



ASp

la revue du GERAS

66 | 2014

**Intersections - l'anglais de spécialité, creuset
multidomaine**

Researching legal terminology: a corpus-based proposal for the analysis of sub-technical legal terms

María José Marín and Camino Rea



Electronic version

URL: <http://journals.openedition.org/asp/4572>

DOI: 10.4000/asp.4572

ISSN: 2108-6354

Publisher

Groupe d'étude et de recherche en anglais de spécialité

Printed version

Date of publication: 1 November 2014

Number of pages: 61-82

ISSN: 1246-8185

Electronic reference

María José Marín and Camino Rea, « Researching legal terminology: a corpus-based proposal for the analysis of sub-technical legal terms », *ASp* [Online], 66 | 2014, Online since 01 November 2015, connection on 03 November 2020. URL : <http://journals.openedition.org/asp/4572> ; DOI : <https://doi.org/10.4000/asp.4572>

This text was automatically generated on 3 November 2020.

Tous droits réservés

Researching legal terminology: a corpus-based proposal for the analysis of sub-technical legal terms

María José Marín and Camino Rea

1. Introduction

- 1 As often agreed by scholars (Mellinkoff 1963; Alcaraz Varó 1994; Tiersma 1999; Borja Albí 2000; Orts Llopis 2006), legal English presents serious difficulties not only to the foreign learner but also to non-specialised natives because of its inaccessibility and pomposity. Mellinkoff's statement about the nature of legal language is self-explanatory in this respect: "The language of the law has a strong tendency to be wordy, unclear, pompous and dull" (1963: 24). In fact, the use of convoluted syntactic structures, Latin and Old French phrases, the continuous appearance of archaic terms or lexical repetition, amongst other features, certainly hinder the understanding of legal texts such as law reports or statutes on which the British legal system is based.
- 2 On a lexical level, the presence of vocabulary shared by the general and specific fields appears to be a particularly relevant trait of legal English terminology, which also contributes to its obscurity due to its often polysemic character. These are the so-called *sub-technical* terms which, according to various authors (Cowan 1974; Trimble 1985; Baker 1988; Farrell 1990; Flowerdew 2001; Chung & Nation 2003; Wang & Nation 2004), either denote a legal concept in both general and specialised usage (words such as "judge"; "solicitor"; "guilty"; "court"), or activate a specialised meaning in the legal field, being frequently used as general words in everyday language ("trial"; "conviction"; "sentence"; "relief").
- 3 Nevertheless, to the best of our knowledge and in spite of the relevance of sub-technical vocabulary in specialised languages, no method has been designed to date to try and quantify the manner in which these terms acquire a new specialised meaning combining qualitative and quantitative techniques. Consequently, this article suggests

a method for the description of the process of specialisation undergone by the sub-technical term “charge”, adopting both a semantic and a corpus-based perspective. With this purpose, Cantos and Sánchez’s (2001) Lexical Constellation (LC) model was applied to its analysis, employing both a qualitative and a quantitative approach, the latter being corpus-based.

- 4 The term “charge” is presented as an example of the methodology suggested herein, which attempts to illustrate the path followed by sub-technical terms towards specialisation, as already stated. Still much remains to be done in this respect and this method of analysis should be implemented on a larger amount of terms for the results to be conclusive. Our proposal thus constitutes a suggestion for the better understanding of the nature of these terms, which should be explored in greater depth in the future applying this procedure.
- 5 Two language corpora were used to exemplify this process in both the general and the legal fields. The two corpora used as reference are *BLaRC* (*British Law Report Corpus*), an 8.85 million-word legal corpus of law reports from the United Kingdom (see Marín & Rea [2012] for an in-depth explanation of the process of compilation and main features), and *LACELL* (*Lingüística Aplicada Computacional, Enseñanza de Lenguas y Lexicografía*), a general English corpus of 21 million words owned and compiled by the LACELL research group at the University of Murcia.
- 6 Section 2 of this article is devoted to the description of these corpora followed by section 3, where the term “sub-technical” is discussed through a literature review. Section 4 includes the description of the methodology applied in this study of sub-technical vocabulary adopting a qualitative and a quantitative perspective. Finally, the last section presents the conclusions drawn from the analysis of the data provided.

2. Description of the two corpora used for the study: *BLaRC* and *LACELL*

- 7 In order for the semantic analysis of the sub-technical term “charge” to be based on actual data, two language corpora were used as source. Their features are described in the following sections.

2.1. *BLaRC*: A legal corpus of British law reports

- 8 *BLaRC*, the *British Law Report Corpus*, was designed and compiled with the purpose of identifying and exploring the nature of legal terminology. This process was carried out abiding by the standards of Corpus Linguistics established by McEnery and Wilson (2001), Sinclair (2005) and McEnery and Xiao (2006) for general corpora, and Pearson (1998), Vargas (2005) and Rea (2010) for specialised ones. It is an 8.85 million-word legal corpus of law reports. Law reports could be defined as written collections of judicial decisions made at the UK courts and tribunals.
- 9 The reasons to focus on this legal genre to study the linguistic properties of its terminology are varied. To begin with, the UK law system belongs to the realm of common law. Unlike civil or continental law systems, such as those found in most European countries, common law is not codified, that is, it could be stated that, in a way, it is judge-made. Judges interpret the law (found in statutes) and apply the

existing case law related to a given case as long as it is similar to that case in its essence. They determine whether a decision made at a higher court should be applicable to a case and how it should be applied. In addition, they must also justify the rationale behind their decision. This is the so-called *ratio decidendi*, which is binding on courts of lower jurisdiction. According to the doctrine of *stare decisis*, which common law revolves around, the judicial decisions (called precedents) made at higher courts such as the Supreme Court or the High Court of England and Wales must be followed by lower courts owing to their binding character, setting a legal precedent within the system, hence their relevance. This is why they are published every year by institutions such as the Incorporated Law Report Council of England and Wales (ILRC) and must be consulted and cited by solicitors, barristers or judges. Therefore, they act as the solid ground which legal practitioners build their arguments on.

- 10 Further to their significance within the UK legal system, judicial decisions might be deemed valuable sources of information about the legal lexicon due to the fact that they not only cover all the branches of law, but might also present full embedded sections of other public and private law genres, thus displaying great lexical density and variety.
- 11 *BLaRC* is a synchronic, monolingual and specialised collection of 1,228 judicial decisions from the UK courts and tribunals system issued between 2008 and 2010 in raw text format. They were downloaded from the database website, edited to eliminate problematic characters and stored as .txt files. Several factors conditioned the decisions made regarding corpus structure: first, the hierarchy of the courts and tribunals the texts were obtained from and secondly, the way that legal terminology varies depending on the system where it is used. This is so because of the laws and regulations that organise the countries which are part of the UK. The judicial systems of Northern Ireland, Scotland, England and Wales do not solely depend on UK institutions, but rather have their own autonomous systems and structure. Except for the Supreme Court (in general terms) and the UK Tribunal Service (with some exceptions), each country is fully independent as regards its judicial system.¹ It should be highlighted that England and Wales have a common system, whereas Northern Ireland and Scotland have their own network of courts and tribunals.
- 12 Therefore, *BLaRC* was structured into five main sections depending on the jurisdictions of the British judicial system: a) Commonwealth countries; b) United Kingdom; c) England and Wales; d) Northern Ireland; e) Scotland. Additionally, each corpus section was divided into different sub-sections coinciding with the hierarchical structure of the courts and tribunals comprised in it. By maintaining this structure, the texts were grouped according to the field of law they belonged to (but for the Supreme Court, most courts and tribunals are organised according the branch of law they pertain to, i.e., criminal law, family law, commercial law, intellectual law, etc.), hence the similarity of their lexicon. Thus, comparing results by studying the sections separately could be easier and respond to a thematic criterion which is fundamental as far as the identification and study of the specialised vocabulary of this legal genre is concerned.
- 13 Owing to the scarceness of legal corpora available and the usefulness of the data provided by them as support for the legal English class, *BLaRC* has recently been made publicly available on Cobb's website *The Compleat Lexical Tutor*,² where it can be selected amongst a list of corpora for legal term queries, providing the concordances and

extended contexts associated with the terms selected. It will also be accessible shortly via Kilgarriff's *Sketch Engine*.³

2.2. LACELL: A general English corpus

- 14 *LACELL* is a 21 million-word general English corpus which comprises mainly texts from the 1990s. It is a balanced synchronic corpus of general English including 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 conversations at different social levels and registers, debates and group discussions, TV and radio recordings, phone conversations, everyday life situations, classroom talk, etc. Nevertheless, as it happens with the *BNC*, approximately 10% of the texts in *LACELL* are transcriptions of oral language, the rest are all written texts.
- 15 Its geographical scope ranges from the USA to Canada, the UK and Ireland. Precisely because of that and the fact that *BLaRC* is a British English corpus, those texts not coming from the UK were removed to avoid skewness⁴ in the results, reducing the original size to 14.9 million words.

3. Sub-technical vocabulary

- 16 As stated in the introduction, the use of vocabulary shared by the general and specialised fields in legal English should be regarded as a relevant feature of this ESP variety. However, there is little agreement amongst specialists on how to define these terms. Generally speaking, authors tend to favour the use of the term “sub-technical” basically defined as vocabulary common to both the general and the specialised fields or amongst scientific disciplines. Only few of them (Cowan 1974; Flowerdew 2001) employ the label “semi-technical” either to refer to the same concept or as a synonym. In addition, most scholars underline the importance of these terms in ESP instruction due to the fact that they might become an obstacle in the learners’ acquisition of the vocabulary in any scientific field.
- 17 Within the group of specialists who have addressed the issue from a didactic perspective we find the work by Cowan (1974) – who coined the term “sub-technical” – Baker (1988) and Flowerdew (2001). They agree on the confusing and obscure character of sub-technical terms, putting special emphasis on the relevance they should be given in the ESP curriculum precisely because of their semantic ambiguity. Flowerdew defined them as “words in general usage [...] which have a special meaning within the technical area” (2001: 82).
- 18 Sub-technical terms have also been explored from a quantitative perspective. Farrell (1990) came to the conclusion that distribution plays an essential role in the identification of sub-technical terms. According to him, these terms are well distributed across a corpus and display high frequency counts. Trimble also focused on frequency but, as Flowerdew did, added a qualitative viewpoint to their definition noticing that “those words [...] in technical contexts, take on extended meanings” (1985: 129).
- 19 Finally, only Chung and Nation (2003) and Wang and Nation (2004) managed to delimit the semantic features of technical and sub-technical vocabulary although they did not

label shared vocabulary as “sub-technical”. In their analysis of Coxhead’s (2000) *Academic Word List* (AWL), Wang and Nation attempted to distinguish the members of the same word family from homographs, which are identical in form but utterly different in meaning. In order to do so, they produced a “semantic relatedness scale” (2004: 291) by means of which they could distinguish whether the different senses of a word were related to each other in a way that it could be regarded as polysemic or if, on the contrary, they were completely unrelated, thus being a clear instance of a homograph. They established six semantic levels which the different meanings of a word may fit into, in relation to their distance from the base meaning. Sub-technical terms could be identified with those members of a word family whose meaning differs from the original one but is not completely unrelated to it.

- 20 As a consequence of the above, our analysis of “charge”, a sub-technical legal term, will try to combine both qualitative and quantitative criteria. Most of the cited authors define sub-technical terms as those which are shared by the general and the specialised fields. This can be quantified focusing on their frequency in both contexts by simply processing both corpora and studying the data. Nevertheless, specialists also insist on the obscurity of these terms owing to the fact that they often activate a specialised meaning in the technical context. Therefore, could statistical information somehow be related to a semantic analysis of sub-technical terms? Our proposal attempts to integrate both perspectives by grounding the semantic analysis of the term “charge” on corpus data. In order to illustrate its process of specialisation, Cantos and Sánchez’s (2001) LC model was implemented to try and account for this process. The following section first focuses on the presence of sub-technical vocabulary in our corpus before proceeding with the analysis itself.

3.1. Relevance of sub-technical vocabulary in BLaRC, our legal corpus

- 21 A comparison was carried out between the term list automatically extracted from *BLaRC* and the general English lists mentioned above with the purpose of demonstrating that sub-technical language can indeed be observed in legal language, as several authors have claimed. This term list was obtained by applying two Automatic Term Recognition (ATR) methods to *BLaRC*, Drouin’s (2003) *TermoStat*⁵ and Nazar and Cabré’s (2012) *Terminus*.⁶ Both authors offer an automatic method to extract specialised terms from a corpus (which must be uploaded to the system) online.
- 22 The output lists of candidate terms were evaluated against a glossary of legal terms comprising 10,088 items, used as gold standard⁷ for comparison. This glossary was compiled by merging together four different online legal glossaries;⁸ its compilation and validation process and its features are described in greater detail in Marín (2014). This comparison resulted into a single-word term inventory of 2,848 items.
- 23 Tables 1 and 2 below include the output lists produced by both *Terminus 2.0* and *TermoStat* before validation (not all the words appearing in each table are true terms),⁹ which contain information such as the lemma (Drouin labels it as “grouping variant”), POS category, frequency and level of specialisation assigned to each word type by both ATR methods. The resulting list was not validated in order to illustrate what both Drouin’s and Nazar and Cabré’s software can do without human intervention in the process. Nonetheless, both lists were compared with the legal glossary mentioned

above and also validated manually to confirm that the 2,848 included items were true legal terms.

Table 1. Output list generated by Nazar and Cabré's *Terminus 2.0*

Rank	Form	Lemma	Frequency	Weighting
1	reasonable	reasonable	4490	397271.62
2	basis	basis	6221	299498.73
3	extent	extent	2341	271501.06
4	payment	payment	3045	243836.79
5	lawful	lawful	705	235189.56
6	witness	witness	1424	230170.33
7	word	word	1027	198149.73
8	facie	facie	333	191377.09
9	context	context	2867	146508.86
10	payable	payable	894	145321.51
11	causation	causation	285	135029.10
12	injunction	injunction	536	121506.21
13	complaint	complaint	1737	112844.92
14	obligation	obligation	2051	112659.45
15	infringement	infringement	377	101451.54
16	wording	wording	626	93573.26
17	presumption	presumption	521	89657.53
18	actual	actual	1030	89491.33

Table 2. Output list obtained with Drouin's *TermoStat*

Candidate (grouping variant)	Frequency	Score (specificity)	Variants	Pattern
section	9694	126.29	section; sections	Common Noun
v	6828	112.55	v	Common Noun

case	11465	111.79	case; cases	Common Noun
para	5973	108.63	para; paras	Common Noun
article	5686	97.39	article; articles	Common Noun
court	6387	88.65	court; courts	Common Noun
appeal	3993	80.30	appeal; appeals	Common Noun
appellant	3102	78.47	appellant; appellants	Common noun
not	22062	75.07	not	Adverb
law	5484	73.55	law; laws	Common Noun
judgment	2862	71.67	judgment; judgments	Common Noun
claim	3293	69.80	claim; claims	Common Noun
right	5795	67.98	right; rights	Common Noun
apply	3542	65.50	apply; applying; applied; applies	Verb

- 24 Subsequently, an Excel spreadsheet was used to compare this list of single-word legal terms with the most frequent 2,570 word families in West's (1953) *GSL* and Coxhead's (2000) *AWL*. It was found that 40.47% of these terms overlapped with both vocabulary inventories. The percentage was slightly higher, 45.41%, if compared with the *BNC* most common 3,000 words, which confirms scholars' intuition that the presence of sub-technical vocabulary in legal English texts is a highly relevant trait of this ESP variety. In fact, almost half of the terms in *BLaRC* are shared with the abovementioned general vocabulary lists in spite of the level of specialisation¹⁰ assigned to those overlapping words by each ATR method.
- 25 Tables 3 and 4 present a sample¹¹ of sub-technical legal terms that are found in both the *BLaRC* and the list of the most frequent 3,000 words of *BNC*. Both tables show the terms as lemmas, not word types, since both methods used to extract them lemmatise the corpus before extracting the terms from it. While table 3 presents a sample of the terms identified by *Terminus* (Nazar & Cabré 2012) which overlap with the *BNC* lists, table 4 comprises a sample of those which were extracted using Drouin's (2003) *TermoStat*. The third and fourth columns in both tables indicate the frequency of occurrence of each term in both the specialised (*BLaRC*) and the general corpus (*BNC*) after normalisation. When the size of two corpora is different, it becomes necessary to normalise the data so that it can be comparable. In this case, the raw frequency data provided by *Wordsmith* was divided by the number of millions of words in each corpus, 8.85 for *BLaRC* and 100 for the *BNC*. The figures shown in the third and fourth columns of tables 3 and 4 correspond with the results obtained after this process.

- 26 As it can be observed from the data provided below, in general, in spite of occurring frequently in both contexts, all the terms selected are more frequent in the specialised corpus. For instance, the frequency of a word such as “claim” is six times as high in *BLaRC* (1,139.77) as it is in *BNC* (188.28).

Table 3. Sub-technical terms extracted by *Terminus* (Nazar & Cabré 2012)

Term	Specificity (<i>Terminus</i>)	Level	Normalised frequency in <i>BLaRC</i>	Normalised frequency in <i>BNC</i>
argument	26106.56		467.9	121.25
claim	13890.29		1,139.77	188.28
argue	10130.73		193.67	141.96
act	5791.31		1930.28	226.57
commit	5359.17		148.58	67.83
appeal	3657.25		2145.87	100.2
criminal	2922.64		390.96	46.02
court	2503.81		2,824.85	299.76
complain	1806.70		88.13	41.35
damage	1638.90		354.35	85.52
allow	858.29		3,700	336.87
authority	795.11		1,039.54	312.31
abuse	545.32		123.16	34.28
charge	427.09		282.48	165.03
civil	385.87		180.33	86.6

Table 4. Sub-technical terms extracted by *TermoStat* (Drouin 2003)

Term	Specificity (<i>TermoStat</i>)	Level	Normalised frequency in <i>BLaRC</i>	Normalised frequency in <i>BNC</i>
case	89.24		3,841.8	635.77
decision	63.86		1,726.32	242.33
contract	31.34		563.95	156.72

consideration	29.54	423.61	78.08
action	17.48	467	268.94
compensation	16.07	169.49	31.91
defence	13.87	314.91	127.6
bind	13.07	169.49	65.17
debt	10.42	101.69	68.35

- 27 The concepts denoted by these terms (included in tables 3 and 4) can be divided into two groups. The first group comprises those terms which are often used in both general and legal English without changing their meaning. For example, the term “authority” can be found in sentences such as “[...] whenever a State exercises legislative, judicial or executive authority in a way that affects an individual's Convention right [...]” (extracted from *BLaRC*) or “[...] government authority should be located in the smallest and most local unit possible [...]” (*LACELL*), referring in both cases to a body or institution which can exercise its power. Similarly, the term “abuse”, meaning “harassment” or “physical maltreatment” occurs in both fields with identical reference, e.g., “[...] that sexual abuse of children remains pervasive in Afghanistan” (*BLaRC*) or “[...] although it doesn't report how these break down into sex abuse, physical abuse, or neglect [...]” (*LACELL*).
- 28 The second group comprises sub-technical terms which have a different meaning when in contact with the legal context. The term “claim” belongs in this group since it can denote a “call for something” (in general), e.g., “Iraq has an old political claim to Kuwait as a former province” (*LACELL*), or rather “the act of legally requesting something”, e.g., “The claimant should be entitled to the benefit of the ordinary rule that he should have the costs of the claim [...]” (*BLaRC*). Likewise, the word “act” might denote “an action” or “a deed” in a sentence such as “to sell the customer list to a third party is an unfair or deceptive act or practice [...]” (*LACELL*), whereas it refers to a law passed at the parliament in the legal field, for instance, “[...] the present document limits the competence of the Scottish parliament whereby an act of this parliament will not modify [...]” (*BLaRC*).

4. Methodology

- 29 Given the high frequency figures associated to sub-technical vocabulary in our legal corpus, as shown above, a novel approach is presented in this section for the analysis of “charge”, a polysemous legal term which could be classified as sub-technical for various reasons. On the one hand, it is included in West's (1953) list of the 2,000 most frequent word families of English and the *BNC* most common 3,000 words, meaning “weight” or “load” as a noun. On the other hand, when in contact with the legal context, it specialises into a highly representative term in our corpus (given its high frequency and text distribution), meaning “an accusation by a judge or competent authority”. Its

analysis was carried out through the implementation of Cantos and Sánchez's (2001) LC model, adopting both a qualitative and a quantitative perspective.

4.1. Cantos and Sánchez's (2001) Lexical Constellation model

- 30 Cantos and Sánchez suggest a new approach to lexical analysis which consists in studying “the way words socialise with other words, forming complex network-like structures or units” (2001: 200). These networks are hierarchically distributed displaying semantic dependencies amongst their constituents, that is, the collocate network generated by the nucleus and its collocates. Such dependencies are organised according to the level of these constituents within the constellation. In fact, the Lexical Constellations (LCs) work similarly to a star system where planets orbit around a central star (the node) which attracts them, being connected, in turn, to other star systems forming constellations, hence their name.
- 31 The LC model manages to overcome limitations such as the establishment of the optimum span to filter the number of collocates to be considered for analysis by setting the sentence as the limit for that span. This model provides information about the most significant collocate¹² of a given word and also about the hierarchical relationship between a word and the constituents of its constellation. Furthermore, it manages to represent those relationships in a visual and multi-dimensional way facilitating to a great extent the understanding of the dependencies existing amongst the elements in each LC.
- 32 As commonly agreed by researchers, the different meanings of a word are neither fully transparent nor clear-cut (Cruse 2004; Almela 2006; Kilgarrieff 2006; Sánchez *et al.* 2010). They could be interconnected forming a semantic network where the central meaning of a given word (often its most frequent interpretation or its oldest meaning) is added new features which stem from it. The LC model could thus be implemented to study the dependencies existing between the general and specialised meanings of sub-technical terms in a very clear and visual manner, showing that “semantic bonds and the configuration of conceptual associations [...] [are] multi-dimensional” (Sánchez *et al.* 2010: 142).

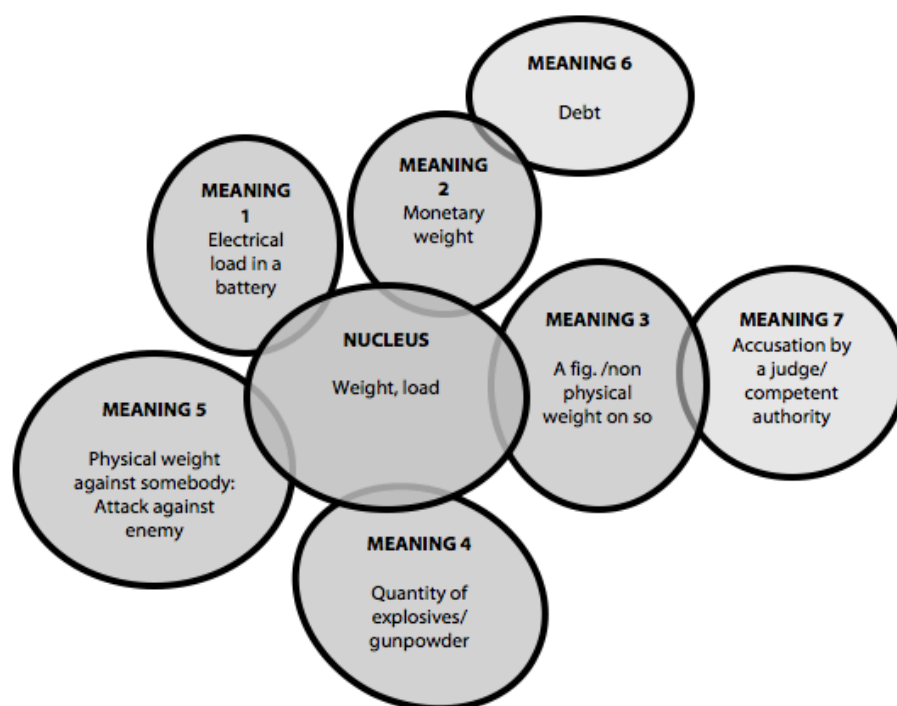
4.2. The qualitative approach

- 33 In order to ground the semantic analysis of “charge” on corpus data, which reflect actual language usage, the implementation of the LC model was carried out adopting both a qualitative and a quantitative perspective. Both processes took place in parallel so that, while we resorted to the different interpretations of the term and its origin and tried to establish the relationship and dependency among its various meanings, we simultaneously consulted corpus data to support the decisions that were made. Section 4.3. exemplifies the quantitative process in greater detail.
- 34 As regards the qualitative approach, several factors were taken into consideration: in the first place, the definitions of the term (as a noun) found in specialised and general dictionaries such as *New Oxford English Dictionary* (Pearsall 1998) and *Dahl's Law Dictionary* (Saint Dahl 1999), whose semantic features were analysed and refined by two specialists in corpus linguistics and legal English; secondly, diachronic criteria, which

were applied for the selection of the nuclear semantic feature at the centre of the LC of “charge”.

- 35 The origins of this term, which derives from Old French, date back to the 13th century, when “charge”, as a noun, was associated to the meaning “load” or “weight”. According to the *New Oxford English Dictionary* (1998), there is no documented use of this word with a legal meaning until the middle of the 15th century, when it was also employed to refer to a formal accusation for having committed a crime. It was not until the beginning of the 16th century that “charge” started being used to refer to a “pecuniary burden” or “cost”. Later in the 18th century, with the discovery of electricity, it developed a new meaning to denote the electrical load necessary for a device to work.
- 36 Having studied a large number of concordances associated to the different collocates of the term “charge” extracted from *LACELL*, the general corpus, it was found that its most frequent meaning could be identified with the price or cost of a service, goods, and the like. This fact is proved by the frequency of some of the most frequent collocates of the term such as “community” ($F^{13}=564$); “free” ($F=235$); “benefit” ($F=138$) or “service” ($F=72$) in sentences like “this booklet is free of charge” or “a service charge will be payable in full with the first instalment”. Nevertheless, a decision was made to place its original meaning at the core of the semantic constellation (see figure 1) owing to the fact that it could be more easily associated with the rest of the definitions than its most frequent meaning in general English.

Figure 1. Semantic Lexical Constellation of the term “charge”¹⁴



- 37 The lexical constellation was organised relying on actual language usage and applying the quantitative method described in greater detail in section 4.3. The hierarchy established for the semantic analysis of the term was related to the distance among its various definitions found in dictionaries and already interpreted and organised with the help of two specialists. The main reason for placing the first five meanings of the

word on a first hierarchical level within the LC was precisely related to the greater semantic proximity between those five definitions of the term and its original meaning. Concerning their order, meanings 1 to 7, the decision was based on the results obtained after implementing the formula suggested in section 4.3.

- 38 As a result, having placed the oldest meaning of the word at the centre of its semantic constellation, the nucleus of the semantic LC of “charge” expands into five main directions which constitute the highest semantic level within it, level 1 in table 6. The original “weight” or “load” denotation can be understood as a “monetary weight; an electrical load in a battery; a quantity of explosives; a physical weight against someone and a non-physical weight on someone”. Except for the second, third and seventh meanings of “charge” shown above, the rest could be understood as rather literal interpretations of its base meaning, owing to the fact that they can somehow be physically felt or measured. In fact, an amount of explosives ready to detonate is something that can be touched or weighed, if necessary, the sentence “detonator exploded but failed to set off the main charge”, taken from *LACELL*, exemplifies this definition. The electrical load in a battery, which generates energy that turns into light or motion, could also be deemed a physical weight, as shown by this concordance line also extracted from *LACELL*: “Alter the main battery leads to use the front battery to start and run the vehicle and wire the battery under the seat to a split charge relay [...]”. Regarding the fifth meaning in figure 1, the physical charge entailed by this word could be read as one which is launched or activated against others, that is, an “attack against the enemy”, e.g., “[...] leaving Fanatics in front of the chargers and between them and their target unit, then the chargers can either remain halted or complete their charge [...]” (*LACELL*).
- 39 On the other hand, a “monetary weight” and a “non-physical weight on somebody” cannot be interpreted *verbatim* but rather as figurative weights that someone has to bear. The cost assigned to goods might be understood as a load which the buyer must carry to purchase them. In fact, in their definitions we still find the word “weight” as indicative of that rather figurative load that someone has to carry, for instance, “[...] again, all absolutely free of charge. We don't charge you a penny” or “I cried a lot coming out of my teens, says Charlotte, now 22, because I realised that I no longer had an excuse to play out the role of mother's beloved charge” (*LACELL*).
- 40 Directly deriving from the idea of “monetary weight”, there is the concept of “debt” (i.e., entry on a bank account which is not instantly paid), whose semantic distance from the nucleus of the constellation is higher due to the fact that the association between the original idea of “weight” and the concept of “debt” is harder to trace. This is why this definition stands at a lower semantic level within the LC (level 2 in table 6), as shown in figure 1. A “debt” might be regarded as a weight on someone's conscience which must be born until it is settled. The following sentence, extracted from *LACELL*, illustrates this point: “One of the advantages of shopping with us is that Barclays Bank will place the credit card charges monthly so that they can be paid in instalments and, if not paid [...]”.
- 41 Concerning the legal meaning of “charge”, it is closely linked to meaning 3 in the constellation, a “non-physical weight on somebody”, however, a further nuance is added to it, that of a burden which is placed on somebody's shoulders when he/she is formally accused of a crime by a legal authority. As well as “debt”, this meaning of “charge” has been placed on a lower semantic level than the one it originates from,

implying further semantic distance from LC nucleus and thus an indirect dependency on it.

- 42 All in all, while there is a clear connection among the meanings of “charge” stemming directly from the nucleus (level 1 meanings) and the nucleus of the LC itself, the secondary meanings which the term acquires bear almost no relationship with the central meaning of the constellation. Actually, a “debt” and “an accusation before a judge” do not apparently seem to have anything in common, yet, thanks to the possibilities offered by the LC model, this connection can be traced back by following an inverse path from the secondary meanings towards the LC nucleus.

4.3. The quantitative approach

- 43 In order to ground our semantic analysis of “charge” on actual language usage, we resorted to the data offered by our general and specialised corpora. After examining and arranging the definitions of “charge” according to their distance/proximity to its original meaning, the meanings in figure 1 were ranked from 1 to 7 applying quantitative criteria. The ranking method consisted in selecting the collocates associated to each of the different meanings of “charge”, which implied reading the concordances generated by those collocates and ensuring that they were correctly related to each of the meanings of the term under examination. Both the collocates and the concordances associated to them were obtained by processing both *BLaRC* and *LACELL* with *Wordsmith 5.0* (Scott 2008). A collocate frequency threshold of 5 was set so as to prevent the number of concordances from becoming unmanageable.
- 44 Subsequently, the average frequency of the collocates related to each of the meanings was calculated. The data was normalised¹⁵ for comparison and the result obtained from *LACELL* was divided by the same datum in *BLaRC* applying the following formula:

$$\frac{\overline{\mu_i^G}}{\overline{\mu_i^T}}$$

where

$$\overline{\mu_i^G}$$

stands for the average collocate frequency in the former corpus and

$$\overline{\mu_i^T}$$

- 45 for the same datum in the latter. The resulting score (labelled as “sub-technicality score” hereinafter) was deemed indicative of the semantic proximity or distance of each of the meanings of “charge” from the general field.
- 46 We departed from the assumption that there is no cut-off point that differentiates terms from non-terms, although many different ATR methods apply thresholds and discard candidate terms simply based on quantitative criteria (see Marín [2014] for a further discussion on this topic). We assumed that sub-technical terms stand somewhere in between general and specialised languages, moving along a continuum of specialisation. Their proximity or distance from each end of this continuum depends on their frequency of usage and their meaning in both contexts. Hence, the higher the average frequency of the collocates associated to each meaning in the general field, the higher the score obtained and, therefore, from our perspective, the closer a given term

would be to the original meaning of the word within the constellation, which is non-specialised. As shown in table 6, the highest ranking meanings of “charge” are those which do not generate any collocates in the specialised corpus, thus, the result obtained after dividing their average collocate frequency by zero tended to infinity (∞).

- 47 Conversely, if the meaning of a word like “charge” was more specialised,¹⁶ it would necessarily be linked to a smaller amount of collocates in the general corpus, thus generating a greater number of them in the specialised one. This is the reason why meaning number 7 in the LC, “an accusation by a judge or a competent authority”, was associated to a very low score (0.47). Subsequently, the formula proposed above was applied and the sub-technicality score quantifying the semantic distance between the different meanings of the term and its original one was calculated. Table 5 illustrates the data used to calculate the score assigned to the different meanings of “charge” as shown in its semantic LC.

Table 5. Data used to calculate the sub-technicality coefficient

Meaning	Normalised collocate frequency in BLaRC	Normalised collocate frequency in LACELL	Sub-technicality score LACELL/BLaRC
Nucleus: A weight, a load	0	12.24	$12.24/0 = \infty$
1: Electrical load in a battery	0	7.63	$7.63/0 = \infty$
2: Monetary weight	0	29.06	$29.06/0 = \infty$
3: A non-physical weight on sb.	0.27	6.3	$6.3/0.27 = 23.4$
4: A quantity of explosives	0.82	13.44	$13.44/0.82 = 16.4$
5: Attack against enemy	0.14	4.5	$4.5/0.14 = 32.14$
6: Debt	1.53	13.15	$13.15/1.53 = 8.6$
7: A formal accusation by a judge	4.22	1.98	$1.98/4.22 = 0.47$

- 48 The constellation was therefore organised applying two methods. Firstly, the qualitative one whereby the semantic proximity/distance of a given meaning to the nucleus was established and secondly, the sub-technicality score obtained after applying the formula above. Both the qualitative and quantitative approaches to the analysis of “charge” took place simultaneously so that they complemented each other for the organisation of the semantic LC of the term.

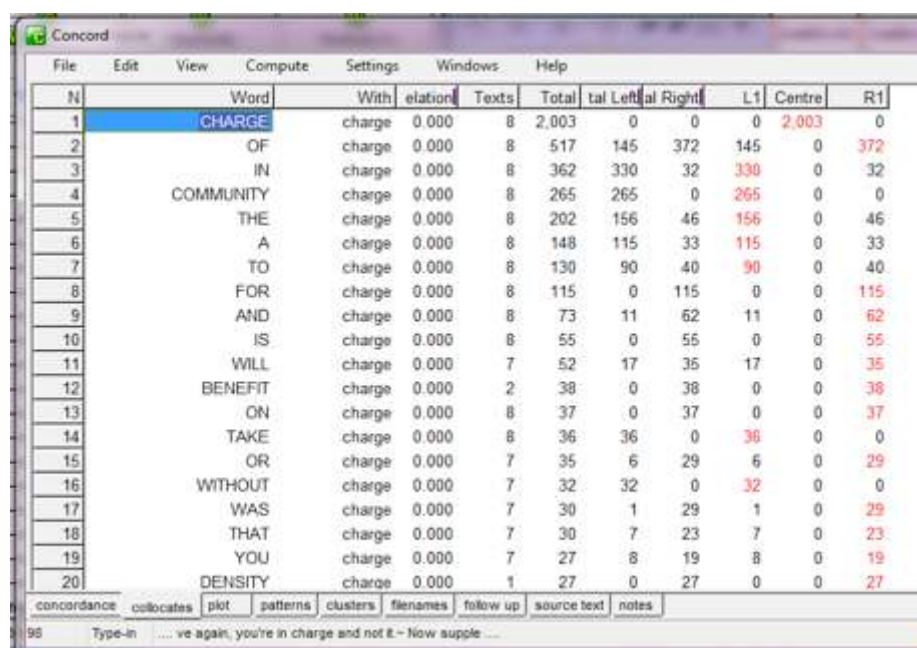
- 49 Both table 6 and figure 1 present the two hierarchical levels in the constellation of “charge”: level 1, which includes those meanings of the term whose semantic proximity to the nucleus is greater, being easily related to the original definition of the word, i.e., “an electrical load in a battery” (meaning 1); “monetary weight” (meaning 2) or “an attack against the enemy” (meaning 3) and level 2, which comprises those meanings of the word whose relationship with the nucleus of the constellation is harder to trace, therefore displaying an indirect semantic dependency on the nucleus of the constellation. This is the case of a “debt” (meaning 6) and an “accusation by a judge” (meaning 7).

Table 6. Structure of the LC of “charge”

LC level	Sub-technicality score LACELL/BLaRC	Position
Nucleus	∞	Nucleus: A weight, a load
LC level 1	∞	Meaning 1: Electrical load in a battery
LC level 1	∞	Meaning 2: Monetary weight
LC level 1	23.4	Meaning 3: A non-physical weight on sb.
LC level 2	16.4	Meaning 4: A quantity of explosives
LC level 2	32.14	Meaning 5: Attack against enemy
LC level 2	8.6	Meaning 6: Debt
LC level 2	0.47	Meaning 7: A formal accusation by a judge

- 50 Once the two levels that the constellation could be organised into were defined, the scores indicated in tables 5 and 6 served as a way to quantify the actual distance of every meaning of “charge” from its original one, “weight” or “load”. Its legal definition was the lowest ranking owing to the lesser average frequency of the collocates associated to it in the general corpus. In fact, as it was expected, almost none of the collocates associated to the term in *LACELL* were legal terms. Figure 2 shows a screenshot of the list of collocates generated by *Wordsmith*. Columns L1 and R1 indicate the number of times that the word “charge” co-occurs with the words listed to the left of the image. As it can be observed, there are no legal terms among its top 20 collocates. The column “centre” shows the total number of times “charge” is used in the corpus. Actually, the only lexical collocates of “charge” within this group are “community” (F=265) and “benefit” (F=38), as exemplified by sentences such as “so what sort of rent were they paying prior to the introduction of the community charge under the old rating system?” or “people will also qualify for a community charge reduction; and/or some charge benefits”.

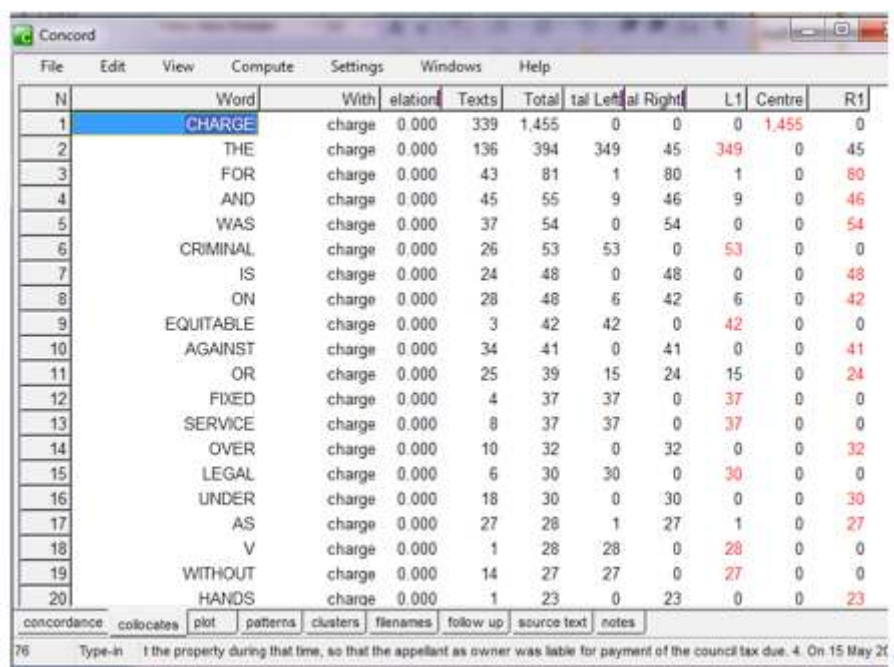
Figure 2. Collocate list of “charge” obtained from LACELL



The screenshot shows the LACELL software interface with a concordance table for the word "charge". The table has columns for N, Word, With, elation, Texts, Total, tal Left, al Right, L1, Centre, and R1. The word "charge" is highlighted in the first row. The table lists various words that collocate with "charge", such as "OF", "IN", "COMMUNITY", "THE", "A", "TO", "FOR", "AND", "IS", "WILL", "BENEFIT", "ON", "TAKE", "OR", "WITHOUT", "WAS", "THAT", "YOU", and "DENSITY". The values in the table represent the frequency of each collocation.

N	Word	With	elation	Texts	Total	tal Left	al Right	L1	Centre	R1
1	CHARGE	charge	0.000	8	2,003	0	0	0	2,003	0
2	OF	charge	0.000	8	517	145	372	145	0	372
3	IN	charge	0.000	8	362	330	32	330	0	32
4	COMMUNITY	charge	0.000	8	265	265	0	265	0	0
5	THE	charge	0.000	8	202	156	46	156	0	46
6	A	charge	0.000	8	148	115	33	115	0	33
7	TO	charge	0.000	8	130	90	40	90	0	40
8	FOR	charge	0.000	8	115	0	115	0	0	115
9	AND	charge	0.000	8	73	11	62	11	0	62
10	IS	charge	0.000	8	55	0	55	0	0	55
11	WILL	charge	0.000	7	52	17	35	17	0	35
12	BENEFIT	charge	0.000	2	38	0	38	0	0	38
13	ON	charge	0.000	8	37	0	37	0	0	37
14	TAKE	charge	0.000	8	36	36	0	36	0	0
15	OR	charge	0.000	7	35	6	29	6	0	29
16	WITHOUT	charge	0.000	7	32	32	0	32	0	0
17	WAS	charge	0.000	7	30	1	29	1	0	29
18	THAT	charge	0.000	7	30	7	23	7	0	23
19	YOU	charge	0.000	7	27	8	19	8	0	19
20	DENSITY	charge	0.000	1	27	0	27	0	0	27

- 51 Conversely, (as it was indeed foreseeable), 58.33% of all the collocates of “charge” found in *BLaRC* refer to legal concepts. Figure 3 reveals that “criminal” (F=53), “equitable” (F=42) or “legal” (F=30) frequently occur with it. Sentences like “[...] in the determination of any criminal charge against him, everyone is entitled to a fair and public hearing”, or “the court should reflect the attempted fraud on Bobs by giving Lexi only an equitable charge for £6.5 million” demonstrate how these words collocate with this legal term.
- 52 In sum, it could finally be stated that the legal meaning of “charge”, using the LC model for its description, is the most unrelated definition of the term to the one selected as the nucleus of the constellation, “weight” or “load”. Actually, the legal meaning of this term follows a process of specialisation which reflects the greater semantic distance from its original meaning. This is why it displays an indirect semantic dependency on the nuclear meaning of its LC, which places it on a secondary level within its semantic constellation. This statement is supported by the collocational data linked to its legal meaning in both the general and the specialised fields, situating it at the very bottom of the ranking used to organise the semantic constellation of this sub-technical term due to its lower sub-technicality score.

Figure 3. Collocate list of “charge” obtained from *BLaRC*


The screenshot shows the Concord software interface with a menu bar (File, Edit, View, Compute, Settings, Windows, Help) and a toolbar. The main window displays a table of collocates for the word 'CHARGE'. The table has columns: N, Word, With, elation, Texts, Total, tail, Left, al, Right, L1, Centre, R1. The data is sorted by elation, with 'CHARGE' at the top (0.000) and 'HANDS' at the bottom (0.000). The 'Texts' column shows the frequency of each word in the corpus, and the 'Total' column shows the total frequency of the word 'CHARGE' in the corpus. The 'tail', 'Left', 'al', 'Right', 'L1', 'Centre', and 'R1' columns show the frequency of each word in the left and right contexts of 'CHARGE'.

N	Word	With	elation	Texts	Total	tail	Left	al	Right	L1	Centre	R1
1	CHARGE	charge	0.000	339	1,455	0	0	0	0	1,455	0	0
2	THE	charge	0.000	136	394	349	45	349	0	45	0	45
3	FOR	charge	0.000	43	81	1	80	1	0	80	0	80
4	AND	charge	0.000	45	55	9	46	9	0	46	0	46
5	WAS	charge	0.000	37	54	0	54	0	0	54	0	54
6	CRIMINAL	charge	0.000	26	53	53	0	53	0	0	0	0
7	IS	charge	0.000	24	48	0	48	0	0	48	0	48
8	ON	charge	0.000	28	48	6	42	6	0	42	0	42
9	EQUITABLE	charge	0.000	3	42	42	0	42	0	0	0	0
10	AGAINST	charge	0.000	34	41	0	41	0	0	41	0	41
11	OR	charge	0.000	25	39	15	24	15	0	24	0	24
12	FIXED	charge	0.000	4	37	37	0	37	0	0	0	0
13	SERVICE	charge	0.000	8	37	37	0	37	0	0	0	0
14	OVER	charge	0.000	10	32	0	32	0	0	32	0	32
15	LEGAL	charge	0.000	6	30	30	0	30	0	0	0	0
16	UNDER	charge	0.000	18	30	0	30	0	0	30	0	30
17	AS	charge	0.000	27	28	1	27	1	0	27	0	27
18	V	charge	0.000	1	28	28	0	28	0	0	0	0
19	WITHOUT	charge	0.000	14	27	27	0	27	0	0	0	0
20	HANDS	charge	0.000	1	23	0	23	0	0	0	0	23

At the bottom of the window, there is a toolbar with buttons for concordance, collocates, plot, patterns, clusters, filenames, follow up, source text, and notes. A status bar at the very bottom shows the page number 76 and the text 'Type-in t the property during that time, so that the appellant as owner was liable for payment of the council tax due. 4. On 15 May 20'.

Conclusion

- 53 The presence and use of terms shared by the general and specialised fields, generally labelled as “sub-technical”, is a highly relevant trait of legal English. As data show, almost half of the term list obtained from a legal corpus of judicial decisions is also present amongst the most frequent words of English. In our view, the study of the linguistic behaviour of these terms might certainly enhance our understanding of the legal lexicon.
- 54 This is the reason why this article has presented a proposal for the study of sub-technical terms adopting both a qualitative and a quantitative perspective on the issue. This analysis was carried out through the application of Cantos and Sánchez’s (2001) LC model for the description of the process of specialisation undergone by the term “charge”.
- 55 Concerning the qualitative approach, the original meaning of the term was singled out as the nucleus of the constellation, constituted by the different definitions of “charge” found in general and specialised dictionaries. As for the latter approach, a sub-technicality score was obtained after applying a formula to try to quantify the semantic distance existing between the different meanings of the term and its original one.
- 56 Using all the data obtained applying both methods, the LC of “charge” was created, which displays both the semantic dependencies amongst its constituents and their hierarchical associations in a visual manner, being organised at two different levels depending on their semantic proximity to the nuclear meaning of the term.
- 57 In short, this model of analysis has aimed to shed some light on the characteristics and linguistic behaviour of sub-technical legal terms, using corpus data as evidence to draw our conclusions from. Nevertheless, we must acknowledge the limitations of a methodology which, in spite of producing reliable results based on real language usage,

might be time-consuming owing to the fact that concordance lines must be read to ensure a correct quantification of the data. Hence, the process should be, at least, partially automated.

- 58 As a result, the aims for our future work in the field will be, on the one hand, to suggest the initial steps to build a method of word sense disambiguation based on corpus data and, on the other hand, to continue with the analysis of sub-technical legal vocabulary using corpus data as a solid base, which may help to illustrate the path followed by these terms towards specialisation given their relevance in legal English.

BIBLIOGRAPHY

- Alcaraz Varó, Enrique. 1994. *El inglés jurídico: textos y documentos*. Madrid: Ariel Derecho.
- Almela, Moisés. 2006. *From Words to Lexical Units: a corpus-driven account of collocation and idiomatic patterning in English and English-Spanish*. Frankfurt: Peter Lang.
- Baker, Mona. 1988. "Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles". *Reading in a Foreign Language* 4/2, 91–105. Retrieved on 2/9/2014 <<http://www.nflrc.hawaii.edu/rfl/PastIssues/rfl42baker.pdf>>.
- Borja Albí, Anabel. 2000. *El texto jurídico en inglés y su traducción*. Barcelona: Ariel.
- Cantós, Pascual & Aquilino Sánchez. 2001. "Lexical constellations: what collocates fail to tell". *International Journal of Corpus Linguistics* 6/2, 199–228.
- Chung, Teresa Mihwa & Paul Nation. 2003. "Technical vocabulary in specialised texts". *Reading in a Foreign Language* 15/2, 103–116.
- Cowan, J. Ronayne. 1974. "Lexical and syntactic research for the design of EFL". *TESOL Quarterly* 8, 389–399.
- Coxhead, Averil. 2000. "A new academic word list". *TESOL Quarterly* 34/2, 213–238.
- Cruse, Alan. 2004. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Drouin, Patrick. 2003. "Term extraction using non-technical corpora as a point of leverage". *Terminology* 9/1, 99–117.
- Farrell, Paul. 1990. *Vocabulary in ESL: A Lexical Analysis of the English of Electronics and a Study of Semi-technical Vocabulary*. Dublin, Ireland: Centre for Language and Communication Studies.
- Flowerdew, John. 2001. "Concordancing as tool in course design". In Ghadessi, M., A. Henry & R. Roseberry (eds.), *Small Corpus Studies and ELT: Theory and practice*. Amsterdam: John Benjamins, 71–92.
- Kilgariff, Adam. 2006. "Word senses". In Aguirre, E. & P. Edmonds (eds.), *Word Sense Disambiguation. Algorithms and Applications*. Berlin: Springer Verlag, 29–46.
- Marín, Maria José. 2014. "Evaluation of five single-word term recognition methods on a legal English corpus". *Corpora* 9/1, 83–107.

- Marín, María José & Camino Rea. 2012. "Structure and design of the BLRC: a legal corpus of judicial decisions from the UK". *Journal of English Studies* 10, 131–145. Retrieved on 2/9/2014 <<http://www.unirioja.es/servicios/sp/ej/jes/jes10.shtml>>.
- McEnery, Tony & Adam Wilson. 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, Tony & Richard Xiao. 2006. *Corpus-based Language Studies: A resource book*. London and New York: Routledge.
- Mellinkoff, David. 1963. *The Language of the Law*. Boston: Little, Brown and Co.
- Nazar, Rogelio & María Teresa Cabré. 2012. "Supervised learning algorithms applied to terminology extraction". In Aguado de Cea, G., M.C. Suárez-Figueroa, R. García-Castro & E. Montiel-Ponsoda (eds.), *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012)*. Madrid: Ontology Engineering Group, Association for Terminology and Knowledge Transfer, 209–217.
- Orts Llopis, María Ángeles. 2006. *Aproximación al discurso jurídico en inglés: las pólizas de seguro marítimo de Lloyd's*. Madrid: Edisofer.
- Pearsall, J. (ed.). *New Oxford Dictionary of English*. 1998. New York: Oxford University Press.
- Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam: John Benjamins.
- Rea, Camino. 2010. "Getting on with corpus compilation: from theory to practice". *ESP World* 1/27, volume 9, Retrieved on 2/9/2014 <http://www.esp-world.info/articles_27/camino%20rea.pdf>.
- Rea, Camino & Aquilino Sánchez. 2010. "Building new meanings in technical English from the perspective of the lexical constellation model". *Iberica*, 20, 107–126. Retrieved on 2/9/2014 <<http://www.aelfe.org/documents/05Rea.pdf>>.
- Saint Dahl, Henry. 1999. *Dahl's Law Dictionary. Diccionario Jurídico Dahl*. New York: William S. Hein and Co., Inc.
- Sánchez, Aquilino, Pascual Cantos & Moisés Almela. 2010. "In search of a new model for WSD: a multi-layered lexical constellation model". In Sánchez, A. & P. Cantós (eds.), *A Survey on Corpus-based Research/Panorama de investigaciones basadas en corpus*. Murcia: Asociación española de lingüística del corpus, 1235–1241. Retrieved on 2/9/2014 <<http://www.um.es/lacell/aelinco/contenido/pdf/86.pdf>>.
- Scott, Mike. 2008. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.
- Sinclair, John. 2005. "Corpus and text: basic principles". In Wynne, M. (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. AHDS Literature, Languages and Linguistics: University of Oxford. Retrieved on 2/9/2014 <<http://ota.ahds.ac.uk/documents/creating/dlc/index.htm>>.
- Tiersma, Peter. 1999. *Legal Language*. Chicago: The University of Chicago Press.
- Trimble, Louis. 1985. *English for Science and Technology: A Discourse Approach*. Cambridge: Cambridge University Press.
- Vargas, Chelo. 2005. "Aproximación terminográfica al lenguaje de la piedra natural: propuesta de sistematización para la elaboración de un diccionario traductológico". Unpublished doctoral thesis. Universidad de Alicante.
- Wang, Karen & Paul Nation. 2004. "Word meaning in academic English: Homography in the academic word list". *Applied Linguistics* 25/3, 291–314.
- West, Michael. 1953. *A General Service List of English Words*. London: Longman. <<http://jbauman.com/gsl.html>>.

NOTES

1. For further information, see the websites of: the Judiciary of England and Wales, <<http://www.judiciary.gov.uk/about-the-judiciary/introduction-to-justice-system/court-structure>>; the Scottish court and tribunal system, <<http://www.scotland.gov.uk/Topics/Justice/legal>>; the Courts and tribunals in Northern Ireland, <<http://www.courtsni.gov.uk/en-GB/pages/default.aspx>>.
2. <http://www.lex tutor.ca/concordancers/concord_e.html>
3. <<https://the.sketchengine.co.uk/login>>
4. The term “skewness” refers to the accuracy of the results obtained after comparing the two corpora automatically. In statistics, a result is skewed if it is not symmetrical about the means. In the field of corpus linguistics, if two corpora are compared, they must be as similar/symmetrical as possible (especially as regards the geographical variants of the language) for the results to be reliable.
5. <<http://termostat.ling.umontreal.ca>>
6. <<http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl>>
7. The term “gold standard” is used to refer to the reliability of the glossary of legal terms compiled by Marín (2014), which is based on the four online glossaries cited below and available online.
8. Available at: < <http://www.legislation.gov.hk/eng/glossary/homeglos.htm>>; < <http://www.judiciary.gov.uk/glossary>>; <http://sixthformlaw.info/03_dictionary/index.htm>; <<http://www.nolo.com/dictionary>>.
9. By “true terms”, we refer to those terms which could be found in the glossary and could thus be confirmed as real ones. This concept has been borrowed from the field of Natural Language Processing where it is used as opposed to “candidate terms”, those terms automatically identified by ATR methods before validation.
10. The weighting or specificity score assigned by each method is calculated by processing both corpora, lemmatising them and implementing their algorithms, which might take into consideration a variety of parameters such as frequency, text distribution, grammatical category, word or sentence length, etc.
11. This sample shows the top 25 sub-technical legal terms included in our inventory (15 in table 1 and 10 in table 2) identified using *Terminus* and *TermoStat*. The resulting lists were compared with the 3,000 most frequent words of *BNC* to find out which legal terms overlapped with the latter. After that, they were separated and arranged according to the ATR method and the level of specialisation obtained for each word type.
12. The ones identified as such by corpus analysis tools such as *Wordsmith* (Scott 2008) owing to their frequent co-occurrence in a given corpus.
13. *F* stands for the frequency of the cluster formed by “charge” and its collocates.
14. The different meanings of the word have been numbered according to the criteria detailed in section 4.3.
15. See section 3.1 for details on normalisation. In this case, the data obtained from LACELL were divided by 14.6, the number of millions of words which form this corpus.
16. The concept “specialised” refers to the presence of a term exclusively in texts which belong to a specialised field, which implies that it would probably not be understood by the layman. The meaning of “charge” cannot be deemed highly specialised because of that fact. It can occur in general and specialised texts denoting a legal concept which anyone could grasp without any difficulty.

ABSTRACTS

This article suggests a new approach for the semantic description of “sub-technical terms” applying Cantos and Sánchez’s (2001) Lexical Constellation model, which allows for a visual representation of the path they follow towards specialisation. Various authors highlight the use of sub-technical terms as an outstanding feature of the legal lexicon. This study attests that almost half of the terms in a legal corpus of 8.85 million words overlap with general English vocabulary lists such as West’s (1953) General Service List (GSL) or the British National Corpus (BNC) most common 3,000 words of English. Therefore, given the relevance of these terms within this ESP variety and the lack of proposals for their analysis, an attempt is made to try and account for such a complex phenomenon adopting a semantic and a corpus-based perspective. This is achieved employing a qualitative and quantitative methodology for the description of the process of specialisation of the sub-technical legal term “charge”, used to exemplify our model of analysis.

Cet article propose une nouvelle approche pour la description sémantique de « termes semi-techniques » à l’aide du modèle de « constellation lexicale » de Cantos et Sánchez (2001), qui permet la représentation visuelle de la trajectoire suivie par ce type de termes vers la spécialisation. Plusieurs auteurs mettent en avant le fait que l’utilisation de termes semi-techniques est une caractéristique saillante du lexique juridique. Cette étude met en évidence le fait que près de la moitié des termes dans un corpus juridique de 8,85 millions de mots recoupe des listes de vocabulaire d’anglais général comme la General Service List (GSL) de West (1953) ou les 3 000 mots d’anglais les plus courants du British National Corpus (BNC). Étant donné l’importance de ces termes dans cette variété d’anglais de spécialité et le manque d’étude dans la spécialité, les auteurs rendent compte de ce phénomène complexe en adoptant une perspective sémantique fondée sur un corpus. Cette étude a été réalisée au moyen d’une méthodologie qualitative et quantitative pour décrire le processus de spécialisation du terme légal semi-technique « charge », utilisé pour illustrer le modèle d’analyse.

INDEX

Mots-clés: analyse de corpus, constellation lexicale, terme semi-technique, terminologie juridique anglaise

Keywords: corpus-based analysis, legal English terminology, lexical constellation, sub-technical term

AUTHORS

MARÍA JOSÉ MARÍN

María José Marín has been a lecturer at the English Department of the University of Murcia, Spain since 2009. She holds a PhD in corpus linguistics and has published several articles in journals such as *Corpora*, *Procedia* or *Miscelánea: A Journal of English and American Studies*. Her areas of interest are ESP, legal English and corpus linguistics. mariajose.marin1@um.es

CAMINO REA

Camino Rea has worked at the Language Centre of the Technical University of Cartagena since 1999. She joined the English Department at the University of Murcia in 2007 as a part-time lecturer. She obtained her PhD in corpus linguistics in 2008, being specialised in ESP within the area of telecommunication English. She has published articles in journals such as *IJES*, *Ibérica* or *ESP World*. camino.rea@um.es