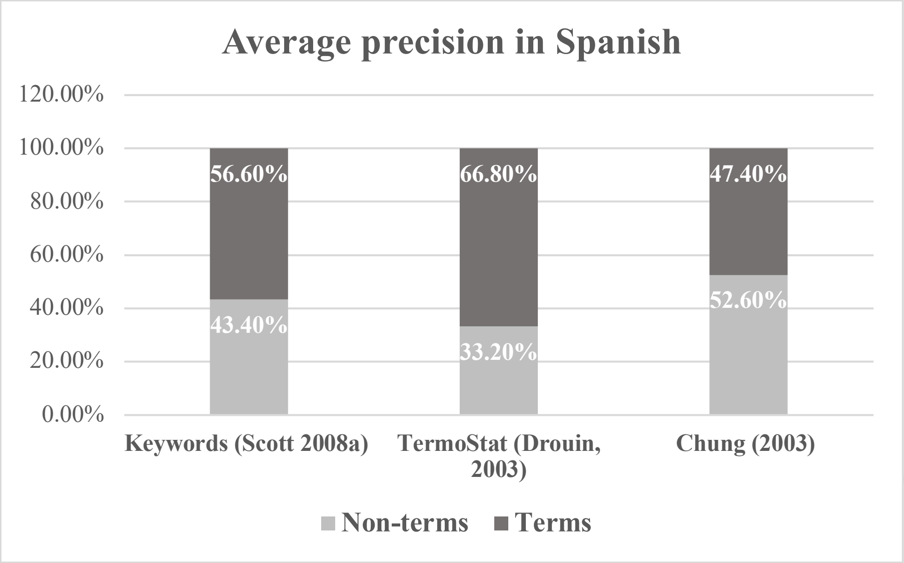


Figure 3. Average precision achieved for the top 500 CTs: Spanish corpus

The results obtained after validating the CT lists in Spanish, as displayed by Figure 3, are slightly higher than the ones described above. In fact, the most precise method, *TermoStat*, identifies 66.8% TTs within the Spanish corpus, while it only retrieves 58.2% from its English counterpart. Although this is mere conjecture, the higher degree of success of this and the other two methods in Spanish might be indicative of the statistical behaviour of Spanish legal terms, which must necessarily differ on average from their English equivalents as regards their frequency and distribution in the general and specialized contexts. However, to confirm this perception, it would be necessary to delve much deeper into the algorithm designed by Drouin and the term retrieval process itself, which does not fall within the scope of this study.

Along these lines, *Keywords* also performs more efficiently in Spanish, standing 5 points above the results obtained in English, achieving 56.6% precision and ranking second. Similarly, Chung's method appears to be more precise in automatic term retrieval when implemented on the Spanish corpus, although the difference is marginal, just 1.20 points higher (47.4%) than in English. Even so, it is the least efficient method in both languages, and also, the most complex to implement, let alone the noise levels generated by the automatic inclusion of elements not found in the reference corpus, which required manual filtering prior to its validation.

In order to reinforce the results shown above, which consisted in calculating precision by means of human validation, recall was also assessed automatically. The automatization of the process was accomplished by comparison with a golden standard, that is, an electronic glossary of legal terms in English stored in an Excel spreadsheet, consisting of a list of 8,715 items taken from different legal term glossaries in raw text format (as defined in Marín, 2014). The term recall refers to the amount of TTs extracted by an ATR method with respect to the entire list of CTs identified in the corpus, not to a single set such as the top 500 CTs displayed above. This parameter could only be measured within the English corpus, as there was no Spanish gold standard to be used as reference.

As illustrated by Figure 4, it is Drouin's technique (2003) which reaches not only the highest precision levels, but also ranks first as regards recall, since it manages to identify 35.3% TTs out of the entire list of items extracted (2,233). It is closely followed by *Keywords*, which obtains 29.2% for this parameter, while Chung's method performs poorly, only managing to recognize 12.5% terms in English.

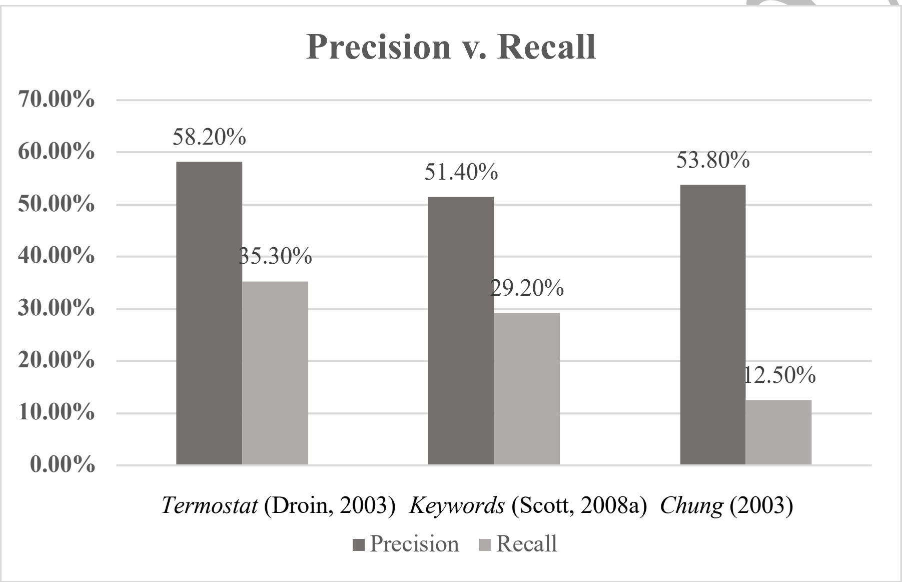


Figure 4. Precision and recall: English corpus

All in all, taking into consideration, not only its user-friendliness as a tool (Scott's software requires greater expertise as many parameters must be adjusted and the options and applications within it are greater), but also the fact that it does not require the use of a reference corpus or wordlist to be uploaded to the system, it is Drouin's method which stands out as the most effective one for legal term extraction out of the three techniques assessed in this study both in Spanish, where it appears to be more efficient, and in English. The results reflected on Figure 4 above reinforce this perception, as *TermoStat* (Drouin 2003) appears to excel the other two methods not only in terms of precision but also regarding recall when implemented on the English corpus.

4.2 Thematic term categories

4.2.1 Corpus-driven semantic classification

In spite of the fact that ATR methods are designed to identify those lexical elements whose statistical salience in specialized contexts make them stand out when put against the general field, they are also useful tools to detect major thematic areas and unveil topics that may otherwise remain unnoticed, particularly when dealing with large text collections. In doing so, they allow for an in-depth examination of the context of usage of those terms automatically identified by the ATR methods selected so, depending on the research objectives established, it might be interesting to select a specific technique not only based on its efficiency in identifying TTs, but also on its capacity to provide a wider picture of the themes or topics a corpus might revolve around, other than solely focusing on specialized terms pertaining to the legal area, as is the case.

In order to determine which of the methods assessed in the present research was capable of signaling a wider range of themes or topics, the top 500 CTs extracted by each method in English and Spanish were examined and classified into five major semantic categories. These thematic categories were defined following a corpus-driven approach, that is, they were identified on the basis of the observation of the items contained in the lists themselves, finding that the largest proportion of such items belonged in the category *general legal terms*, as shown in Table 5. The rest of the themes identified amongst the top terms in each output list were *territory*, *evaluative items*, *family* and *crime/punishment.* From a qualitative perspective, the study of the last four groups might complement the legal term inventory as they point at relevant topics in both corpora other than legal terms *stricto sensu*. Let us remind the reader that the measure which was employed to identify and rank the terms indicates their statistical saliency in comparison with a general language corpus, hence the greater presence of these items amongst the top terms.

Those elements comprised within the theme *territory* are closely linked to the major topic that the texts revolve around, immigration, finding words such as *asylum*, *deportation*, *nacionalidad* (*nationality*) or *extranjero* (*foreigner*) amongst its constituents in both languages. Secondly, the group *evaluative items* embraces those terms which could potentially express the speaker's attitude towards the propositional content of the texts, including words like *vulnerable*, *degrading* or *inhuman* in English and *grave* (*serious*) or *indefensión* (*helplessness*) in Spanish. In the third place, the concept *family* gathers words which point at familiar issues or concerns that relate to immigration. This is the case of words like *marriage*, *spouse*, *matrimonio* (*marriage*) or *tutela (guardianship)*. The last category, *crime*, comprises those items which either explicitly refer to crime itself, for instance, *trafficking*, *torture*, *offence, trata (human trafficking)* or *infracción* (*breach*) or rather signal the consequences of committing a crime: *detain*, *imprisonment*, *multar* (*fine*) or *sanción* (*penalty*).

Table 5. Thematic categorization of terms

	TermoStat (Drouin 2003)				
	Legal terms	*Territory*	*Evaluative items*	*Family*	*Crime*

English Corpus	**73.6%** *appellant* *decision* *administrative* *appellate* *cross-examination* *mandatory* *affidavit*	**9.24%** *asylum-seeker* *returnee* *entrant* *extradition* *resident* *migrant* *return* *immigrant* *refugee* *domestic*	**10.95%** *vulnerable* *manifestly* *reasonableness* *degrading* *inhuman* *unfounded* *unfairness* *irrational* *inconsistency*	**1.02%** *marriage* *father* *spouse*	**4.79%** *criminal* *trafficker* *offender* *offend* *breach* *criminal* *detainee* *torture* *trafficking* *offence*
	Keywords (Scott 2008a)				
	Legal terms	*Territory*	*Evaluative items*	*Family*	*Crime*
	82.1% *decision* *appeal* *tribunal* *appellant* *court* *judge* *article* *case*	**7.39%** *immigration* *asylum* *deportation* *residence* *jurisdiction* *refugee* *nationality* *entry*	**4.66%** *error* *proportionality* *reasonable* *erred* *arguable* *credibility* *disproportionate* *proportionate*	**0.38%** *spouse*	**5.44%** *detention* *detained* *trafficking* *criminal* *persecution* *imprisonment* *offence* *torture*
	Chung (2003)				

	Legal terms	*Territory*	*Evaluative items*	*Family*	*Crime*
	72.52% *respondent* *appellants* *petitioner* *tribunal* *UNHCR* *submits* *UT* *cpr*	**13.06%** *deportation* *immigration* *asylum* *deport* *deporting* *reside* *relocation* *EU* *stateless*	**9.9%** *proportionate* *insurmountable* *disproportionate* *mistreated* *unfairness* *mistreatment* *erroneously* *defamatory* *fraudulently*	**0%**	**4.5%** *detention* *detainee* *trafficking* *detaining* *detained* *breach* *breaches* *detainees* *infringed*
	TermoStat (Drouin 2003)				
	Legal terms	*Territory*	*Evaluative items*	*Family*	*Crime*

Spanish Corpus	**87.12%** *sentencia* *recurso* *apelación* *artículo* *art* *jurisprudencia* *contencioso* *auto* *juzgado* *administrativo* *tribunal* *sección*	**7.48%** *territorio* *permanencia* *retorno* *estancia* *residencia* *extranjero* *extranjería* *empadronado*	**1.19%** *irregular* *proporcional* *indefensión* *privativa* *grave* *agravantes*	**3.59%** *matrimonio* *reagrupación* *esposo* *familiar* *reagrupante* *cónyuge* *matrimonial* *tutela* *arraigar*	**0.59%** *multar* *sanción* *lesión* *criminal*
	Keywords (Scott 2008a)				
	Legal terms	*Territory*	*Evaluative items*	*Family*	*Crime*
	89.4% *sentencia* *recurso* *administrativo* *contencioso* *jurisprudencia* *tribunal* *directiva* *procedimiento*	**4.22%** *retorno* *residencia* *extranjero* *estancia* *nacionales* *nacionalidad* *asilo* *Schengen*	**2.45%** *irregular* *proporcionalidad* *controvertida* *irregularmente* *grave* *pretensiones* *debidamente* *proporcionada*	**1.4%** *arraigo* *familiar* *matrimonio* *reagrupante*	**2.46%** *multa* *penal* *delito* *indocumentado* *infracciones* *trata* *pena* *criminal*
	Chung (2003)				
	Legal terms	*Territory*	*Evaluative items*	*Family*	*Crime*

	96.42%	**0.51%**	**0.51%**	**2.55%**	**0%**
	apelante	*empadronado*	*desvirtuado*	*reagrupante*	
	apelada			*reagrupada*	
	impugnada			*reagrupar*	
	roj			*reagrupado*	
	cendoj			*ascendientes*	
	stsj				
	tjue				
	jurisprudencial				
	loex				

Table 5 also displays the proportion of terms included in each category (expressed in percentages) with respect to the top 500 terms in each language for each method. As a whole, general legal terms such as *appellant* or *cross-examination* in English or *recurso* (*appeal*) and *auto* (*court order*) in Spanish, as was to be expected, stand out as the most numerous category. The rationale behind this result is that the principal technique which the three assessed methods rely upon is corpus comparison. Regardless of the greater or lesser degree of sophistication of the algorithms employed in ATR, the comparison of a specialized corpus against a general one, using frequency as the major parameter for term retrieval, necessarily implies that highly specialized terms, whose frequency of usage in general language is low, will be pushed to the top of the term ranking. This becomes more evident in Spanish, finding that it is the predominant thematic category and contains practically the entirety of the terms retrieved (96.42%), especially after applying Chung's (2003) method.

The results were similar in English, although the proportion of general legal terms was lower. *Keywords* (Scott 2008a) is the method that identified the largest amount of these, 82.1%, although it also achieves to bring to the forefront other thematic areas like *territory*, *evaluation* or *crime,* including 7.39%, 4.66% and 5.44% items respectively. As a whole, although ATR method precision might be higher in Spanish, as demonstrated in Section 4.1, judging by the figures displayed in Table 5, the capacity of these methods to identify a wider array of topics is not so high in this language. Nonetheless, *TermoStat* (Drouin 2003) appears to be the one that extracts a greater proportion of items belonging in the groups *territory* (7.48% elements), *family* (3.59%) and *evaluation* (1.19%). On the contrary, Chung's method, which only contrasts term frequency without considering other parameters like distribution or probability (broadly speaking), identifies a marginal number of terms other than those in the legal term group. As illustrated in Table 5, except for the category *family*, where we find 2,55% of the terms extracted, the remaining three, *territory*, *evaluation* and *crime* do not even reach 1%.

Therefore, leaving aside the thematic group *legal terms*, which, as stated above, clearly refers to the legal genre the corpus texts belong in, that of judicial decisions (the terms *appellant*, *judge*, *case*, *court* or *tribunal* instantiate this fact), except for Chung's output list in Spanish and, in general, the category *family* in English, the three methods offer a wide variety of examples that might act as a point of departure for the further exploration of the corpora at hand.

To begin with, the thematic group *territory* in English clearly highlights the relevance of asylum requests as a major subject which the English corpus revolves around, given that the terms *asylum* and *refugee* are amongst the top terms extracted by the three ATR methods. Similarly, other terms like *extradition*, *deport, deportation* or *relocation* relate to this topic and can be found in the three term lists. In a similar fashion, the concept of *residence* connects with *asylum* and *deportation*, as well as other realizations of that lemma, namely, *resident*, *residence* or *reside*.

On the contrary, the Spanish group *territory* is less populated and does not seem to demonstrate such a strong connection with the notion of *asylum* as the English corpus does. In fact, the term *asilo* was only retrieved by *Keywords* (Scott 2008a) in Spanish. However, the term *residence* deploys itself throughout the Spanish corpus as well as it does in its English counterpart although its presence is more relevant, covering a considerable proportion of the items in the category, terms like *residencia (residence)*, *empadronado (registered as resident)*, *permanencia (permanence)*, *nacional (national/domestic)*, *nacionalidad* (*nationality)* or *estancia (stay)* exemplify this circumstance.

Similarly to *territory*, the category *evaluative terms* is considerably numerous in English, as presented in Table 5, containing 8.5% terms on average in contrast with the Spanish set, where we only find 1.38% of these elements. Even so, the items comprised in both text collections have something in common, their negative connotations. Terms like *degrading, inhuman, irrational, disproportionate, insurmountable* or *mistreatment* in English and *indefensión (helplessness), grave (serious), controvertida (controversial)* or *desvirtuado (distorted)* in Spanish convey the attitudinal positioning on the part of the speaker that might be worth further scrutiny, since these elements may point at sensitive topics in connection with immigration and help to characterize this phenomenon as seen through the eyes of the judiciary.

On the other hand, the degree of representativeness of the category *family* in the English corpus is barely inexistent, comprising only 0,45% terms on average, yet, the elements within this group and the statistical data associated with their usage might signal the relevance that family issues have in migration processes. Words like *marriage*, *father* or *spouse* illustrate this trend. Likewise, the data provided by the Spanish corpus in relation to familiar issues (including 2.5% terms on average), which partially overlap with the items retrieved from the English text collection, enrich our perception of the fundamental role played by families in migration processes and their connection with the legal scenario. As well as other items like *matrimonio* (*marriage*), *familiar* (*familiar*) or *esposo* (*husband*), the lemma *reagrupar* (bringing the members of a family back together) and all its variants, coupled with *tutela* (*guardianship*) and *arraigar* (take root in a country), insist on the need migrants express to keep their families reunited and the essential role that children play in legal processes related to immigration. Still, a closer examination of the context of usage of all these terms would be necessary to reach sound conclusions in relation to this and other topics enumerated in this section. However, such analysis falls out of the scope of the present research.

Lastly, the category *crime*, as was to be expected, stands third as regards the number of terms it gathers in English (4.8% on average), whereas in Spanish it roughly reaches 1%. Let us insist on the fact that Chung's method does not extract any of these elements from this text collection. The terms which the three ATR methods at hand identified as members of this category basically revolve around two axes, on the one hand, general legal terms associated to criminal behaviour and its punishment such as *offender*, *breach*, *imprisonment*, *lesión* (*injuries*) or *multa* (*penalty*) and, on the other hand, specific terms referring to actual criminal activities like *trafficking*, *torture* or *persecution*, which might deserve specific attention. Their context of usage should be explored further though, so as to clarify the specific conditions displayed in judicial decisions that might present migrants as victims of human trafficking or torture, being persecuted in their home countries or threatened and forced to be part of this criminal activity, or as an active part of human trafficking networks and members of criminal organizations.

4.2.2 Semantic categorization using UMUTextStats

As suggested by Bisceglia, Calabrese, and Leone (2014), and Jumaquio-Ardales, Oco, and Madula (2017), the use of Natural Language Processing (NLP) tools in combination with more standard corpus analysis techniques such as keyword analysis or collocate extraction might also enhance our knowledge of the semantic and morphological categories of the lexicon in a corpus. This is why, this section introduces the automatic categorization of the lexical items found in both corpora using the software *UMUTextStats* (García-Díaz et al. 2018; García-Díaz, Cánovas-García, and Valencia-García 2020), a text classification software, built on similar technology to the well-known *Language Inquiry and Word Count –LIWC* (Pennebaker, and Francis 1999), which could be regarded as a useful tool to examine the emotional, cognitive and structural components contained in language on a word-by-word basis by determining the percentage of words which belong in those categories. The major difference between *LIWC* and *UMUTextStats* lies in the fact that the latter adds a linguistic basis of European Spanish and also several categories which are not word-based. The software described herein can process large amounts of text and the result is a vector consisting of different features which range from grammatical information such as the total amount of pronouns, negations, or auxiliary verbs (amongst other) to other psycholinguistic categories like emotions, named entities, or cognitive processes.

It is worth noting that in the dictionaries used by the software, lexical items were formalized by means of regular expressions, that is to say, search strings that can be used to specify sequences of characters to be extracted from a text or corpus (Jurafsky, and Martin 2019). Thus, for instance, *doméstico/a/os/as* (*domestic*) was formalized as *doméstic[oa]s*?, which is interpreted by the software as the string of characters *domestic-* followed either by *-o* or *-a*, and after that sequence, an optional *-s*. Some other examples comprise broader possibilities, such as the regular expression *abraz\w**, which matches the string *abraz-* followed by any repetitions (*) of any alphanumeric character *(\w*), allowing for the retrieval from the corpus of the whole verbal conjugation of *abrazar* (*to hug*), the noun *abrazo(s*) (*hug/s*), or, in general, any word built on the stem *abraz-*.

Let us briefly examine the most relevant categories identified by the software in the Spanish and English corpora. As shown in Table 6, the top 5 Spanish categories that reflect the semantic content of the items comprised in them relate to topics labelled as *social-analytic* (21.3%), a very broad category which includes terms[15] like *absolución* (*acquittal*), *abogado* (*solicitor/lawyer*) but also *cacao* (*cocoa*) or *mariposa* (*butterfly*); organizations (17.84%), exemplified by *tribunal supremo* (*supreme court*), *ONG* (*NGO*) or *PP/PSOE* (major political parties in Spain) and *locations* (6.79%), for instance, country names, cities or more specific places.

Table 6. Top 5 categories identified by UMUTextStats in the Spanish corpus

	mean
lexical-social-analytic	21.30%
lexical-organizations	17.84%
lexical-locations	6.79%
lexical-persons	6.79%
lexical-social-relativity-space	5.64%

Table 7 reflects the top 5 categories resulting from the automatic processing of the English corpus. Although the proportion of items in each category is considerably lower than the data displayed above, there is a coincidence between the top two categories, *organizations* and *social-analytic,* although, in this case, *organizations* ranks first in English. As regards the actual percentage of items comprised in each category, *organizations* represents 10.49% of the types found in the corpus (with words such as *court, conservatives* or *labour*), followed by *social-analytic*, which covers 3.27% of the types (varied terms as *sentence, trial, loneliness* or *prostitute* belong in this category), and *lexical-social-relativity-movement*, ranking third with 2.68% of the types found in the corpus (*approach, exit,* or *flee* are included within this thematic group).

As illustrated by the examples provided, only two of these categories partially coincide with the ones defined in Section 4.2.1, namely, *movement* and *locations*, which might be paired with *territory*. However, if the major purpose of classifying the lexicon in a text collection was to try and find out what major topics a legal corpus revolves around, such broad categories as *social-analytic*, although they may reveal some interesting themes in connection with immigration like *prostitution*, are far too inclusive to be able to actually signal specific thematic areas for further analysis.

Table 7. Top 5 categories identified by UMUTextStats in the English corpus

	mean
lexical-organizations	10.49%
lexical-social-analytic	3.27%
lexical-social-relativity-movement	2.68%
lexical-social-cognitive-insight	2.38%
psycholinguistic-processes-positive	1.82%

In sum, a software like *UMUTextStats* offers the possibility of determining the proportion of terms/lexical items falling into each of the morphosemantic and psycholinguistic categories defined in it, which range from words containing different types of affixes, to functional and lexical word classes or words referring to persons, locations, time, space or movement, amongst other. Yet, it does not extract the specialized terms in a corpus and then classify them according to their features, but rather determines the percentage of types in a text collection which fall into each of these categories with respect to the entire type count. Thus, although it does provide a much broader characterization of the lexicon (performed in a fully automatic manner) than the one presented in the previous section, it does not facilitate the actual examination of the items in each category, as it is solely focused on the quantification of such items, rather than on their extraction or their context of usage. Moreover, some of the categories included in it are far too broad to actually point at specific themes or topics susceptible of further analysis.

From a quantitative perspective, the fact that a tool like *UMUTextStats* manages to obtain the percentage of lexical items that fall into each of these categories without considering other parameters such as distribution, might push to the top of the rank some thematic categories which may not be representative of the corpus in its entirety, but rather of a set of texts where certain words are used recurrently. On the contrary, Drouin's (2003) and Scott's (2008) methods (this is not so for Chung's) pinpoint those lexical elements whose statistical relevance make them stand out within a corpus as a whole, deeming distribution a fundamental parameter to determine their position within the term ranking and thus potentially pointing at their degree of representativeness and their thematic relevance.

5. Conclusion

This chapter has sought to raise awareness on the need to apply ATR Methods to the lexical profiling of legal texts. For that purpose, three of these methods (Drouin 2003; Scott 2008a; Chung 2003) were implemented on two corpora of Spanish and English judicial decisions to measure their degree of reliability in automatically identifying legal terms. Two of them (Drouin's *TermoStat,* 2003 and Scott's *Keywords*, 2008a) allow the user to process corpora by simply uploading a specialized text collection to the system (Drouin's method) or rather processing it with the software tool included in a software package (Scott's method). The major difference between these two methods as regards implementation lies in the fact that, on the one hand, Drouin's software is freely available online (Scott's requires a license) and, on the other hand, it does not involve the use of general language corpora on the part of the user, as the software already includes some in several languages. Concerning the degree of expertise implied in managing both software packages, it is Drouin's method which appears to be more straightforward and easier to manage by the user, who can process a corpus quite intuitively without requiring any further assistance.

Firstly, The application of Chung's method was more complex since it required the manual implementation of the algorithm proposed by the author, an elaborate task that was facilitated greatly by using an Excel spreadsheet. Even so, the process was time-consuming because it forces the user to be relatively proficient in managing this type of software (it requires the use of complex formulas to search the results and then determine a term's frequency ratio).

Secondly, the degree of efficiency achieved by each of these methods was calculated after obtaining the CT lists and then validating them by determining the percentage of TTs contained amongst the top 500 CT extracted. The results were similar across languages, although slightly higher in Spanish than in English. *TermoStat* (Drouin 2003), *Keywords* (Scott 2008a) and Chung's ratio method (2003) reached 66.8%, 56.6% and 47.4% precision respectively in Spanish, standing, on average, 5 points above the results obtained in English.

Thirdly, the 500 terms identified by the three methods were classified into five thematic categories, *legal terms* being the most populated one and containing 76.07% and 90.98% items in English and Spanish respectively, as was to be expected. These figures clearly indicate that the terms retrieved by these ATR methods, in spite of them being more efficient in Spanish, are better distributed into thematic areas in English. However, Chung's method failed to detect different topics amongst the terms it identified, as it pushed to the top of the term list highly specialized legal terms almost in their entirety in Spanish. The other four categories, namely, *territory, evaluative items, family* and *crime* distributed themselves unevenly across these two languages, finding *territory* as the second most populated group followed by *evaluative items*. Out of the three ATR methods assessed, *TermoStat* (Drouin 2003) excelled the other two as regards its ability to embrace a larger proportion of terms within each thematic category, as shown in Table 5, where these four areas contain a more balanced proportion of terms both in Spanish and in English.

Finally, in order to reply to the question posed in the title *Do ATR methods provide a shorter path to lexical profiling?*, both text collections were also processed with *UMUTextStats* (García-Díaz et al. 2018; García-Díaz, Cánovas-García, and Valencia-García 2020), an unsupervised text classification tool which facilitates the automatic analysis of corpora for the classification of their lexicon into morphosemantic categories, which are represented in relation to the proportion of lexical items falling into each of these categories with respect to the entire type list. The process of implementation of this procedure was certainly faster and easier than the one described in Section 4.2.1., yet, the software could solely point at the most relevant themes in the corpus based on the amount of elements comprised in each thematic category, regardless of their distribution throughout the corpus or their salience with respect to other non-specialized texts collections. This type of tools might be excellent for automatic text classification or authorship attribution, as they work fully automatically and do not require any supervision, but their application to discourse studies based on thematic categorization might be limited.

Conversely, the implementation of ATR methods facilitated the identification of the most relevant themes in both corpora after creating *ad* *hoc* categories to classify the top 500 terms extracted and comparing them. Although the thematic classification took longer, given the fact that *UMUTextStats* does not produce any term lists and does not give access to their context of usage (apart from it not focusing on such parameters as distribution), it is recommendable to resort to ATR methods as a point of departure (particularly Drouin's (2003) for the different reasons stated above) for the lexical profiling of legal texts, as it seems to be the shortest path to do so in a more reliable manner, particularly when the major aim is studying legal discourse, only requiring manual work for the thematic categorization phase.

To conclude, as regards future research, this proposal presents a working methodology which may allow for a deeper and more comprehensive understanding of legal texts. As authors acknowledge, the literature devoted to the assessment of ATR methods is scarce, even more so within the legal field, thus, using this proposal as reference might facilitate considerably the scrutiny of other corpora by firstly identifying the terms in them and then moving onto the definition of the major topics they revolve around. In fact, this methodology might be applicable to the study of other public legal genres such as legislative or administrative texts which may relate to the topic of immigration and could be compared to judicial decisions such as the ones at hand, in search of different perspectives from which such a complex phenomenon could be depicted.

References

Ahmad, Khurshid, Andrea Davies, Heather Fulford, and Margaret Rogers. 1994. "What is a Term? The Semi-automatic Extraction of Terms from Text." In *Translation Studies: An Interdiscipline,* edited by Mary Snell-Hornby, Franz Pöchhacker and Klaus Kaindl, 267-278. Amsterdam: John Benjamins.

Alcaraz Varó, Enrique. 1994. *El inglés jurídico: textos y documentos*. Madrid: Ariel Derecho.

Ananiadou, Sophia. 1988. *A Methodology for Automatic Term Recognition.* PhD Thesis, University of Manchester Institute of Science and Technology: United Kingdom.

Ananiadou, Sophia. 1994. "A Methodology for Automatic Term Recognition." In *COLING. Proceedings of the 15th International Conference on Computational Linguistics*, 1034-1038. https://doi.org/10.3115/991250.991317

Anthony, Laurence. 2020. AntConc (Version 3.5.9) [Computer Software]. Tokyo: Waseda University. https://www.laurenceanthony.net/software

Arora, Chetan, Mehrdad Sabetzadeh, Lionel Briand, and Frank Zimmer. 2016. "Automated Extraction and Clustering of Requirements

Glossary Terms[J].” *IEEE Transactions on Software Engineering* 43(10):918-945.

Aronson, Alan R. and François-Michel Lang. 2010. “An Overview of MetaMap: Historical Perspective and Recent Advances.” *Journal of American Medical Informatics Association* 17(3):229-236.

Astrakhantsev, Nikita. 2018. “ATR4S: Toolkit with State-of-the-art Automatic Terms Recognition Methods in Scala.” *Language Resources and Evaluation* 52:853–872.

Marín, María José. 2014. “Evaluation of Five Single-Word Term Recognition Methods on a Legal Corpus.” *Corpora* 9(1):83–107.

Marín, María José. 2016. “Measuring the Degree of Specialisation of Sub-Technical Legal Terms through Corpus Comparison: a Domain-Independent Method.” *Terminology* 22(1):80-102.

Barrón-Cedeño, Alberto, Gerardo E. Sierra, Patrick Drouin, and Sophia Ananiadou. 2009. “An Improved Automatic Term Recognition Method for Spanish.” In *International Conference on Intelligent Text Processing and Computational Linguistics,* edited by Alexander Gelbukh, 125-136. Berlin: Springer.

Bernier-Colborne, Gabriel. 2012. “Defining a Gold Standard for the Evaluation of Term Extractors.” In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12*), edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis, 15-18. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/index.html

Biel, Łucja and Jan Engberg. 2013. “Research Models and Methods in Legal Translation.” *Linguistica Antverpiensia* 12:1–11.

Bisceglia, Bruno, Rita Calabrese, and Ljubica Leone. 2014. “Combining Critical Discourse Analysis and NLP Tools in Investigations of Religious Prose.” *LRE-REL2. Proceedings of the 2nd Workshop on Language Resources and Evaluation for Religious Texts. 31 May 2014, Reykjavik, Iceland*, edited by Claire Brierley, Majdi Sawalha and Eric Atwell, 24-29. http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014Workshop-LRE-Rel2%20Proceedings.pdf

Borja Albí, Anabel. 2000. *El texto jurídico en inglés y su traducción*. Barcelona: Ariel.

Breeze, Ruth. 2015. “Teaching the Vocabulary of Legal Documents: A Corpus-driven Approach.” *ESP Today* 3(1):44–63.

Bourigault, Didier. 1992. “Surface Grammatical Analysis for the Extraction of Terminological Noun Phrases.” In *COLING 1992 - Volume 3: The 14th International Conference on Computational Linguistics,* 977-981. https://aclanthology.org/C92-3150.pdf

Cabré Castellví, Maria Teresa. 1999. *Terminology: Theory, Methods and Applications*. Amsterdam: John Benjamins.

Cabré Castellví, Maria Teresa, Rosa Estopà Bagot, and Jordi Vivaldi Palatresi. 2001. “Automatic Term Detection: A Review of Current Systems.” In *Recent Advances in Computational Terminology 2*, edited by Dider Bourigault, Christian Jacquemin and Marie-Claude L’Homme, 53-87. Amsterdam: John Benjamins.

Chung, Teresa Mihwa. 2003. “A Corpus Comparison Approach for Terminology Extraction.” *Terminology* 9(2):221-246.

Church, Kenneth and Patrick Hanks. 1990. “Word Association Norms, Mutual Information, and Lexicography.” *Computational Linguistics* 16(1):22-29.

Coxhead, Averil. 2000. “A New Academic Word List.” *TESOL Quarterly* 34(2):213-238.

Dagan, Ido and Kenneth Church. 1994. “Termight: Identifying and Translating Technical Terminology.” In *Fourth Conference on Applied Natural Language Processing*, 34-40. Association for Computational Linguistics. https://doi.org/10.3115/974358.974367

Daille, Béatrice. 1996. “Study and Implementation of Combined Techniques for Automatic Extraction of Terminology.” In *The Balancing Act: Combining Symbolic And Statistical Approaches To Language,* edited by Judith L. Klavans and Philip Resnik, 49-66. Cambridge, MA: MIT Press.

David, Sophie and Pierre Plante. 1990. *Termino* 1.0. Research Report of Centre d'Analyse de Textes par Ordinateur. Montréal: Université du Québec.

Drouin, Patrick. 2003. "Term Extraction Using Non-technical Corpora as a Point of Leverage." *Terminology* 9(1):99-115.

Dunning, Ted E. 1993. "Accurate Methods for the Statistics of Surprise and Coincidence." *Computational Linguistics* 19(1):61-74.

Fahmi, Ismail, Gosse Bouma, and Lonneke Van Der Plas. 2007. "Improving Statistical method using known terms for automatic term extraction." (conference talk). Conference: *Computational Linguistics in the Netherlands* (*CLIN 17*), November 2007 (unpublished).

Frantzi, Katerina T. and Sophia Ananiadou. 1996. "Extracting Nested Collocations." In *COLING 1996 - Volume 1: The 16th International Conference on Computational Linguistics*, 41-46. USA: Association for Computational Linguistics. https://aclanthology.org/C96-1009.pdf

Frantzi, Katerina T. and Sophia Ananiadou. 2000. "Automatic Recognition of Multi-Word Terms: the C-value/NC-value Method." *International Journal on Digital Libraries* 3:115-130.

Gabrielatos, Costas and Anna Marchi. 2011. "*Keyness: Matching Metrics to Definitions" (conference talk). Conference: Corpus Linguistics in the South: Theoretical-methodological challenges in corpus approaches to discourse studies - and some ways of addressing* *them,* 5th November, Portsmouth (unpublished). http://eprints.lancs.ac.uk/51449/4/Gabrielatos_Marchi_Keyness.pdf.

García-Díaz, José Antonio, Mar Cánovas-García, and Rafael Valencia-García. 2020. "Ontology-driven Aspect-based Sentiment Analysis Classification: An Infodemiological Case Study Regarding Infectious Diseases in Latin America." *Future Generation Computer Systems* 112:641-657. doi:10.1016/j.future.2020.06.019

García-Díaz, José Antonio, María Pilar Salas-Zárate, María Luisa Hernández-Alcaraz, Rafael Valencia-García, and Juan Miguel Gómez-Berbís. 2018. "Machine Learning Based Sentiment Analysis on Spanish Financial Tweets." In *Trends and Advances in Information Systems and Technologies (WorldCIST'18 2018). Advances in Intelligent Systems and Computing,* Vol. 745, edited by Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis and Sandra Costanzo, 305-311. Springer: Cham.

Heylen, Kris and Dirk De Hertog. 2015. "Automatic Term Extraction." In *Handbook of Terminology,* edited by Hendrik Kockaert and Frieda Steurs, 203-221. Amsterdam: John Benjamins.

Jacquemin, Christian. 2001. *Spotting and Discovering Terms through NLP.* Massachusetts: MIT Press.

Jumaquio-Ardales, Alona, Nathaniel Oco, and Rowell Madula. 2017. "Click-analysis of a Lesbian Online Community in Facebook Using the Critical Discourse Analysis and Natural Language Processing." *Humanities Diliman: A Philippine Journal of Humanities* 14(1):46-68.

Jurafsky, Daniel and James H. Martin. 2019. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River: Els autors.

Justeson, John S. and Slava M. Katz. 1995. "Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text." *Natural Language Engineering* 1(1):9-27.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1:7-36.

Kit, Chunyu and Xiaoyue Liu. 2008. "Measuring Mono-word Termhood by Rank Difference via Corpus Comparison." *Terminology* 14(2):204-229.

Lemay, Chantal, Marie-Claude L'Homme, and Patrick Drouin. 2005. "Two Methods for Extracting 'Specific' Single-word Terms from Specialised Corpora: Experimentation and Evaluation." *International Journal of Corpus Linguistics* 10(2):227-255.

Loginova, Elizaveta, Anita Gojun, Helena Blancafort, Marie Guégan, Tatiana Gornostay, and Ulrich Heid. 2012. "Reference Lists for the Evaluation of Term Extraction Tools." In *Proceedings of the 10th Terminology and Knowledge Engineering Conference: New Frontiers in the Constructive Symbiosis of Terminology and Knowledge Engineering (TKE 2012)*, edited by Guadalupe Aguado de Cea, Mari Carmen Suárez-Figueroa, Raúl García-Castro and Elena Montiel-Ponsoda. https://www.researchgate.net/publication/236686487_Reference_Lists_for_the_Evaluation_of_Term_Extraction_Tools

Maynard, Diana and Sophia Ananiadou. 2000. "TRUCKS: A model for Automatic Multi-word Term Recognition." *Journal of Natural Language Processing* 8(1):101–125.

Meijer, Kevin, Flavius Frasincar, and Frederik Hogenboom. 2014. "A Semantic Approach for Extracting Domain Taxonomies from Text." *Decision Support Systems* 62:78-93.

Mellinkoff, David. 1963. *The Language of the Law*. Boston: Little, Brown & Co.

Mondary, Thibault, Adeline Nazarenko, Haïfa Zargayouna, and Sabine Barreaux. 2012. "The Quaero Evaluation Initiative on Term Extraction." In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis, 663-669. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/index.html

Nazar, Rogelio and María Teresa Cabré. 2012. "Supervised Learning Algorithms Applied to Terminology Extraction." In *Proceedings of* *the 10th Terminology and Knowledge Engineering Conference: New Frontiers in the Constructive Symbiosis of Terminology and Knowledge Engineering (TKE 2012)*, edited by Guadalupe Aguado de Cea, Mari Carmen Suárez-Figueroa, Raúl García-Castro and Elena Montiel-Ponsoda, 209-217. Madrid: Universidad Politécnica de Madrid.

Pazienza, Maria Teresa, Marco Pennacchiotti, and Fabio Massimo Zanzotto. 2005. "Terminology Extraction: An Analysis of Linguistic and Statistical Approaches." In *Knowledge Mining. Studies in Fuzziness and Soft Computing,* Vol. 185, edited by Spiros Sirmakessis*,* 255-279. Berlin: Springer.

Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam: John Benjamins.

Pontrandolfo, Gianluca and Stanisław Goźdź-Roszkowski. 2014. "Exploring the Local Grammar of Evaluation: The Case of Adjectival Patterns in American and Italian Judicial Discourse." *Research in Language* 12(1):71-91.

Rayson, Paul, and Roger Garside. 2000. "Comparing Corpora Using Frequency Profiling." In *WCC '00: Proceedings of the Workshop on Comparing Corpora,* Vol. 9, 1-6. https://doi.org/10.3115/1117729.1117730

Reese, Samuel, Gemma Boleda, Montse Cuadros, Lluís Padró, and German Rigau. 2010. "Wikicorpus: A Word-Sense Disambiguated Multilingual Wikipedia Corpus." In *Proceedings of 7th Language*

Resources and Evaluation Conference (LREC'10), edited by Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner and Daniel Tapias, 1418-1421. European Language Resources Association (ELRA).

Rema Rossini, Favretti, Fabio Tamburini, and Enrico Martelli. 2001. “Words from the Bononia Legal Corpus.” *International Journal of Corpus Linguistics* 6(3):13-34.

Schmid, Helmut. 1999. “Improvements in Part-of-Speech Tagging with an Application to German.” In *Natural Language Processing Using Very Large Corpora*, edited by Susan Armstrong, Kenneth Church, Pierre Isabelle, Sandra Manzi, Evelyne Tzoukermann, and David Yarowsky, 13-25. Springer.

Scott, Mike. 2008a. *WordSmith Tools*, Version 5. Liverpool: Lexical Analysis Software.

Scott, Mike. 2008b. *WordSmith Tools Help*. Stroud: Lexical Analysis Software.

Shang, Jingbo, Jialu Liu, Meng Jiang, Xiang Ren, Clare R. Voss, and Jiawei Han. 2018. “Automated Phrase Mining from Massive Text Corpora.” *IEEE Transactions on Knowledge and Data Engineering* 30(10):1825-1837.

Spasic, Irena, Sophia Ananiadou, John McNaught, and Anand Kumar. 2005. “Text Mining and Ontologies in Biomedicine: Making Sense of Raw Text.” *Brief Bioinform* 6(3):239-251.

Tiersma, Peter. 1999. *Legal Language*. Chicago: The University of Chicago Press.

Vivaldi, Jorge, Luis Adrián Cabrera-Diego, Gerardo Sierra, and María Pozzi. 2012. "Using Wikipedia to Validate the Terminology Found in a Corpus of Basic Textbooks." In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk and Stelios Piperidis, 3820–3827. European Language Resources Association (ELRA). Available at: http://www.lrec-conf.org/proceedings/lrec2012/index.html

[1] For more information on the LACELL research group see: https://curie.um.es/curie/catalogo-ficha.du?seof_codigo=1&perf_codigo=4&cods=E020*02

[2] Available at: https://www.cs.upc.edu/~nlp/wikicorpus/

[3] Available at: http://termostat.ling.umontreal.ca/index.php?lang=fr_CA

[4] See p. 18 on filtering the CTs obtained with Chung's method.

[5] In this case, the result was multiplied by 10,000 to make the figures more manageable and avoid an excessive number of zeros and decimals.

[6] See Section 3.1.3. for details on normalisation.

[7] http://www.poderjudicial.es/search/indexAN.jsp

[8] http://www.bailii.org/

[9] For the compilation of both legal corpora, the query terms related to the topic of migration.

[10] For more details on the calculation method see:

https://ucrel.lancs.ac.uk/people/paul/publications/rg_acl2000.pdf and

https://ucrel.lancs.ac.uk/llwizard.html (UCREL's log-likelihood site)

[11] K= Keyness score

[12] S=Specificity score

[13] FR= Frequency Ratio

[14] Terms like these were left out of the validated lists applying quantitative and also qualitative criteria. In fact, they are very common and frequent in the general context, yet, they were not perceived strictly as legal terms by the referees.

[15] These examples have been extracted from the dictionary library included in the software tool.