

A CORPUS-BASED PROPOSAL FOR THE GRADING OF VOCABULARY TEACHING MATERIALS FOR THE LEGAL ENGLISH CLASS

María José Marín Pérez, Camino Rea Rizzo, University of Murcia. Spain

Abstract

The introduction of legal English as a compulsory subject in the curriculum of the Law Degrees taught at Spanish universities due to the implementation of the Bologna Reform has led to the design of syllabuses which are intended to enable students to become proficient users of this English variety with both academic and professional purposes. This paper presents a corpus-based proposal for the grading of materials for the teaching of legal vocabulary which can be extrapolated to other varieties of English for Specific and Academic Purposes (ESAP). In order to exemplify it, a sample list of 33 crime nouns (obtained from the legal English textbooks consulted) has been examined in terms of their frequency, keyness and text range values in an *ad hoc* legal corpus of 2.6 million words, *UKSCC*. After doing so, Chung's (2003) automatic term recognition (ATR) method has been applied so as to establish their level of specialisation. Our proposal relies on the assumption that the information obtained after taking these different parameters into consideration might be helpful for the ESAP instructor to rank the vocabulary inventories obtained from specialised corpora so that the materials derived from them can be graded according to the students' needs.

Keywords: Legal English, specialised corpora, ESAP, ATR methods

Introduction

The motivation of the present study originates from the introduction of Legal English as a core subject within the Degree in Law at the University of Murcia (Spain) after the Bologna Reform (implemented in the academic year 2009/2010) which implied, amongst other aspects, a greater internationalisation of university degrees as well as their adaptation to the Higher Education standards in the European Union.

The course integrates both English for General Academic purposes (EGAP) and English for Specific Academic Purposes (ESAP) to attain the main objectives established, being English the language of instruction. On the one hand, students must gain the academic skills which enable them as effective communicators in formal settings and as autonomous learners for lifelong learning. On the other hand, as legal practitioners, they might develop their career in Spain where the number of British expatriates is considerably high in certain areas. Therefore, they may have to communicate in English in order to give legal advice to this type of clients. Likewise, they should be able to explain the distinctive features of the Spanish and British legal systems, the court system and structure, or the process of a civil or criminal claim using the corresponding terminology.

The main goals of the course agree with the B2 or Vantage level on the scale of the Common European Framework of Reference for Languages (CEF) according to which language users:

Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options. (Council of Europe, 2001: 5).

In fact, the course attempts to help students advance from B1, their initial level, to B2 with a focus on the specialised register. For learners who are moving on to special purposes study, it is advisable to concentrate on specialised vocabulary since it offers greater text coverage (Nation and Hwang, 1995).

This paper revolves around crime nouns in order to find out how they are actually used by professionals and how relevant they are in legal English. The most frequent crime nouns are usually introduced in B2 general purpose textbooks but we must deal with all of them in an English for Academic Specific Purposes (EASP) setting depending on their relevance in the specific domain. Establishing the level of specialisation of these items can help to grade the materials based on them according to the students' needs.

Under these premises, crime nouns are analysed adopting a corpus comparison approach in terms of frequency, range and keyness and also classified into levels of specialisation. The combination of such parameters may provide an important clue about the behaviour of these words to help to plan and design materials for our legal English class. Thus, section two presents the main features of the *United Kingdom Supreme Court Corpus (UKSCC)* compiled *ad hoc* to study the behaviour of the lexical items selected. It is followed by section

three where the methodology employed for the analysis of the data and the results of such analysis are offered. To conclude, section four summarises the major conclusions drawn from the close examination of the information provided by the corpus texts and explains the future research questions posed by this study.

UKSCC and other legal corpora: State of the art

One of the major obstacles encountered for the elaboration of specialised materials to teach legal English in Higher Education contexts is the scarce amount of resources available, as it usually happens in other branches of ESAP (Rea, 2010). Resorting to specialised corpora might be an option, as McEnery and Wilson (1996) affirm: "... such corpora can be used to provide many kinds of domain-specific material for language learning" (p. 121). Schmitt (2002) believes that their use might be beneficial regarding them as a valuable teaching resource as well as a useful tool to assess vocabulary acquisition. In addition, Gilquin and Granger (2010) insist on the importance of learners' exposure to authentic materials based on corpora which also offer "a large number of authentic instances of a particular linguistic item" (p. 359) thus helping to disambiguate meanings. Nevertheless, to our knowledge, the number of legal corpora is reduced, and access to them, except for a few cases, is not complete. Let us then review the most relevant ones existing to date.

BoLC, the *Bononia Legal Corpus*¹, is, by far, the most comprehensive legal corpus available due to its size (the English subsection amounts to 59 million words) and generic diversity. It can be freely accessed online (though not downloaded) and has an English and an Italian section. Barring *BoLC*, the rest of the corpora studied herein do not satisfy our needs either because they are too small to act as reference or because they focus on aspects of the language we are not interested in.

The *JRC-Acquis* corpus is another legal corpus which can be consulted online. It is a multilingual comparable collection of legal texts in 22 different languages focusing on European legislation. Likewise, the *CorTec* corpus is a scientific-technical collection of texts comparing English and Brazilian Portuguese devoted to commercial law. Conversely, the *House of Lords Judgment* corpus (*HOLJ*), is monolingual and includes judicial decisions from the House of Lords focusing on the definition of rhetorical role labels. Finally, Cambridge University Press owns the Cambridge International Corpus which is neither accessible nor downloadable.

For these reasons, *UKSCC*, the *United Kingdom Supreme Court Corpus*, was created. It is a monolingual legal English corpus of 2.6 million words whose main aim is to act as a reliable source of specialised vocabulary. It is composed of 193 judicial decisions from the UK Supreme Court and the House of Lords issued from 2008 to 2010.

Judicial decisions have been selected as the genre to base *UKSCC* on due to the pivotal role they play in common law legal systems, acting as the major source of information for legal practitioners. The Supreme Court of the United Kingdom has been chosen as a source to obtain the texts owing to its position at the top of the judicial pyramid and the fact that its

¹ All the corpora referred to in this section (except the one owned by CUP) are available online as indicated in the reference section.

decisions always set precedent being most often cited by judges and barristers at court. It also deals with all branches of law providing rich and varied texts as far as their lexicon is concerned.

Methodology

There exist different methods to analyse the information obtained from linguistic corpora. The literature on the subject shows how authors employ such methods to analyse and classify specific lexicon (Yang, 1986; Farrell, 1990; Coxhead, 2000; Nation, 2001; Rea, 2008, amongst others). The use of stop lists to discriminate general from specific terms is present in most of them. Michael West's (1953) *General Service List* of the most frequent 2,000 word families of English is one of the earliest general English inventories. Other more recent general vocabulary listings employed with this purpose are the *Academic Word List* (AWL) (Coxhead, 2000), or the *British National Corpus* lists (2007).

As observed in the literature reviews (Maynard and Ananiadou, 2000; Cabré et al., 2001; Lemay et al., 2005; Chung, 2003, to name but a few), most ATR (Automatic Term Recognition) methods concentrate on multi-word terms neglecting single-word ones to a certain extent. One of the methods which focus on the latter is Chung's (2003) who establishes a cut-off point to discriminate terms (words with a specialised meaning) from non-terms (both general and also sub-technical words –those shared by both the specialised and general contexts–) after comparing a general and a specific English corpus. Chung reaches the conclusion that “a type had to occur at least 50 times more often in the technical text than in the comparison corpus, or only occur in the comparison corpus” (Chung, 2003: 259) to be considered specific through the validation of her method by comparison with a qualitative one: the *rating scale approach*. The results of the comparison yield 86% coincidence on average, particularly regarding highly specialised words and non-terms.

The 33 crime nouns selected for this study were obtained from the legal English textbooks consulted (Fernández, 1994; Rice, 2007; Krois-Linder, 2008; Frost, 2009; Callanan, 2010; Orts, 2010) as they constitute an example of general, sub-technical and technical vocabulary whose treatment and analysis might be extrapolated to other word types not only in the legal field but also in other specialised branches of ESAP. These nouns were identified applying Chung's (2003) ATR method to the corpus compiled after scanning and processing the texts obtained from the legal English textbooks above (including 196,245 tokens), which resulted into a term inventory of 1,570 legal terms. After that, only a group of them was selected for this study owing to their specificity level (as determined by Chung's method) and also to their wide text distribution across the corpus (this datum was obtained using *Wordsmith 5.0* (Scott, 2008)), which could be regarded as an objective indicator of their representativeness.

Data analysis

Frequency, text range and keyness

The classification of crime nouns according to their level of specialisation could be employed to produce a *to-be-taught* list of terms that might be used as reference to plan and design didactic materials, as stated above. As suggested by Nation (2001) and Hwang and Nation (1995), once learners have mastered the general English vocabulary and wish to move onto the specialised field, it might be recommendable to concentrate on its lexicon as a way of

improving text coverage, thus enhancing their understanding of specialised texts (either oral or written).

The information provided after applying Chung's ATR method may also be complemented by the examination of other parameters in order to make decisions on the raking method to apply when arranging the items to be included in our *to-be-taught* vocabulary list. Let us then examine those parameters before applying this ATR method proper.

Frequency is one of them. It indicates how often a wordform appears in a given corpus, however, it cannot point towards the relevance of a word type, or how well distributed it is within a collection of texts. That is why high frequency scores just inform us about how many times a word type repeats in the corpus although, especially in corpora like *UKSCC*, comprising long texts, this may happen in just one or two of the texts in it. Therefore, if we intend to select vocabulary items to plan and design materials based on them, we should concentrate not only on the most frequent ones (which might be helpful) but also on the most representative, definitely contributing to greater text coverage.

Table 1 below shows different data associated to the terms selected. The first and second columns indicate the relative frequency of each term in both the corpora used as reference, *UKSCC* (our legal corpus) and *LACELL*, a general English corpus of 21 million words compiled and owned by the *LACELL* research group at the University of Murcia, which both authors of this article are members of. Due to the different size of both corpora, the data had to be normalised for comparison². This was done by multiplying the raw frequency of the data obtained with *Wordsmith* by 1,000 and then dividing it by the number of tokens in each corpus, namely, 2,628,915 for *UKSCC* and 21,016,504 for *LACELL*. The third column of the table indicates the term's text range, that is, its distribution across *UKSCC*, for instance, *conspiracy* appears in 22 out of the 193 texts which form the corpus.

As regards the fourth column of the table, it shows the ratio obtained after applying Chung's (2008) ATR method, the value employed to rank the terms, which points to the level of specialisation of each term by comparing its frequency in the specialised corpus with the same value in the general one. Finally, the fifth column presents the *keyness* value calculated by *Wordsmith 5.0* selecting the log-likelihood algorithm amongst one the possible options offered by Scott's (2008) software.

² This is why the frequency value is *relative* and not *raw* or *absolute*.

Table 1. *Crime nouns: Relative frequency (RF) in UKSCC and LACELL, text range (TR), Chung's ratio (CR) and keyness values*

Word type	RF	RF	TR	CR	Keyness
	UKSCC	LACELL			
Conspiracy	0,1357	0,0005	22	237,8320	838,205
Trafficking	0,1065	0,0009	14	117,8110	870,371
Bribery	0,0076	9,5163E-05	3	79,9436	26,861
Perjury	0,0057	9,5163E-05	6	59,9577	565,04
Fraud	0,1700	0,0028	28	59,5580	1006,226
Nuisance	0,0216	0,0003	13	56,9598	63,157
Intimidation	0,0213	0,0003	12	55,9605	101,689
Manslaughter	0,0243	0,0004	8	51,1639	54,940
Arson	0,0015	4,7582E-05	3	31,9774	356,98
Threats	0,0327	0,0016	22	20,2210	40,403
Battery	0,0627	0,0038	4	16,4880	204,735
Abduction	0,0038	0,0002	3	15,9880	123,98
Forgery	0,0064	0,0011	3	5,6626	24,704
Torture	0,0448	0,0095	16	4,7166	146,088
Larceny	0,0007	0,0002	1	2,6647	567,83
Mayhem	0,0003	0,0001	1	2,6647	899,23
Assault	0,0612	0,0244	30	2,5089	87,134
Robbery	0,0114	0,0054	14	2,0854	11,06
Theft	0,0201	0,0107	16	1,8831	15,02
Burglary	0,0053	0,0029	5	1,7765	3,34
Murder	0,0893	0,0525	28	1,7001	48,541
Rape	0,0304	0,0197	14	1,5410	11,36
Blackmail	0,0030	0,0022	8	1,3607	0,6
Violence	0,0947	0,0735	46	1,2875	12,89
Infanticide	0,0003	0,0003	1	0,9992	0
Neglect	0,0083	0,0110	10	0,7580	1,66
Kidnapping	0,0022	0,0030	5	0,7494	0,49
Homicide	0,0030	0,0042	7	0,7185	0,88
Harassment	0,0060	0,0097	10	0,6270	3,7
Extortion	0,0003	0,0007	1	0,5329	0,45
Slavery	0,0019	0,0050	1	0,3735	6,32
Vandalism	0,0003	0,0031	1	0,1211	9,56
Incest	0,0003	0,0035	1	0,1080	11,22

As illustrated in table 1, only two, *fraud* and *conspiracy*, are above the mean frequency value scored by the 6674 word types in *UKSCC* computed by *Wordsmith 5* (Scott, 2008), that is, 353.16. The rest of them are, except for four items, far below it. If put in contrast with other similar corpora like the two million-word legal section of *BNC*³ (The *British National Corpus*),

³Available at Tom Cobb's website: http://www.lex Tutor.ca/concordancers/concord_e.html

the figures are very similar (after normalizing the data per 1,000 words due to their difference in size), 0.035 for *fraud* and 0.031 for *conspiracy* respectively.

As for their distribution (or text range), they appear in 10.81 texts on average, three times less than the whole corpus mean value for this parameter, 32.31. Only six of them: *fraud*, *conspiracy*, *violence*, *murder*, *assault* and *threats*, occur between 22 to 46 texts. Their text coverage counts, calculated using Heatley and Nation's (1996) software *Range*, are also low, covering 0.001% of all the running words (tokens) in *UKSCC*, especially if compared with the mean value for the first 2,000 specialised terms in *UKSCC*, namely, 0,24%. However, in this case, owing to the small size of the sample list employed for this study, this percentage should not be decisive for the selection of the vocabulary items to be included in our *to-be-taught* list, as the amount of text covered by it could not possibly reach any higher levels.

Regarding keyness, that is, "how unusually frequent or infrequent a given type is within a corpus" (Scott 2008, p. 184), after comparing *UKSCC* to *LACELL*, it appears that crime nouns score noticeably high in comparison with the whole list of word types in the study corpus. While crime nouns show 108.75 on average, the mean value for this parameter in the whole of *UKSCC* is 116.08. Therefore, in spite of their relatively low frequency counts, text distribution and coverage, it appears that they are reasonably representative of the genre under examination.

General and academic vocabulary

Having examined this preliminary information, crime nouns still remain unclassified as regards their level of specialisation. In order to identify those which are more representative of the general and academic fields, *UKSCC* was processed with *Range* to obtain the list of the word types present in the first 3,000 words of *BNC*. Then, the data were compared with the list of crime nouns using an excel spreadsheet so as to identify the ones falling within this inventory. Only five of them appear in it, namely, *violence*, *murder*, *battery*, *threats* and *robbery*. However, after checking the concordances⁴ of *battery* and *threats*, these two wordforms were removed from the general category since they acquire a specialised meaning in the specific corpus while they have a different one in *BNC*.

Regarding academic vocabulary, having processed *UKSCC* with *Range* using *AWL* (Coxhead, 2000) as the only baseword list and compared the results with the list of crime nouns, it appears that none of them belongs to the academic group.

All in all, if only 3 out of 31 crime nouns have been found in the general list of vocabulary – which represents 9.67% of the total – and none of them in *AWL*, it could be claimed that their level of specialisation might be high. Chung's (2003) technique will thus be implemented to establish the degree of specialisation of each item in the list and separate those elements which are highly technical from those which are shared both by general and specific contexts.

⁴Checked at Mark Davies' website: <http://corpus.byu.edu/bnc>

Terms and non-terms

Chung classifies the vocabulary in her corpus into six different groups depending on their ratio of occurrence. She calculates this value by dividing a word's relative frequency in the specialised corpus by the same value in the general one. Once she validates her results through a comparison with a qualitative ATR method, she reaches the conclusion that the wordforms whose ratio is equal or higher than 50 and those not in the reference corpus are terms, whereas the rest are non-terms.

After calculating the corresponding ratios (as shown in table 1) and checking the concordances of the most doubtful cases employing the *Concordance* tool included in *Wordsmith 5*, only three elements, namely, *abduction*, *larceny* and *mayhem*, were moved to the group of terms since their concordances proved that they were only employed in specialised texts in *LACELL*. The ones found in the *BNC* general list were put in a separate group. But for *abduction*, *larceny* and *mayhem*, the rest of terms were correctly identified thus reaching 90% precision. The list produced goes as follows:

- General vocabulary: *violence*, *murder*, *robbery*.
- Non-terms: *Arson*, *threats*, *battery*, *abduction*, *forgery*, *torture*, *assault*, *theft*, *burglary*, *rape*, *blackmail*, *neglect*, *kidnapping*, *infanticide*, *homicide*, *harassment*, *extortion*, *slavery*, *vandalism*, *incest*.
- Terms: *Conspiracy*, *trafficking*, *bribery*, *perjury*, *fraud*, *nuisance*, *intimidation*, *manslaughter*.

Consequently, crime nouns could be said to be specialised to a certain extent, 24.24% of them are technical against 75.76% which belong to the category of general vocabulary and non-terms being employed in different types of contexts, not only in the specific one.

As a final suggestion, the contexts of usage of vocabulary items may also be explored using specialised and general corpora as reference. It is precisely the group of non-terms which could be of greater interest for the ESAP lecturer as there are several wordforms in this set which acquire a specialised meaning or are employed as a given word class within the legal context, differing from their usage outside it. This is the case of *battery*, *threats*, or *forgery*. As *battery* is the most outstanding case of all, below is an example of the different contexts where it appears and the senses it acquires in both *LACELL* and *UKSCC*:

LACELL:

- ... because they contain no cadmium, while the NI-CAD people point out that all rechargeable battery packs should be returned to a certified battery recycling depot when they no long... (*power source*)
- The swing by La Fureur was the latest in a battery of public appearances to plug the new CD, including the Oprah Winfrey show and CNN ... (*a high number of items*)

UKSCC:

- First, the Ashleys say that they should be entitled to seek to establish their claim in battery, as otherwise they may not recover their costs of that issue. (*criminal offence*)

- Mr James had previous convictions from the age of 17 for, amongst other things, battery, common assault, ... (*criminal offence*)

Potential applications of the terms list obtained

As already stated above, the classification of the terms selected into different categories might help the ESAP instructor to grade the vocabulary to be taught at different stages in a legal English course. Designing activities based on a specialised corpus might be beneficial for the students owing to the authentic character of the texts contained in it. As regards *UKSCC*, our legal corpus, it comprises judicial decisions made by British judges, whose relevance within the UK legal system is fundamental, therefore, resorting to this kind of material could be highly motivating for the students. In fact, *data-driven learning (DDL)* experiments carried out within the area of legal English, as stated by Boulton (2011), are scarce, so measuring the results and the benefits obtained using this kind of activities could also be regarded as a relevant contribution to the area. Tim Johns (1997), who coins the term *DDL*, also underlines the motivating character of this teaching-learning method which encourages learners' autonomy. By discovering the rules of the language underlying real samples, the students become "language detectives" (Johns 1997: 101) and learn how to learn.

The programming of the corpus-based vocabulary activities to be inserted within the course syllabus could be based on the level of specialisation of the terms used in each of them, that is, the more general the term, the sooner it could be taught. Actually, such terms as *violence*, *murder* or *robbery*, which can be considered as general words in spite of their meaning, might have already been acquired by the students before starting the course and could be included as part of introductory vocabulary activities. Similarly, such terms as *conspiracy*, *bribery*, *perjury* or *intimidation* could be introduced later in the course, owing to their more specialised character.

One of the advantages of the use of *DDL* methods is that the material obtained from corpora is highly versatile, allowing the activity designer to exploit different linguistic levels such as the morphological, syntactic or semantic ones (Marín, 2014). The activity presented below exploits the semantic level of the language and focuses on the possible interference of the students' L1 with the acquisition of new terms which present a certain resemblance with their mother tongue. The students selected would be a group of 50 Spanish Law undergraduates in the first year of the Law degree doing a legal English course for the first time. Their competence level would be B1 to B2 according to the CEFR which establishes as a main goal the capacity to understand and express yourself in a specialised environment.

Students would be provided with different samples obtained from both corpora (*UKSCC* and *LACELL*) where the word *abduction* (ranking 12th on the list of terms in table 1) was employed. Due to its Latin origin, *abduction* could be associated with its Spanish cognate *abducción* (commonly used to refer to the kidnapping of human beings by aliens and not to child kidnapping, its legal sense). They would be asked to translate the different meanings of the term in the specialised and the general corpora in order to try and measure their capacity to infer the legal sense of *abduction* in spite of the interference of the L1. The concordances below offer some contexts of usage of the term in the legal and general fields. The general

meaning of the term could coincide with the meaning of the term in Spanish, therefore, the difficulty involved in this case would be the acquisition of a new sense of the word, often employed as a collocate of *child* to refer to the kidnapping of a minor.

UKSCC:

- Hence, if a child is wrongfully removed from his country of ordinary residence, the courts of the country to which he is taken should not ordinarily exercise jurisdiction save for the purpose of sending the child back (para 2). The same should apply if a child is brought from one country to the other for the purposes of contact, and is then wrongfully retained (para 3). This very largely reflects the principal provisions of the Hague Convention on **Child Abduction**.
- This was a **child abduction** case in which a Jordanian married to a British woman took their daughter to Jordan.

LACELL:

- But when a story about the **abduction** of a Britney Spears cardboard cut-out makes the cover of a supposed "news" paper, I simply cannot hold my tongue another second. In fact, it makes me feel like screaming until my throat bleeds.
- One ufologist who collects information with more purpose than most is Ken Phillips, a downbeat ex-teacher who runs a support group for British abductees. For about a year he's been analysing **abduction** experiences, with the aim of putting together a psycho-social profile of the classic victim.

In order to measure their degree of understanding, now using only the L2, the students would be asked to match several sentences (also obtained from both corpora) with their corresponding meanings depending on the context being specialised or not. Moreover, they could also be requested to produce their own examples as a way to show their understanding of the polysemous character of the legal term *abduction*.

Final remarks and further research

This study has shown several possible applications of the use of specialised corpora in the ESAP class. They can be a reliable source of information for the planning and design of didactic materials by resorting to the vocabulary inventories extracted from them and their contexts of usage. Establishing a ranking method for these word lists might be a useful way to concentrate on those items which may contribute to a greater extent to the understanding of specialised texts. This paper proposes to take into consideration different parameters such as their level of specialisation, frequency, representativeness or text distribution. However, much remains on the part of the ESAP instructor to decide how to apply these criteria in order to contribute to the acquisition of specialised vocabulary most adequately, always adapting to their students' characteristics and needs.

This paper has thus presented a sample list of crime nouns whose method of analysis can be applied to any other word lists from either legal English or any other ESAP branches in an automatic way. The results of our analysis reveal evidence of the relevance of crime nouns in legal English. The classification carried out in relation to the different parameters considered and the authentic samples provided by both corpora lead to lay particular emphasis on those words which may activate a different meaning depending on the context

where they are used. Such characteristic poses considerable difficulties both to students when facing and producing legal texts and to the analyst when first dealing with quantitative criteria and then changing the classification on a qualitative basis.

Thus, some activities have been suggested to exploit corpus materials for the teaching of this type of vocabulary. Section four presents the advantages of using DDL methods with that purpose, owing to their motivating character, together with the actual proposal of three different activities for the teaching of the legal term *abduction*, a polysemous word which may pose difficulties for its understanding due to its formal similarity (though not semantic) with the students' L1, Spanish.

To conclude, as further research, we would like to focus on the classification of not only crime nouns but also the whole specialised terminology in our corpus so that it could be introduced and better sequenced in teaching depending on our students' needs and according to the actual usage of language in the discourse community. The present study may be considered as an incipient trial of a comprehensive and detailed analysis of legal vocabulary.

Biodata

María José Marín has been a lecturer at the English Department at the University of Murcia, Spain, since 2009. She holds a PhD in Corpus Linguistics and has published several articles in relevant 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.

Camino Rea has worked at the Language Centre of the Technical University of Cartagena since 1999 and 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*.

References

- BoLC. *Bononia Legal Corpus*. (Last accessed, July 2012). Available from:
http://corpora.dslo.unibo.it/bolc_eng.html
- Boulton, A. (2011). Data Driven Learning: the Perpetual Enigma. In Roszkowski, S. and Lewandowska-Tomaszczyk, B. (eds.) *Explorations across Languages and Corpora*. Frankfurt: Peter Lang.
- Cabré, M. T., Estopà, R., Vivaldi, J. (2001). Automatic term detection: a review of current systems. Bourigault, D., Jacquemin, C., L'Homme, M.C. (Eds.). *Recent Advances in Computational Terminology*, 2, 53-87. Amsterdam: John Benjamins, Natural Language Processing.
- Callanan, H., Edwards, L. (2010). *Absolute Legal English*. London: Delta.
- Chung, T. M. (2003). A corpus comparison approach for terminology extraction. *Terminology* 9(2), 221-246.
- CorTec corpus. (Last accessed, July 2012) Available from:
http://www.fflch.usp.br/dlm/comet/consulta_cortec.html

- Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Retrieved from: <http://www.teemeurope.eu/documents/CEFR.DOC>
- Coxhead, A. (2000). A new academic word list. *TESOL, Quarterly*, 34(2), 213-238.
- Fernández, R.L., Almendárez, I. (1994). *A guide to Legal English/ Inglés para Juristas*. Madrid: Síntesis.
- Farrell, P. (1990). *Vocabulary in ESL: A Lexical Analysis of the English of Electronics and a Study of Semi-technical Vocabulary*. Dublin: Centre for Language and Communication Studies.
- Frost, A. (2009) *English for Legal Professionals*. New York: Oxford University Press.
- Gilquin, G., Granger, S. (2010). How can data-driven learning be used in language teaching? O'Keefe, A., McCarthy, M. (Eds.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge.
- Heatley, A., Nation, I.S.P. (1996). *Range* (computer software). Wellington: Victoria University of Wellington. Retrieved from: <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- HOLJ corpus.(Last accessed, July 2012).Available from: <http://www.ltq.ed.ac.uk/SUM/CORPUS/index.html>
- Hwang, K. and Nation, P. (1995).Where would general service vocabulary stop and special purposes vocabulary begin? *System*,23(1), 35-41.
- Krois-Linder, A., Firth, M. (2008). *Introduction to International Legal English : A Course for Classroom or Self-study Use*. Cambridge: Cambridge University Press.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In Wichmann, A., Figelston, S., McEnery, T. and Knowles, G. (eds.) *Teaching and Language Corpora*. London: Longman.
- Lemay, C., LHomme, M.C., Drouin, P. (2005). Two methods for extracting "specific" single-word terms from specialized corpora: experimentation and evaluation. *International Journal of Corpus Linguistics*, 10 (2), 227-255.
- Marín, M.J. (in the press, 2014). A Proposal to Exploit Legal Term Repertoires Extracted Automatically from a Legal English Corpus. *Miscelánea: A Journal of English and American Studies*, 49. Zaragoza: Universidad de Zaragoza.
- Marín, M.J., Rea, C. (2011). "Design and compilation of a legal English corpus based on UK law reports: the process of making decisions". Carrió Pastor, M. L., Candel Mora, M.A.(Eds.). *Las Tecnologías de la Información y las Comunicaciones: Presente y Futuro en el Análisis de Córpora. Actas del III Congreso Internacional de Lingüística de Corpus*(101-110). Valencia: Universitat Politècnica de València.
- Maynard, D. and Ananiadou, S. (2000). TRUCKS: A model for automatic multi-word term recognition. *Journal of Natural Language Processing*8(1), 101–125.
- McEnery, T. and Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Orts, M.A. (2010). *Inglés jurídico y económico : manual para su aprendizaje y traducción*. Murcia: Diego Marín.
- Rea, C.(2008). *El Inglés de las Telecomunicaciones: Estudio Léxico Basado en un Corpus Específico*. Tesis doctoral. Murcia: Universidad de Murcia.
Retrieved from <http://www.tdx.cat/handle/10803/10819>



- Rea, C. (2010). Getting on with corpus compilation: from theory to practice". *ESP World*, 1 (27).
- Rice, S. (2007). *Professional English in Use: Law*. Cambridge: Cambridge University Press.
- Schmitt, N. (2002). Using corpora to teach and assess vocabulary. Tan, M. (Ed.). *Corpus Studies in Language Education*. IELE Press.
- Scott, M. (2008). *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.
- Steinberger R., Pouliquen, R., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D., Varga, D. (2006). The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa, Italy, 24-26 May. JRC-Acquis corpus available from: <http://langtech.jrc.ec.europa.eu/JRC-Acquis.html>
- West, M. (1953). *A General Service List of English Words*. London: Longman.
- Yang, H. (1986). A new technique for identifying scientific/technical terms and describing science texts (an interim report)". *Literary and Linguistic Computing*, 1(2), 93-103.