A PROPOSAL TO EXPLOIT LEGAL TERM REPERTOIRES EXTRACTED AUTOMATICALLY FROM A LEGAL ENGLISH CORPUS

MARÍA JOSÉ MARÍN PÉREZ
Universidad de Murcia
mariajose.marin1@um.es

1. Introduction

The potential applications of specialised vocabulary inventories are manifold. They can be employed by linguists, translators or ESP (English for specific purposes) instructors as reliable sources of information for linguistic analysis, translation or language teaching. This research explores possible ways of exploiting a legal corpus and the vocabulary lists automatically extracted from it for the teaching of legal English terminology.

The role played by language corpora within the field of teaching ESL (English as a second language) and ESP is profusely discussed in the literature on the subject. On the whole, authors tend to favour their use as a learning tool or reference source acknowledging their advantages but also their limitations. McEnery and Xiao (2010: 364-365) accurately summarise the different areas of convergence between corpus linguistics and ESL:

The indirect use of corpora in teaching (reference publishing, materials development, and language testing), the direct use of corpora in teaching (teaching about, teaching to exploit, and exploiting to teach) and further teaching-oriented corpus development (languages for specific purposes (LSP)) corpora, first language (L1) developmental corpora and second language (L2) learner corpora).

Concerning the advantages of the use of corpora in language teaching, various scholars (Boulton 2012b; Hunston 2007; Johns 1986, 1991; McEnery and Wilson

1996; Sinclair 2003) envisage them as a highly motivating and valuable resource which exposes learners to genuine instances of language usage. Moreover, McEnery and Wilson (1996) and Boulton (2012b) underline their useful and current character as they "keep up with the changes as the terminology evolves within specific areas" (Boulton 2012b: 262) and suggest their use as a tool to review and update already existing teaching materials (McEnery and Wilson 1996).

According to Johns (1986, 1991), who coins the term *data-driven learning* (*DDL*), another advantage of the use of corpora in language teaching is their contribution to the development of learning strategies. By discovering the rules of the language underlying real samples, the students become "language detectives" (Johns 1997: 101) and learn how to learn. In Boulton's words, DDL methods contribute to develop learners' autonomy, that is, by handling and analysing corpora, learners "come to their own conclusions" (Boulton 2011: 563).

Conversely, Flowerdew (2009) criticises the predominantly inductive character of DDL methods which tend to offer decontextualised language samples extracted from corpora. She agrees with Swales (1990) that "corpus linguistics techniques encourage a more bottom-up rather than top-down processing of text in which truncated concordance lines are examined atomistically" (Flowerdew 2009: 395).

In spite of this criticism, the use of corpora is considerably widespread for the teaching of ESL and ESP. Nonetheless, as far as legal English is concerned, the scarcity of didactic materials based on legal corpora is manifest. Furthermore, as shown in Boulton's (2010)[1] review of over a hundred different empirical evaluations of DDL, carried out in the last twenty-five years, only two of them (Fan and Xun-feng 2002; Hafner and Candlin 2007) explore the field of legal English teaching.

Therefore, in order to try and bridge the methodological gap existing in the area, this study presents the proposal of four different corpus-based activities for the teaching of legal terminology. These activities have been developed using an 8.85 million-word legal corpus, *BLaRC* (the *British Law Report Corpus*), designed and compiled by the author, which is described in section two of this article. This section also presents a taxonomy for the classification of legal terminology using both qualitative and quantitative criteria. The third section offers a comparison between the list of terms identified in our legal corpus and the one produced from a corpus of legal English textbooks with the aim of demonstrating the relevance and usefulness of the former corpus as support material for teaching legal English vocabulary. A pedagogical research method is also suggested within this section for the future implementation of the activities proposed in it. Finally, section three ends with the actual proposal of the activities mentioned above followed by the conclusion to this study, in section four.

2. Corpus Description and Word Categories

2.1. *BLaRC:* A Legal Corpus of British Law Reports

In spite of the large number of authors who support the use of corpora as a useful ESP teaching and learning resource, the number and accessibility of legal corpora is small[2]. This is the reason why a legal 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 Sinclair (2005) for general corpora, and Pearson (1998), Rea (2010) for specialised ones.

BLaRC is an 8.85 million-word legal corpus of law reports, that is, written collections of judicial decisions made at British courts and tribunals. The reasons for focusing on this particular genre to study the linguistic properties of legal terminology are varied. To begin with, the UK belongs to the realm of common law. In common law systems, case law stands at their very basis relying on the principle of binding precedent for it to work, that is to say, a case tried at a higher court must be cited and applied whenever it is similar in its essence (the *ratio dicendi*) to the one being heard. Another fact that makes law reports outstanding within the legal field is that they not only cover all the branches of law, but they might also present fully embedded sections of other public and private law genres such as statutes, lease agreements, wills, deeds of property and the like, displaying therefore great lexical richness and variety.

As for its structure, *BLaRC* is a synchronic, monolingual and specialised collection of 1228 judicial decisions from the UK court and tribunal system issued between 2008 and 2010 in raw text format. Two elements conditioned our structuring of the corpus: the need to attend to the geographical origin of the texts under consideration and the need to abide by hierarchical criteria. The legal vocabulary varies according to which part of the UK has jurisdiction: 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[3].

Consequently, *BLaRC* was structured into five main sections depending on the jurisdictions of the British judicial systems, that is, the geographical scope of its courts and tribunals: 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 involved. By maintaining this structure, the

texts were grouped according to the field of law they belonged to (most courts and tribunals, with the exception of the Supreme Court, 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 would 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.

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 *Lextutor*[4], where it can be selected from 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 on Kilgarriff's *Sketch Engine*[5].

2.2. Word Categories according to their Meaning and Frequency

Applying specialised vocabulary inventories to the teaching of ESP terminology would require a clear classification of the lexical items included in them, since the distinction made in the literature on the subject between terms and non-terms is often blurred and overlapping. This is the reason why a taxonomy is presented below by which the vocabulary found in legal texts can be classified into different categories according to its meaning and frequency both in the specialised and general fields.

The pedagogical advantages of using a vocabulary taxonomy are related to the sequencing of corpus-based activities designed for the acquisition of legal vocabulary. Using this taxonomy, the ESP instructor can classify the terms obtained from a specialised corpus and grade the activities based on those terms according to their level of specialisation, their relevance both in the general and specialised fields, or the different senses a word may acquire in both contexts. Consequently, integrating those activities within the course syllabus will become an easier task for the instructor interested in incorporating DDL methodology as support to textbook-based teaching methods.

The distinction between highly specialised words in any subject field, that is, words which are exclusively employed in the specialised context on the one hand, and highly general ones on the other, such as the ones found in West's (1953) *General Service List of English Words* (*GSL*), appears to be clear-cut.

Nevertheless, there are words standing somewhere in between general and highly specialised vocabulary whose level of specialisation is hard to define, especially using quantitative criteria. A large number of Automatic Term Recognition methods (Chung 2003; Drouin 2003; Nazar and Cabré 2012; to name but a few)

are based on corpus comparison techniques for the extraction of the terms in a specialised corpus. They often employ such parameters as the frequency or distribution of those terms in both general and specialised corpora for the implementation of their algorithms. Therefore, attempting to quantify the level of specialisation of shared vocabulary, which can frequently be found in both contexts, is a complex task because of the statistical data involved.

From a pedagogical point of view, several authors (Baker 1988; Cowan 1974; Chung and Nation 2003; Flowerdew 1993; Wang and Nation 2004) express their concern about the difficulties caused by sub-technical words, that is, words which are employed both in the specialised and general fields, often acquiring a new technical meaning when in contact with the specialised context. Sub-technical vocabulary may have already been acquired as part of the learners' general vocabulary stock in which case it may have to be re-learnt as it often activates a new sense when in contact with the specialised context.

On the whole, authors tend to favour the use of the term *sub-technical* to refer to those types of words which are basically defined as shared vocabulary both by the general and the specialised fields or by scientific disciplines. Some authors also stress the relevance of the different senses of sub-technical words which acquire new meanings in technical areas. In addition, most of them underline their relevance in ESP instruction and the greater importance they must be given within the syllabus due to the fact that they might become an obstacle to the learners' acquisition of the vocabulary in any scientific field. Only Chung and Nation (2003) and Wang and Nation (2004) are more exhaustive as regards the delimitation of the semantic features of technical and sub-technical vocabulary in an attempt to analyse this lexical phenomenon from a different perspective, yet they do not employ the label *semi-* or *sub-technical*.

Thus, taking all these different perspectives into consideration and having observed a wide sample of highly specialised and sub-technical words taken from our legal corpus, a taxonomy is offered for the classification of legal vocabulary. This taxonomy answers to both quantitative and qualitative criteria, that is, it takes into consideration the frequency of usage of sub-technical words in the general[6] and specialised fields and also their meaning in both contexts. The words in bold correspond to the ones employed in the activities designed in the third section of this study. This proposal resembles Wang and Nation's (2004) as regards the semantic criteria employed in its design:

1) Words denoting a legal concept which are frequently used both in the general and specialised fields not changing their meaning in the legal context: *judge, court, tribunal, **law**, jury, legislation, robbery, theft, guilty, solicitor.*

2) Words often employed both in the general and specialised fields which change their meaning in the legal context sharing some semantic features with their original meaning: *charge*, ***offence***, *sentence*, ***claim***, *decision*, *grounds*, *complaint*, *dismiss*, *evidence*, *relief*, *record*, *trial*, *battery*.

3) Words occurring more frequently in the specialised field than in the general one. These words change their meaning in the specialised context, their new meaning being quite distant or completely unrelated to their general one: ***appeal***, *conviction*, ***party***, *warrant*, *terms*, *act*.

4) Highly technical words which appear almost exclusively in the legal field. If these words are employed in a general context, they also convey a legal concept in that field: ***breach***, *appellant*, *tortious*, *respondent*, *grantor*, *dicta*, *jurisprudence*, *tortfeasor*.

3. Direct Applications of the Term Lists Obtained From *BLaRC*

3.1. Term extraction and relevance of *BLaRC* term lists

BLaRC, our legal corpus, was processed using Drouin's (2003) automatic term recognition (ATR) method, *TermoStat*, with the aim of producing a reliable single word term (SWT) inventory which could be employed as reference for the design of vocabulary activities for the legal English class. A list of candidate terms having been obtained, it was compared with a legal glossary of 10,088 entries used as a gold standard to validate the method automatically. The method managed to identify 73% of true terms, on average finding its peak of precision[7] at 87% for the top 400. The resulting term list comprised 2,848 legal terms.

Before actually starting with the proposal of activities, we decided to compile a corpus using three legal English textbooks: *Professional English in Use: Law* (Brown and Rice 2007), *Introduction to International Legal English* (Krois-Linder and Firth 2008) and *Absolute Legal English* (Callanan and Edwards 2010), in an attempt to attest the usefulness and representativeness of the term lists obtained from *BLaRC* to be employed as support material to legal English textbooks. As Harwood affirms, corpora should "be used as a launch pad for classroom research into how the linguistic item in question is used by experts and students in the learners' local context" (Harwood 2005: 158).

The three textbooks mentioned above were selected owing to their comprehensive topic coverage, embracing a wide range of law areas, which guaranteed great lexical variety. Such variety also ensured that, if the overlap percentage found between the term lists obtained from *BLaRC* and the textbook corpus was high, the former would be representative not only of a single legal genre, but also of the whole language variety.

The first step of this comparison consisted in scanning and processing the textbooks using an OCR software. Then, the texts obtained, which contained 196,245 tokens, were stored in raw text format and processed with *Wordsmith 5.0* (Scott 2008) resulting into a type list[8] of 14,686 items that could be analysed and compared with the ones based on *BLaRC*, our legal corpus (the corpus obtained by scanning the three textbooks will be referred to as *LeG-TeXT* henceforth). We concentrated solely on SWTs to facilitate the comparison and the automatic search for concordance lines employing the *Concord* tool included in Scott's (2008) *Wordsmith's* package.

After extracting and validating the STWs in *LeG-TeXT* applying Drouin's (2003) ATR method (following the same steps as in the processing of *BLaRC*), it was attested that 67% of the SWTs identified were already present in the term lists obtained from our legal corpus using the same ATR method, a considerably high percentage taking into account the fact that the textbooks employed as reference deal with many different types of both private and public legal documents and topics apart from law reports. Furthermore, the documents used in the textbooks examined are usually adapted to fit into levels B2 to C1, to use the CEFR system, which makes such a high percentage of overlap even more relevant owing to the fact that *BLaRC* is made up of authentic language samples. Authentic legal texts are not adapted to suit students' needs but rather reflect real usage by legal practitioners. However, in this case, given the high overlap percentage found between our term list and the one obtained from the textbook corpus, our list could safely recommended as suitable for a B2 to C1 level course on legal English.

LeG-TeXT was also analysed with Heatley and Nation's (1996) software *Range* adapting our term list (the SWTs identified in *BLaRC*) to become a base-word list used as reference by the software (instead of employing the ones provided by default with the software programme from *GSL, AWL* or *BNC*[9]) with the purpose of establishing the percentage of running words in *LeG-TeXT* covered by our list. Surprisingly, the specialised terms found in *BLaRC* covered 12.37% of the running words in the textbook corpus, nearly three times as much as the expected percentage of text coverage established by Nation and Waring (1997) for specialised vocabulary.

According to Nation and Waring (1997), knowing the most frequent 2,000 words included in West's (1953) *GSL* enables us to understand approximately 80% of the words in any text. Nation (2001) classifies vocabulary into four different categories depending on their level of specialisation: general words, which provide ca. 80% of text coverage (or *text range*, as Nation puts it); academic words, included in Coxhead's (2000) *AWL,* which can cover around 10% of the words in any text;

technical words, which cover approximately 5% of the tokens in the corpus; and low frequency words, that is, those which do not fit into any of these categories, which would cover the remaining 5% of words.

Even so, the specialised terms in *BLaRC*, which would fit into Nation's category of technical words (5% predicted text range), covered almost three times as many words as might be expected according to Nation (2001). Probably, the representativeness of law reports within legal English, coupled with the fact that legal terminology is often employed outside the legal domain, can explain this finding. As a matter of fact, after processing the lists of terms identified in *BLaRC* with *Range,* almost half of the specialist vocabulary in those lists was present in West's (1953) *GSL* and Coxhead's (2000) *AWL* (40.47% of *BLaRC* term list) as well as amongst the most frequent 2,000 words in the *BNC* (45.41% of it).

The overall number of coinciding terms and the percentage of text coverage provided by the term inventory extracted from *BLaRC* explains why we decided to employ it as a source to design a set of activities for the teaching of legal terminology.

3.2. Didactic Exploitation of Term Inventories

Numerous authors (Boulton 2012a; Brodine 2001; Johns 1991; Landure 2013; Leech 1997; Rodgers et al. 2011, amongst many others) have carried out experiments using DDL methodology to plan and evaluate different types of activities which focus on diverse language levels and learning skills.

Following these experiments and owing to the scarcity of corpus-based proposals for the teaching of legal English terminology in the literature, as already stated, several activities were designed, which could be employed as a complement to other existing ESP teaching materials such as the textbooks mentioned above. The integration and sequencing of these activities within the course syllabus would depend on their relationship with the topics of the didactic units or content blocks, especially those activities focusing on the semantic and discursive levels of the language. For instance, the third and fourth tasks presented below, focusing on the various meanings of the terms *party* and *offence* and the concept of *claim* in the UK, could be included in a didactic unit entitled "Criminal justice and criminal proceedings"[10], given the relevance of these concepts within criminal law. Owing to their greater conceptual complexity of these activities and the fact that they integrate several language skills, they should be planned as final tasks within the unit for the revision and reinforcement of the contents studied in it.

As regards those activities devoted to the teaching of morphological and syntactic aspects of the language (activities 1 and 2 below): their planning could be more flexible due to the fact that they are form-focused. Nevertheless, it would be

recommendable to implement them earlier in the course since learning word formation mechanisms and simple syntactic patterns could contribute to the acquisition and production of more complex structures and patterns as the course progresses. As Schmitt asserts (2010: 36), "learning a word is incremental in nature, [which] means that vocabulary programs need to build recycling into the curriculum".

Prior to the implementation of these corpus-based tasks, the students should also be instructed on the use of concordancers[11] so as to be able to access easily the information requested from the corpus used as support for the legal English class (*BLaRC* in this case). They should learn how to generate concordances, identify collocates, sort the concordance lines depending on their preferences, apply stop lists whenever it was required, adjust the settings for the identification of collocates, and so on, so that the data provided by the corpus could be handled by them autonomously and exploited in as many ways as possible. This process could be time-consuming if not properly planned. Therefore, one session at the beginning of their language course should be given over to learning how to use corpus management tools properly. This would suffice for the carrying out of corpus-based activities during the whole course, as long as this type of tasks is carried out on a regular basis. In fact, using corpora as a means to access the language and to discover the rules governing it should not be a major problem for ESP students. As Boulton (2012b) illustrates through a survey of twenty experimental corpus-based studies focusing on ESP teaching, "overall, it seems that the participants do manage to deal with corpus data quite successfully" (2012b: 277).

3.2.1. Pedagogical Research Design

The activities presented below were conceived not as a substitute for already existing materials but rather as an option for the legal English practitioner interested in experimenting with DDL methodology. Consequently, they are adapted to suit the competence level established for the three legal English textbooks employed as reference, that is, CEFR level B2 to C1 (upper-intermediate to advanced level). The ESL learner who has fully achieved CEFR B2 or *Vantage* level, generally speaking:

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: 24)

The research method suggested for the evaluation of these activities would be a pre- and post-test design including a randomised group of fourth year Spanish students of translation, whose curriculum offers legal English as one of their compulsory subjects[12]. The students would be selected at random from amongst the members of the group. The use of initial pre-tests (they would be done at the beginning of the language course) would allow the instructor to divide the group in two/three different sub-groups depending on the results obtained. At that point of their academic training, fourth year translation students would be expected to have fully attained the language objectives established for CEFR B2 level. Nevertheless, owing to the fact that the members of the main group were selected at random, forming different level sub-groups would allow the instructor not only to compare the effects of the activity proposal but also to control the competence level variable. The initial pre-tests would include reading and listening comprehension activities as well as writing and speaking tasks which would be adapted to CEFR B2 level. If some of the individuals selected failed to pass the tests, they would be discarded from the control group, since they would not fulfil the requirements set for this experiment.

Once the main group was clearly organised into different competence level sub-groups, a second set of specific pre-tests would be given to them. In this case, they would be identical to the post-tests so that the effectiveness of the activities proposed could be measured (Hamp-Lyons 1985). The specific pre-tests would be administered at the beginning of each of the content blocks with which the activities below were associated. Each of the tests would assess the students' previous knowledge of the legal terminology presented in each of the activities. For instance, activity 2 consists in discovering the prepositional patterns associated with frequent legal terms such as *appeal*, *claim* or *breach*, therefore, as part of the specific pre-test, the students could be given multiple choice questions offering different items to link these terms with; they could be asked to correct phrases or sentences including these prepositional patterns; they could also be requested to fill in the gaps with the correct preposition in each case, and so on.

Subsequently, the post-tests could be carried out at the end of each of the didactic units or content blocks where the activities were included. They would be identical to the specific pre-tests, as already described. The results of the post-tests would thus be compared to the ones obtained with the pre-tests, allowing the instructor to quantify the sub-groups' average progress. Likewise, the results obtained by each of the sub-groups could also be compared as a way of controlling and measuring the competence level variable introduced after dividing the main group into different level sub-groups.

3.2.2. Activity Proposal

Activity 1

One of the activities which were planned to develop the learner's awareness of the morphological structure of legal terms consisted in asking them to guess what terms would stem from a list of the most relevant ones found in their textbooks, that is, asking them to try and form part of their word families (Bauer and Nation 1993). Thus, the students would necessarily have to explicitly reflect on the processes underlying the formation of words, which could also contribute to the development of strategies for the proper understanding and usage of the legal lexicon. As Bauer and Nation (1993: 253) observe, "... once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort".

Before starting with the activity, the instructor would necessarily have to make certain morphological rules explicit as regards the use of prefixes and suffixes putting special emphasis on typically legal ones such as *counter-*, *cross-*, *-ant*, etc. to facilitate and control the task.

Words like *appeal, claim* or *law* form other terms by derivation whose usage learners would have to attest through the search of concordance lines in *BLaRC*. These concordances would serve not only to confirm their guesses, but also to study the context of usage and meaning.

A follow-up activity could be suggested consisting in offering them different sentences (obtained from the corpus) with gaps in them. The students would have to decide which of the contexts would be more suitable for each of the members of a given word family, thus proving their understanding of the terms whose morphological structure they have reflected on. The concordances below, extracted from our legal corpus, illustrate the use of some of the legal terms belonging to the word families of:

APPEAL

Concordance

Notice. WAG was joined as a party. There has been no cross-appeal against the Commissioner's findings set out in
with Article 14 of the Convention. * Although there is no cross-appeal by the respondents on the issue of jurisdiction,
him, dismissed the originating applications and dismissed a cross-appeal by HMRC as the points that it sought to raise
to the wife. She was ordered to pay the costs of the cross-appeal. Discussion. The circumstances in which the
have gone back to the judge on this basis, as well as on the cross-appeal, for we are not in a position to make findings of
have no defence to Bocardo's claim. I would dismiss the cross-appeal. (b) Damages: 37. The parties are agreed that,

Concordance

name through a dormant company with the name of the Appellant, Reids Transport Minishant Limited, which was
the appeal against revocation of the operator licence of the Appellant company be DISMISSED. The order of the Traffic
the time of the public inquiry and of matters in favour of the Appellant company and Mr Laidlaw and Mr Fraser. He
to revoke the licence. 3. At the hearing of the appeal the Appellant company and Mr Laidlaw and Mr Fraser were
and A. FRASER & R. LAIDLAW. Attendances: For the Appellant: Tim Nesbitt of Counsel. Heard at Eagle Building,
Judge Frances Burton, Leslie Milliken, Stuart James. Appellant REIDS TRANSPORT MINISHANT LIMITED and A.
His Hon. Michael Brodrick, David Yeomans, Stuart James. Appellant: S & A BRUFORD TRANSPORT. Attendances: For

CLAIM

Concordance

contention was that, since Mr Morris had not made a counterclaim in the county court proceedings for repayment
to them that there was a County Court claim, met by a counterclaim. At a case management hearing on the 13th of
was pronounced in terms of the conclusion to the landlords' counterclaim. This was to the effect that the contract resulting
pleaded in para 18(1)(i) of the re-re-amended defence and counterclaim and are summarised in the agreed statement of
in para 18(1)(i) of TMT's re-re-amended defence and counterclaim by references to "sleeves". TMT pleads that both
* In para 18(1)(ii) of the re-re-amended defence and counterclaim TMT pleads that, in the context of the relevant

Concordance

not accepted at all, then the statutory rule requires that the claimant should pay the acquiring authority's costs from the
adopting the period during which, under CPR pt 36.11, a claimant may accept a pt 36 offer or payment without
order. It does seem to me, however, that where the claimant does accept the offer, and the tribunal has a
namely until the 3rd November 2004, on which date the claimant "declined" the sealed offer. I would award the
the offer, and the tribunal has a discretion as to costs, the claimant should be entitled to the benefit of the ordinary rule
period, but there would be little room for argument that the claimant should be entitled to any further costs thereafter."

LAW

Concordance

it followed that in determining the disturbance element it was lawful also to consider the open market value of the land itself.
by the withdrawal of support from land in connection with lawful coal-mining operations. A letter on costs accompanies
absence of planning permission or a certificate of existing lawful use would affect the minds of potential tenants in the
in 1975/76. (4) No planning permission. (5) No certificate of lawful use. (6) No guarantee planning will be granted. (7) No
The building had no planning permission or certificate of lawful use. It had no water supply, drainage, sewerage or
that the absence of planning permission or a certificate of lawful use for warehousing would not have affected the rent

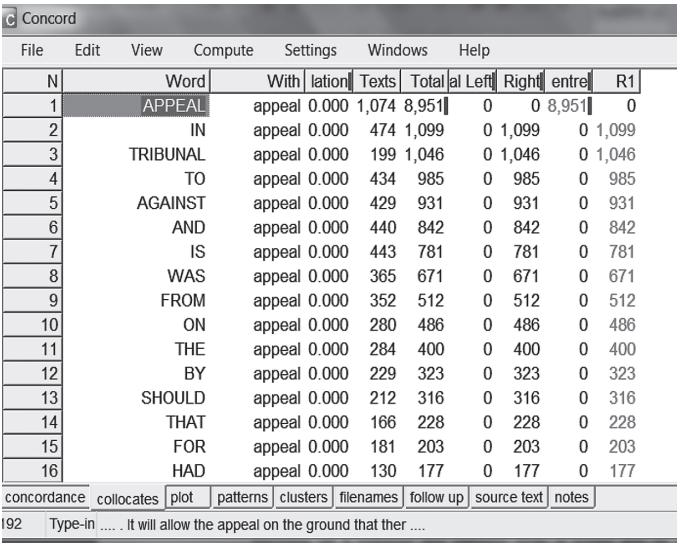
Concordance

I further guarantee that I will not indulge myself into any unlawful or illegal activity in the United Kingdom." 17. The
and the claimant ought to have been held the victim of unlawful discrimination by reference to his status as a
this context the word "lawful" means more than merely not unlawful but should be understood to connote the
consistently. The Home Office had treated his presence as unlawful and it would be inconsistent for the Department for
that the legislation could easily have made clear that even unlawful residence would be taken into account. H. Mr
different part of the criminal law, a person who commits an unlawful act towards another, with the necessary mens rea,

Activity 2

On a syntactic level, the study of lexical patterns is another potential application of specialised corpora within DDL. As pointed out by Lewis, "teachers should present words in the classroom in sequences whenever possible" (in Schmitt 2007: 754). In fact, lexical patterning has been made more accessible thanks to the use of corpora (Schmitt 2007), which clearly illustrate patterning constraints in real language use, hence the relevance of proposing this type of activities for the teaching of ESL and, in turn, ESP.

This activity would consist of asking learners to focus on the most frequent prepositions accompanying a set of legal terms such as *appeal*, *claim*, *right* or *breach* by examining a set of unfiltered concordance lines associated with them and concentrating on their collocate lists. They would be requested to study their main



Concord

File Edit View Compute Settings Windows Help

N	Word	With	lation	Texts	Total	al Left	Right	entre	R1
1	APPEAL	appeal	0.000	1,074	8,951	0	0	8,951	0
2	IN	appeal	0.000	474	1,099	0	1,099	0	1,099
3	TRIBUNAL	appeal	0.000	199	1,046	0	1,046	0	1,046
4	TO	appeal	0.000	434	985	0	985	0	985
5	AGAINST	appeal	0.000	429	931	0	931	0	931
6	AND	appeal	0.000	440	842	0	842	0	842
7	IS	appeal	0.000	443	781	0	781	0	781
8	WAS	appeal	0.000	365	671	0	671	0	671
9	FROM	appeal	0.000	352	512	0	512	0	512
10	ON	appeal	0.000	280	486	0	486	0	486
11	THE	appeal	0.000	284	400	0	400	0	400
12	BY	appeal	0.000	229	323	0	323	0	323
13	SHOULD	appeal	0.000	212	316	0	316	0	316
14	THAT	appeal	0.000	166	228	0	228	0	228
15	FOR	appeal	0.000	181	203	0	203	0	203
16	HAD	appeal	0.000	130	177	0	177	0	177

concordance collocates plot patterns clusters filenames follow up source text notes

192 Type-in .... . It will allow the appeal on the ground that ther ....

FIGURE 1. List of immediate right collocates attracted by appeal.

collocates with the aim of identifying those prepositions which the concordancer would present as their most relevant functional collocates. In order for the activity to accomplish its goal, only the immediate right collocates would be considered. After doing so, they would be offered different examples extracted from *BLaRC* to fill in the gaps with the appropriate prepositions to guarantee the validity of their observations. Following a similar process to activity 1, the students could assess their own performance by consulting a set of concordance lines previously selected and filtered by the instructor, which would act as feedback to their answers.

The figure below illustrates the lists of the main collocates generated by *appeal* in *BLaRC* using *Wordsmith 5.0* (Scott 2008).

Activity 3

Focusing on a semantic level, it would be recommendable to study the contexts of usage of sub-technical words, which characterise legal language and partly explain the great percentage of shared vocabulary between the legal and general fields. Sub-technical words frequently specialise when in contact with the legal

PARTY	
MEANINGS *(OED)*	CONCORDANCES
1. A social gathering especially for pleasure or amusement	… if I were to plan a big party, or an anniversary or something like that, and I'd hope those would be jolly enjoyable days too …(*LACELL*)
2. An established political group organized to promote and support its principles and candidates for public office	I witnessed this over and over again, until I decided to join the Green Party in 1993 (*LACELL*)
3. A person or group involved in a legal proceeding as a litigant	Each party shall pay its own costs in respect of the issue of costs (*BLaRC*)
OFFENCE	
MEANINGS *(OED)*	CONCORDANCES
1. Annoyance, displeasure, or resentment	… I didn't take offence at the question. I think it was a perfectly fair question. … (*LACELL*)
2. A violation or breach of a law, custom, rule, a crime	The Director may withdraw or restrict access to the facilities in response to an offence or a suspected offence against these rules or to protect the services (*BLaRC*)

TABLE 1. General and specialised meanings of party and offence and concordance line examples

environment causing confusion for ESP learners who might already have acquired them as part of their general vocabulary.

The students would be given a list of these words taken from the corpus. They would be asked to match a set of selected concordance lines with the different senses of those words taken from both a general and specialised dictionary: *The Oxford English Dictionary* (2002) and *Dahl's Law Dictionary* (Saint-Dahl 1999). For this activity, they should make use of a general corpus with the aim of identifying the general meanings of the words given. There is a plethora of options: for instance, they could access other general corpora either stored on their computers or offered online. Prof. Mark Davies' website[13] provides online access to varied general English corpora which could serve this purpose. Table 1 illustrates the most frequent senses of the sub-technical terms *party* and *offence* selected from amongst the ones defined in the *OED* (2002) and the concordances obtained from *BLaRC*, our legal corpus, and *LACELL*, the general one.

Activity 4

To finish with this section devoted to the didactic exploitation of corpus data, a content-focused activity is presented. Using the multi-word term (MWT) list generated by Nazar and Cabré's (2012) online ATR software, *Terminus 2.0,* a number of these items was selected with the purpose of helping the students understand and acquire, for example, such a concept as "types of claim" in the UK. In order to do this, they would have to delimit the concept clearly by differentiating the categories or types it comprised. This could be achieved by providing them with concordance examples with gaps which they could fill by using the compound term list provided. Resorting to a specialised dictionary would be recommendable as support for this task. In addition, they would probably have to consult the co-text of the concordances for a fuller understanding of the examples.

Error correction could also be suggested as a follow-up activity. The students would be provided with a set of sentences which they would have to amend by employing the appropriate compound term in each case. Through this task, the students would confirm the conclusions they may have reached thanks to the observation of the concordance lines provided initially.

Owing to the greater linguistic and conceptual complexity of this activity, pair-work is to be recommended as a way of lightening the load for the student. In fact, collaborative learning is highly motivating and promotes linguistic accuracy (Fernández 2012; Storch 1999), apart from providing greater chances of using the L2.

Table 2 shows the MWTs associated with *claim* and identified by *Terminus*[14] (Nazar and Cabré 2012). They have been arranged according to the level of specialisation calculated by the ATR method applied. Further below, table 3 presents some of the concordances for these compound terms found in *BLaRC*.

CLAIM TYPES (MWTs)	SPECIALISATION LEVEL (*Terminus*)
FRESH CLAIM	2767.60
POSSESSION CLAIM	1480.37
DERIVATIVE CLAIM	1426.45
UNFAIR DISMISSAL CLAIM	987.766
BATTERY CLAIM	939.60
PROPRIETARY CLAIM	597.777
CIVIL CLAIM	587.884
ILLIQUID CLAIM	456.557
RENEWAL CLAIM	414.608
EQUAL PAY CLAIM	407.413
PERSONAL INJURY CLAIM	380.729

TABLE 2. MWTs associated with claim as ranked by Nazar and Cabré's (2012) Terminus

CONCORDANCE LINES FROM *BLaRC*
… I don't believe it and therefore I am not going to regard it as a **fresh claim** …
… he could review the Council's decision to bring and maintain the **possession claim** on normal judicial review principles …
… there were special circumstances justifying the **derivative claim** which he seeks to bring …
… if he had been of the view that the unfair **dismissal claim** had not been made in time …
… prosecution in a civil court of the **assault and battery claim** would be "manifestly unfair" to him
… her family was to ensure that the husband acquired no **proprietary claim** to shares in the wife's family …
… It is not necessarily an abuse to proceed with a **civil claim** in tort against a defendant who …
… that the rule that an illiquid claim cannot be pleaded by way of compensation to a **liquid claim** …
… the applicant's own evidence in his **renewal claim** form to benefit where the phrase 'emotional support' was used …
… in what circumstances, if any, can a claimant in an **equal pay claim** show that she is in the same

TABLE 3. Concordance lines illustrating the MWTs generated by claim

To conclude, a final project is proposed to foster and assess the acquisition of this set of compound terms. The students would be given a home assignment consisting in writing an essay which would discuss the concept of *claim* and *claim types* in UK law. A number of questions would be formulated acting as cues for the planning and development of the task, for instance:

— "Who are the participants which take part in the development of a claim in the UK?
— Is this process similar to the one followed in your country?
— What are the most frequent claim types in the UK?
— Do they belong to the criminal or the civil fields?
— Do you know any relevant case belonging to any of these categories? Could you explain it?"

4. Conclusion

This article has presented a proposal to employ a specialised corpus based on law reports and the term lists obtained automatically from it as support material for the teaching and learning of legal English terminology. The formulation of this proposal is the result of the methodological void in the area where only a few experiments have been carried out employing DDL methodology. The pedagogical relevance of the use of DDL activities in SL instruction lies in the role which the learner adopts in the learning process, becoming a central part and an active participant within it. By examining language samples extracted from corpora, learners become researchers themselves trying to solve tasks as if they were

language detectives, to use Johns' terms (1997). In fact, this kind of task promotes linguistic awareness as well as learning autonomy, which are key to success in language learning (Boulton 2011). Studies reveal that once learners have become familiar with corpus linguistics analysis tools, DDL appears to be particularly effective, for instance, in the acquisition and use of specialised vocabulary in context (Lin 2008). Authors like Kaur and Hegelheimer (2005) attest that the experimental group used in their evaluation of DDL methods "used the target items more frequently and more accurately in writing" (Boulton 2011: 278).This is the reason why the use of corpus-based activities could become a useful complement to already existing materials, positively contributing to the acquisition of specialised terminology.

The second section of this article describes the design criteria and features of *BLaRC,* the legal corpus designed by the author and employed as the source to obtain the necessary information from.This section also presents a legal vocabulary taxonomy aimed at facilitating the integration of specialised vocabulary activities within the ESP course syllabus.

In the third section, after reflecting on the subject of DDL and justifying the use of our corpus and specialised vocabulary inventories, some activities are suggested to complement the ones offered in three different legal English textbooks used as references. They focused on several linguistic levels, namely, morphological (concentrating on derivational processes for word formation), syntactic (grammatical patterns associated with certain legal terms), semantic (study of polysemic terms) and discursive (proposal of a written project).

To conclude, as further research, a pedagogical research method is suggested with the aim of implementing these activities in the future and testing the resultant learning outcomes.

Notes

[1]. A supplement to this reference with a detailed explanation of all the experiments can be consulted at: http://corpuscall.eu/file.php/5/0_DDL_empirical_survey_2012_July.pdf

[2]. See Author and Rea (2012) for a fuller review of the amount and availability of legal corpora.

[3]. For further information, see:

— The website of the Judiciary of England and Wales: http://www.judiciary.gov.uk/about-the-judiciary/introduction-to-justice-system/court-structure

— Scottish court and tribunal system: http://www.scotland.gov.uk/Topics/Justice/legal

— Courts and tribunals in Northern Ireland: http://www.courtsni.gov.uk/en-GB/pages/default.aspx

[4]. http://www.lextutor.ca/concordancers/concord_e.html

[5]. https://the.sketchengine.co.uk/login

[6]. The general corpus employed for the observation of the frequency of the words included in this study is *LACELL,* a 21 million-word general English corpus designed and compiled by the *LACELL* research group of the English Department at the University of Murcia. This corpus is not publicly available.

[7]. Precision is calculated by determining the percentage of true terms (terms found in the gold standard) identified by an ATR method with respect to the whole list of candidate terms extracted by it.

[8]. The term *type* refers to any word identified by the software regardless of the number of times it occurs in the corpus, every time it repeats itself it is referred to as a *token.*

[9]. *British National Corpus*

[10]. This is the title of unit 5 of Brown and Rice's (2007) *Professional English in Use: Law.*

[11]. The concordancer employed in this case has been the *Concord* tool included in Scott's (2008) *Wordsmith 5.0.*

[12]. http://www.um.es/web/letras/contenido/estudios/grados/traduccion-ingles/plan/asignaturas

[13]. http://corpus.byu.edu/

[14]. Although Drouin's software allows the user to identify MWTs as well, the precision levels achieved after testing by Nazar and Cabré's software were higher (35.86% and71.5% on average respectively) in the automatic recognition of compound terms. The latter was therefore selected to perform this task.

Works Cited

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.

Bauer, Laurie and Paul Nation. 1993. "Word Families". *International Journal of Lexicography* 6 (4): 253-279.

Boulton, Alex. 2010. "Learning outcomes from corpus consultation". In Moreno Jaén, M., F. Serrano Valverde and M. Calzada Pérez (eds.) *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*. London: Equinox: 129-144.

—. 2011. "Data Driven Learning: the Perpetual Enigma". In Roszkowski, S. and B. Lewandowska-Tomaszczyk (eds.) *Explorations across Languages and Corpora*. Frankfurt: Peter Lang: 563-580.

—. 2012a. "Beyond concordancing: Multiple affordances of corpora in university language degrees". In Macaire, D. and A. Boulton(eds) *Languages, Cultures and Virtual Communities. Procedia: Social and Behavioral Sciences* 34: 33-38. <http://www.sciencedirect.com/science/article/pii/S1877042812003138> Accessed 31 December, 2013.

—. 2012b. "Corpus consultation for ESP. A review of empirical research". In Boulton, A., S. Carter-Thomas and E. Rowley-Jolivet (eds.) *Corpus-Informed Research and Learning in ESP*. Issues and Applications. John Benjamins Publishing Company: 261-292.

Brodine, Ruey. 2001. "Integrating Corpus Work into an Academic Reading Course". In Aston, G. (ed.) *Learning with Corpora*. Bologna: Athelstan: 333-344.

Brown, Gillian and Sally Rice 2007. *Professional English in Use: Law.* Cambridge: Cambridge U.P.

Callanan, Hellen and Lynda Edwards. 2010. *Absolute Legal English.* London: Delta.

Chung, Teresa Miwha. 2003. "A corpus comparison approach for terminology extraction". *Terminology* 9 (2): 221-246.

Chung, Teresa Miwha and Paul Nation. 2003. "Technical Vocabulary in Specialised Texts". *Reading in a Foreign Language* 15 (2): 103-116.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge U.P. <http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/CEFR_EN.pdf> Accessed 30 December, 2013.

Cowan, J. Ronayne. 1974. "Lexical and syntactic research for the design of EFL". *TESOL Quarterly* 8: 389-99.

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

Drouin, Patrick. 2003. "Term extraction using non-technical corpora as a point of leverage". *Terminology* 9 (1): 99-117.

Fan, May and Xu Xun-Feng. 2002. "An evaluation of an online bilingual corpus for the self-learning of legal English". *System* 30(1): 47-63.

Fernández, Ana. 2012. "Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work". *Journal of Second Language Writing* 21: 40-58.

Flowerdew, John. 1993. "Concordancing as a tool in course design". *System* 21 (2): 231-244.

Flowerdew, Lynne. 2009. "Applying corpus linguistics to pedagogy: A critical evaluation". *International Journal of Corpus Linguistics* 14 (3): 393-417.

Gilquin, Gaëtanelle and Sylvianne Granger. 2010. "How can data-driven learning be used in language teaching?". In O'Keefe, A. and M. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics.* London: Routledge: 359-370.

Hafner Christoph and Christopher Candlin. 2007. "Corpus tools as an affordance to learning in professional legal education". *Journal of English for Academic Purposes* 6(4): 303-318.

Hamp-Lyons, Liz. 1985. "Two Approaches to Teaching Reading: A Classroom-Based Study". *Reading in a Foreign Language* 3 (1): 363-73.

Harwood, Nigel. 2005. "What do we want EAP teaching materials for?". *Journal of English for Academic Purposes* 4: 149-161.

Hunston, Susan. 2007. *Corpora in Applied Linguistics.* Cambridge: Cambridge U.P.

Johns, Tim. 1986. "Microconcord: A language-learner's research tool". *System* 14 (2): 151-162.

—. 1991. "Should you be persuaded: Two examples of data-driven learning". *English Language Research Journal* 4: 1-16.

—. 1997. "Contexts: The background, development and trialling of a concordance-based CALL program". In Wichmann, A., S. Figelston, T. McEnery and G. Knowles (eds.) *Teaching and Language Corpora.* London: Longman: 100-115.

Kaur, Jagdish and Volker Hegelmeier. 2005. "ESL students' use of concordance in the transfer of academic world knowledge: An exploratory study". *Computer Assisted Language Learning* 18 (4): 287-310.

Krois-Linder, Amy and Matt Firth. 2008. *Introduction to International Legal English: A course for classroom or self-study use.* Cambridge: Cambridge U.P.

Leech, Geoffrey. 1997. "Teaching and language corpora: A convergence". In Wichmann, A., S. Fligelstone, T. McEnery and G. Knowles (eds.) *Teaching and language corpora.* London: Longman: 1-23.

Landure, Corinne. 2013. "Corpus et dictionnaires: prise de conscience linguistique chez des apprenants LANSAD". *Bulletin VALS-ASLA* 97: 59-76.

Lin, Ming Chia. 2008. "Building a lexical syllabus on Moodle with web concordancers for EFL productive academic vocabulary". In *Proceedings of WordCALL 2008.* Fukuoka: Fukuoka University.

McEnery, Tony and Andrew Wilson. 1996. *Corpus Linguistics.* Edinburgh: Edinburgh U.P.

McEnery, Tony and Richard Xiao. 2010. "What corpora can offer in language teaching and learning". In Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning.* London: Routledge: 364-380.

Nation, Paul. 2001. *Learning vocabulary in another language.* Cambridge: Cambridge U.P.

Nation, Paul and Robert Waring. 1997. "Vocabulary Size, Text Coverage and Word Lists". In Schmitt, N. and M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge U.P.: 6-19.

Nazar, Rogelio and María Teresa Cabré. 2012. "Supervised Learning Algorithms Applied to Terminology Extraction". In Aguado de Cea, G., M.C. Suárez-Figueroa, R. García-Castro and 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.

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): 1-23.

Rodgers, Ornaith, Angela Chambers and Florence Le Baron-Earle. 2011. "Corpora in the LSP classroom: A learner-centred corpus of French for biotechnologists". *International Journal of Corpus Linguistics* 16 (3): 391-411.

Saint Dahl, Henry. 1999. *Dahl's Law Dictionary.* New York: William S. Hein and Co. Inc.

Scott, Mike. 2008. *WordSmith Tools version 5* [computer software]. Liverpool: Lexical Analysis Software.

Schmitt, Norbert. 2007. "Current perspectives on vocabulary teaching and learning". In Cummins, J. and C. Davison (eds.) *The International Handbook of English Language Teaching, Vol.* New York: Springer: 745-759.

—. 2010. "Key issues in teaching and learning vocabulary". In Chacón-Beltrán, R., C. Abello-Contesse and M. Torreblanca-López (eds.) *Insights into Non-native Vocabulary Teaching and Learning*. Bristol: Multilingual Matters: 28-40.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford U.P.

—. 2003. *Reading Concordances: An Introduction.* London: Longman.

—. 2005. "Corpus and Text: Basic Principles". In Wynne, M (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford University: AHDS Literature, Languages and Linguistics.<http://ota.ahds.ac.uk/documents/creating/dlc/index.htm> Accessed 27 September 2013.

Storch, Neomy. 1999. "Are two heads better than one? Pair work and grammatical accuracy". *System* 27: 363-374.

Swales, John. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge U.P.

Vargas, Chelo. 2005. *Aproximación terminográfica al lenguaje de la piedra natural: propuesta de sistematización para la elaboración de un diccionario traductológico*. Doctoral thesis. Universidad de Alicante.

Wang, Karen, and 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.

Received: 27 September 2013
Accepted: 2 February 2014