STRUCTURE AND DESIGN OF THE BRITISH LAW REPORT CORPUS (BLRC): A LEGAL CORPUS OF JUDICIAL DECISIONS FROM THE UK

Mª José Marín Pérez
Camino Rea Rizzo
University of Murcia

ABSTRACT. *The aim of this paper is to describe and justify the structure and design criteria to create a legal English corpus of judicial decisions. The authors, lecturers of this ESP variety, decided to engage into specific corpus design due to the small variety of teaching materials and corpora available. Judicial decisions are essential wheels in the legal machinery of common law systems and, precisely because of that fact, they are fundamental as a legal genre. This is why we intend to compile a 6m word legal corpus of UK judicial decisions in order to establish the core vocabulary of the genre and use it for further linguistic analysis and the elaboration of didactic materials.*

Keywords: Corpus linguistics, law reports, ESP, representativeness, common law.

María José Marín Pérez / Camino Rea Rizzo

ESTRUCTURA Y DISEÑO DEL BLRC: UN CORPUS DE INGLÉS JURÍDICO BASADO EN SENTENCIAS JUDICIALES

RESUMEN. *La finalidad de este artículo es describir y justificar la estructura y los criterios de diseño para la creación de un corpus de inglés jurídico basado en sentencias judiciales. Las autoras, profesoras de esta variedad de ESP, decidimos involucrarnos en el diseño de un corpus específico debido a la escasez de materiales y corpus existentes. Las sentencias judiciales son una pieza fundamental en los sistemas judiciales common law y, precisamente por este motivo, son esenciales como género dentro del inglés jurídico. Así pues, pretendemos compilar un corpus de sentencias judiciales del Reino Unido de 6 millones de palabras para así establecer el vocabulario básico de éstas y utilizarlo también para posteriores análisis lingüísticos y elaboración de materiales didácticos.*

Palabras clave: Lingüística del corpus, sentencias judiciales, ESP, representatividad, *common law*.

1. INTRODUCTION[1]

The implementation of the Bologna Reform has brought about a substantial change in the status of English as a subject in Higher Education programmes excepting degrees in English studies and Translation. The new European Higher Education system aims to qualify graduates for professional competences among which the mastering of a second language, particularly English, is a must. In fact, the European Commission and the European Language Council actively promote language learning through the development and implementation of language policies, according to which all future graduates in Europe should be able to communicate in at least two languages other than their mother tongue (Räisänen and Fortanet 2008: 21).

The restructuring of university programmes according to European language standards entailed the learning of English as a transversal competence. Its presence in the current university programmes has resulted from the choice between two possible ways of integration: the adoption of English as the language of instruction in a considerable part of some compulsory subjects, or the offer of English for specific purposes (ESP) courses, as a separate subject

---

[1] A preliminary version of this proposal was presented at the III Congreso Internacional de Lingüística de Corpus, held at the Universidad Politécnica de Valencia (7-9th April, 2011).

independently of content courses. The latter is the case of Legal English incorporated into the degree in Law at the Law Faculty of the University of Murcia which the authors have been and will be in charge of teaching.

It was a hard task to decide on teaching materials when first facing the subject. Legal English is a particularly obscure variety of ESP. Jonathan Swift would state in *Gulliver's Travels* as early as 1726 that it is "a peculiar Cant and Jargon of their own, that no other Mortal can understand" (Mellinkoff 1963: 5). In addition to this, the amount of available materials, especially text books, was considerably scarce, as usually happens in other branches of ESP (Rea 2010). Resorting to specific corpora could have been an option:

> Such corpora can be used to provide many kinds of domain-specific material for language learning, including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora (McEnery and Wilson 1996: 121).

Nevertheless, to our knowledge, the amount of written legal corpora is also reduced, and access to them, except for a few cases, is not complete. As a consequence of the scarce amount of such corpora and the methodological void derived from it, we engaged into ESP corpus design and decided to create the British Law Report Corpus (BLRC): a legal English corpus of judicial decisions that could act as a reliable source of specific vocabulary for the development of new teaching materials, and information for further language analysis.

The aim of this paper is to present the process of design and compilation of the BLRC, according to Corpus Linguistics standards (Wynne 2005) for general corpora and its adaptation to specific corpora (Rea 2010). First, the legal corpora found are introduced; next, we give a detailed account of the design process and justify the reasons that lead to the selection of this legal genre, the mode of the texts, the organization of the corpus into different categories, and the distribution of texts per category; to finish with some final remarks on further corpus applications and future research.

2. THE BLRC AND OTHER LEGAL CORPORA

The research into specific corpora availability led us to a short list of legal corpora which did not satisfy our needs. The first corpus worth mentioning is the Bononia Legal Corpus (Rossini, *et al.* 2001), since this is probably the most comprehensive legal corpus existing due to its selection of texts from varied genres and topics, and also the closest to ours especially regarding the genres it covers. It was "set up with representative documents of a legislative, judicial

and administrative nature" –as stated on their website–. It is a multilingual comparable Italian-English corpus, which aims at "representing the two different legal systems, in particular the differences between the civil law and the common law systems" –on the website –. Its English section, whose initial target was 10m words, covers several legal genres, namely, UK statutes, law reports and statutory instruments. It can be freely accessed through the internet but not downloaded.

However, the rest of corpora we have found were either too small to act as a normative reference for us, or inaccessible. They focused on aspects of the language we are not interested in, or were conceived as parallel corpora with a translational or comparative purpose. The JRC-Acquis Corpus is one of them. It is a multilingual parallel corpus which includes European Union legislative texts affecting all Members States in 22 different languages. The English section contains 23,545 texts and 34,588,383 words. It is fully accessible and downloadable.

Next, the CorTec corpus is a scientific-technical parallel corpus divided into four sections, one of them deals with commercial law and includes agreements and contracts in English and Brazilian Portuguese. It has 1m words per section and has been developed by the Translation and Terminology Centre of the University of St. Paul, Brazil.

As for the HOLJ corpus, it is a monolingual synchronic one comprising 188 judgments of the House of Lords from 2001 to 2003. The number of words is approximately 3,000,000 and its aim is to define a set of rhetorical role labels.

To finish, the Cambridge International Corpus, owned by Cambridge University Press, has a legal corpus section of 20m words. It is not accessible or commercialised. It has been employed by CUP to design their legal English books.

There also exist legal sections or materials within some of the best known general British English corpora like the BNC or the COBUILD, but they could not serve our purpose either as they are non-specific.

3. LAW REPORTS AND THEIR ROLE IN COMMON-LAW-BASED LEGAL SYSTEMS

Establishing the *sampling frame,* that is, "the entire population of texts from which we [would] take our samples" as McEnery and Wilson put it (1996: 78), was our first objective, and law reports were selected due to the pivotal role they play in the UK judicial system as well as in any other common law countries. We are aware that the conclusions drawn from the study of one single genre

cannot be extrapolated to the whole variety. However, judicial decisions or judgements – as reflected on law reports – stand at the very core of common law systems acting as the main source of law followed by statutes and equity, and thus hold a prominent position in legal ESP.

If representativeness is crucial for the design of any corpus (Sinclair 1991; Biber 1993; Sánchez *et al*. 1995; McEnery and Wilson 1996; Tognini-Bonelli 2001; Hunston 2002; Wynne 2005, etc.), narrowing the boundaries of our object of study became a must, as we soon realised how legal language is intertwined with everyday language, how it is present both in the public and private fields, and consequently how the vastness of this ESP branch could not be covered or managed in a project of this nature.

Orts (2006, 2009) offers a comprehensive review of different approaches to *legalese* and legal genres both from the field of law (Mellinkoff 1963; Jackson 1985; Tiersma 1999; etc.) and linguistics (Crystal and Davy 1969; Danet 1980; Swales 1985; Kurzon 1984; Maley 1987; Alcaraz 1994; Bathia 1993, 2004; etc.). The number of legal genres authors have identified varies depending on the perspective of their analysis, and law reports, that is, written reports of judicial decisions or judgments, appear in generic classifications as part of the oral mode (Danet 1980), within the category: *recording and law making* (Maley 1994), or as public unenacted law (Orts 2009), amongst others.

Sinclair states that "the contents of the corpus should be selected […] according to their communicative function in the community in which they arise" (2005: 1), so it appeared to us that the selection of law reports as our object of study was fully justified since they are an essential wheel in the British legal machinery, and their status within it is unquestionable.

The United Kingdom belongs to the realm of common law, as opposed to civil or continental law. Western European law, except for the UK, is based on the civil law system. Although it may refer to and apply the existing jurisprudence, it mostly relies upon the law pertaining to the criminal or civil fields -amongst others- which is codified following the Roman law tradition. On the other hand, in common law countries like USA, Canada, Australia, and so forth, and specifically in the UK, law decisions were based on previous cases always abiding by the principle of *stare decisis* – to stand by what has previously been decided –, and not on acts passed at the parliament.

However, common law systems have evolved in different ways: some of them are mixed like Québec or Scotland, where the law is both codified and uncodified. The majority of them is mostly jurisprudential and complies with the principle of binding precedent, that is to say, the decisions made at a higher tribunal should act as binding precedent as long as they are related to the case

in question in their essence. Determining what the essence of a given case is – establishing the *ratio dicendi* – is part of the judge's role.

Nonetheless, in purely common law systems, the acts passed at their parliaments have gained greater importance being most often cited in case decisions. In the last 150 years (Orts 2006), enacted law has become essential as a source of law, albeit judicial decisions, as far as they interpret the law and the existing precedents, stand out as the major one.

Another fact that makes law reports an outstanding genre in common law legal systems is that they not only cover all the branches of law, but also touch upon other genres like statutes, or others like wills, contracts, and so forth, when such text types are referred to as facts, evidence, or any other section within the judicial decision.

Law reports are written collections of judicial decisions on cases that solicitors, barristers, judges, or any other legal professionals need to know. They must be cited and act as the solid ground on which they will build their arguments. This is why, in those systems based on case law, case reports are published by councils, that is, The Incorporated Council of Law Reports of England and Wales (ICLR), publishing houses like Butterworth or Lloyds, and so forth, every year. Due to the widespread use of information technologies and particularly the internet, there is a tendency towards digitalising these texts and storing them in online databases. Using search engines can make case citation a really easy task that used to take ages for legal practitioners to become fully informed about the existing jurisprudence.

Access to most of these databases is often restricted; there are such popular ones as Justis.com, LexisNexis.com, and so forth – they are really expensive due to the amount of time they save, so law firms, law faculties, and the like are subscribed users precisely because of that –. They offer different possibilities to the legal practitioner to locate and cite cases depending on the court they were heard at, their main topic, the judges who heard them, the identity of the parties, and so forth. However, the British and Irish Legal Information Institute (BAILII.org) has created a completely free and comprehensive online database with more than 200,000 cases available and classified them according to the court where they originated and the jurisdiction they belong to.

Although we are not subscribed users of the databases mentioned above, we have enjoyed free access to some of them and confirmed the fact that, leaving aside the numerous possibilities and applications they provide to legal practitioners, they offer a smaller amount of texts than the free-access BAILII database. BAILII has become a really useful and free source of not only case decisions, but also statutes and some scientific legal texts. It is supported by a number of sponsors like the Inns of Court –barristers' professional associations–,

law faculties –Cambridge, Oxford, Glasgow, Edinburgh, Cork, etc.– , law firms and other prestigious institutions, hence its importance and recognition by professionals. This is precisely why we have decided to use it as our main source to obtain the legal texts that form the BLRC.

4. CRITERIA APPLIED IN BLRC COMPILATION

4.1. GEOGRAPHIC CRITERIA AFFECTING THE SELECTION OF TEXTS

As well as abiding by hierarchical criteria when organizing the corpus, one of the first elements that conditioned our choice was the way that legal vocabulary varies according to the system where it is used. This is so because of the laws and regulations that organise the countries which the UK is divided into. 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. But for the Supreme Court –in general terms– and the UK Tribunal Service –except for some cases–, each country is fully independent as regards its judicial system. Therefore, we decided to structure the BLRC into five main branches depending on the jurisdictions of their judicial systems, that is, the geographical scope of their courts and tribunals: Commonwealth countries, United Kingdom, England and Wales, Northern Ireland, and Scotland.

Special attention is deserved by the first section, that of Commonwealth countries. The Judicial Committee of the Privy Council is a UK institution whose main role is acting as the "highest court of appeal for many current and former Commonwealth countries, as well as the UK's overseas territories, crown dependencies, and military sovereign base areas" –as stated on their website–. The cases heard at this court may come from such varied origins as Mauritius, Jamaica, Trinidad and Tobago, and so forth. Since such geographical variation necessarily implies terminological changes due to their different legal systems, we considered it interesting to devote one of the sections of the corpus to the texts coming from such varied sources, in spite of them not being too numerous.

As regards the second section, it comprises those institutions which are competent to judge cases from all over the UK –with certain exceptions–. This category includes the court of last resort of Great Britain: The Supreme Court as well as the net of administrative courts.

The other three sections are organised in the same way as their judicial systems, that is, except for England and Wales who share the same structure and laws, the justice of Northern Ireland and Scotland work independently from the other two but for the net of administrative tribunals –excepting some cases– , and the Supreme Court.

4.2. CHRONOLOGY

The BLRC is a specific synchronic monolingual corpus of legal English texts which aims at establishing the core vocabulary of law reports in the UK. Following Pearson, "a specific corpus compiled for terminological studies [...] delivered in the last 10 years prior to the date of compilation" (1998: 51). This is why we decided to compile the texts produced at UK courts and tribunals from 2008 to 2010 as we expect to finish gathering them before the end of 2011. The texts were always gathered randomly yet always belonging to the time span just mentioned, and we will not include any that was added to the database after December 2010.

Moreover, due to the changes that the structure of these courts has experienced for the recent modifications of the law that regulates it, we considered that, if the structure of the corpus responds to the structure of UK courts and tribunals because of thematic and hierarchical reasons, it should adjust to the latest modifications it has experimented.

We are specifically referring to the *Constitutional Reform Act, 2005*, by which the Supreme Court of the UK was created and started to work on 1 October 2009 –its role was formerly performed by the so-called Law Lords of the House of Lords–, and the *Tribunals, Courts and Enforcement Act, 2007* which regulates the structure of these institutions thus affecting the structure of the BLRC itself.

4.3. MODE

The mode of the texts included in the BLRC is written. Initially, the corpus is not intended to be tagged so the written samples are stored as raw text. Obtaining oral samples of legal language that reflected the arguments, facts, and decisions dealt with at court, would have implied having access to courtrooms and permission to record the trial sessions, a certainly complicated objective for Spanish researchers merely interested in linguistic data. Furthermore, supposing we had been granted access and permission to do so, obtaining an amount of texts that could make our conclusions representative of the variety, would have taken ages. Besides, the range of the text selection included in the BLRC would have required going to one and every courtroom belonging to all the jurisdictions and levels the corpus has been structured into, a definitely unattainable task for a project like this one.

Regarding the texts themselves, they are authentic transcriptions of judicial decisions whose structure may vary depending on the nature of the case and the hierarchical position of the court where it was heard, that is, cases heard at the Supreme Court follow a complex and long route of appeal that implies much greater argumentation and case citation than a case tried at a first-tier tribunal –at the bottom of the judicial pyramid–.

They are full texts obtained in digital format from BAILII, which was certainly a great advantage that saved much time as regards the compilation phase. The size of the texts themselves varies from really long ones of 20,000 words, to really brief ones of about 600. However, the average is 2,000 to 2,500 words. They have all been produced, though not necessarily transcribed, by British judges and reflect their decisions about the cases in question as well as the facts, arguments, prior decisions made at other courts, and any other kind of information relevant to the case.

4.4. CORPUS SIZE AND REPRESENTATIVENESS: ESTABLISHING THE WORD TARGET

Representativeness is central to corpus design, as shown above, and the size of a corpus may determine whether it is representative of the variety of language it aims at covering, or simply an illustrative sample of it with no predictive value.

Authors do not agree regarding the recommended size for a specific corpus. Whereas Pearson (1998) proposes a million words as a reasonable number –she poses that the limit should rather be established by the number of texts available and convertible into digital format–, Sinclair (1991) believes that corpora must be as large as possible establishing 10 to 20m words as the recommendable target for a specific one. On the other hand, Kennedy (1998) does not think that a big corpus necessarily represents the language better than a small one. In general, some 40% or 50% of the lemmas in a corpus occurs once, and more than one occurrence is required to establish a comparison.

In addition to these arguments, we took into consideration the availability of legal texts and their high numbers –16,612 texts in total from 2008 to 2010– in order to decide on the word target. Also the relevance of law reports in the judicial system coupled with the great amount of topics covered by these texts, was determining when we had to make the first decisions on the size of our corpus.

As a consequence, we established that, although this is a specific corpus based just on one legal genre, the target should be 6,000,000 words approximately, six times as big as Pearson proposes, essentially because of the easy access to already digitalized texts in either .rtf or .pdf format and, naturally, of what has just been stated above.

4.5. WIDE TOPIC VARIETY

Law reports should not only be paid special attention within ESP because of their essential function in common law systems, but also because of their vast topic coverage. This corpus has been organised according to the source where

the corpus texts originated, that is, what court or tribunal cases were heard at and decided on.

Tribunals and courts are specialized in a given branch of law: criminal law, family law, commercial law, intellectual law, and so forth, and law reports touch upon one and every branch of both the private and public fields. Judges are in charge of judging cases by both interpreting the law itself –the statutes passed at the parliament–, and fundamentally taking into consideration the existing precedents, therefore their judgments, as reflected on law reports, pertain to all the fields of law.

4.6. *THE STRUCTURE OF COURTS AND TRIBUNALS AND THEIR PARALLELISM WITH THE BLRC*

In *Corpus Linguistics,* McEnery and Wilson cite Biber when highlighting the importance of establishing a clear structure for the design of a corpus prior to its compilation and analysis: "Biber [...] emphasises the advantage of determining beforehand the hierarchical structure or strata of the population, that is, defining what different genres, channels and so on it is made up of" (1996: 79). Hence, we deem it fundamental to justify the categorization method followed to organise the BLRC.

Our corpus retains the current UK tribunal and court structure after its recent modifications as reflected on BAILII due to several reasons, the first one being the relevance of the hierarchy of courts and tribunals in the UK legal system. The principle of binding precedent, which the British judicial system revolves around, establishes that any decision made at a higher court or tribunal will set binding precedent as long as the case is similar to the one under examination in its essence –the *ratio dicendi*–.

Secondly, if we maintain this structure, the texts will be grouped according to the field of law they belong to, so they will be similar in lexical terms, and comparing results by studying the categories separately will be easier and respond to a thematic criterion we consider fundamental as far as our further objective is concerned, that of establishing the core vocabulary of law reports.

In the third place, the route of appeal for a case also responds to this hierarchy. One single case could be heard at more than one tribunal or court if it obtained leave of appeal, that is to say, when a decision is not favourable to any of the parties involved in a trial, it may be appealed to and, if granted permission, it could be heard at higher instances. This fact implies that there are similar tribunals and courts belonging to the same field of law at different levels of the judicial structure, that is, the UK Upper tribunal of Finance and Tax and the First Tier Tax Tribunal deal with similar cases, yet the former is at a higher

level and would either have jurisdiction over certain cases which imply, say, greater amounts of evaded money, or others that come from First Tier tribunals and have been granted leave of appeal.

The same case could go up the structure to the court of last resort of the UK: the Supreme Court, although, as far as the lexical content of the texts is concerned, it would be modified and argued in greater depth every time it is revised and heard at a higher level, hence becoming a different text as it follows the route of appeal.

To finish with the enumeration of the criteria that have conditioned the organization of the corpus, we would like to refer to the distribution of the population in the UK. As it is shown in the UK official census 2011, elaborated by the Office of National Statistics, it appears that almost 90% of the population of the whole territory is concentrated in England and Wales while Northern Ireland only has about 3% and Scotland 9%. Although we have not mathematically distributed the number of texts and word targets per category and subcategory depending on these figures, we did take them into account in order to reinforce the representativeness of the texts obtained from English and Welsh sources that amounted to approximately 55% of the total. The structure and categorisation of the BLRC will be shown in the following section where its distributional criteria and word targets will also be explicated.

4.7. *DISTRIBUTIONAL CRITERIA AND TARGETS PER CATEGORY*

The amount of texts forming the BLRC is not evenly distributed amongst its categories. We found great variation depending on the text source: court or tribunal. Whereas there were sections where the overall number of texts was remarkably high –The Administrative Chamber of England and Wales high Court section of BAILII offered 1922 cases between 2008 and 2010–, there were others which were also exceptionally low with 1 or 2 at most –The UK VAT & Duties Tribunals, Insurance Premium Tax section–.

The reasons for the irregular distribution of the texts available are varied, in some cases, especially regarding tribunals, they have either started working recently or disappeared due to the *Tribunals, Courts and Enforcement Act, 2007*. In some others, the high figures coincide with a densely populated area –one of the criteria supporting text distribution within the corpus– or with a court that, due to its high status in the hierarchy –i.e. any of the chambers of the High Court of Justice of England and Wales–, is in charge of hearing a very high number of cases.

In addition to this, we assume that the fact that a court or tribunal is less productive in terms of text availability implies that there are fewer cases being heard at it. Whether this is true or not, it is beyond our knowledge since BAILII is a free online database supported by authoritative institutions and built up by

reputed contributors who altruistically donate the transcriptions of judge's decisions obtained from official sources.

Nonetheless, and regardless of the possible explanation we may have found for this phenomenon, we believe that the targets established for the sections and subsections of the corpus, had to be proportional to the total number of texts available within the covered time span. Therefore, all the subtargets were set according to this criterion: if the number of texts in a section was higher, we believe they necessarily had to keep a greater proportion of words, and thus be more representative of the language as that is the proportion they keep in real life, or at least we assume this to be so.

The total amount of texts available between 2008 and 2010 is 16,612, and the distribution of texts and targets per section and subsection is exemplified in the following graph for category 4: Northern Ireland Courts and Tribunals. Each main category of the corpus, namely, Commonwealth countries, UK, England and Wales, Scotland and Northern Ireland, has been organised in the same way. After having calculated the word targets per section and subsection, the overall corpus size will be 6,249,000 words.
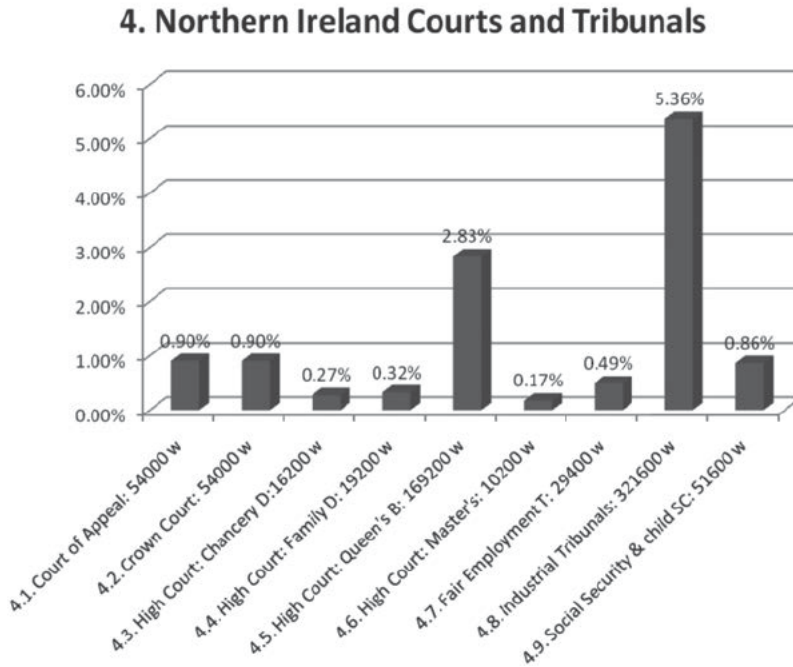
Graph 1. Distribution and targets in the Northern Ireland section of the corpus.

5. CONCLUSION AND FURTHER RESEARCH

The processes of design and compilation of the BLRC have been carried out according to Corpus Linguistics standards so that the results obtained from its subsequent analysis could be worthy, reliable and useful for language teaching and learning. We have attempted to adhere as far as possible to both the criteria proposed for text selection and the guidelines for the compilation of specialized corpora from the literature available.

A well-designed corpus creates an excellent opportunity to look into language evidence and perform quantitative and qualitative analyses. The BLRC is structured in such a way that it will allow multiple contrastive analyses in relation to the different parameters governing in its projection, namely topic variety, types of courts and tribunals, and geographical situation.

Once accomplished the compilation, the following step is the processing of the corpus. WordSmith.5 is the programme chosen for this task. It is first intended to provide direct information about the basic computational characteristics of the corpus. The programme generates immediate data about types, tokens, type-token ratio, frequency lists and so forth. In a second stage, a corpus comparison approach will be adopted in a bid to identify the most significant keywords in legal English in comparison to general English, since the ultimate goal of our research consists in extracting the essential vocabulary of legal English. Our objective fits in the long tradition within lexical studies of developing word lists (Thorndike and Lorge 1944; West 1953; Nation 1990; Coxhead 2000) for teaching and learning English as a second language. In addition, many authors have claimed the adequacy of developing discipline-based lexical repertoires (Nation 2001; Hyland and Tse 2007; Read 2007; Rea 2008) in order to guide materials writers and exams developers, assist instructors' teaching, and meet students' specific needs.

Finally, the BLRC has been envisaged to serve multiple purposes in the long term. The BLRC is intended to be available on line as a source of language examples for students, as a terminological database, even as a limited content-disciple source for students and lawyers and, needless to say, as the solid ground for linguistic research.

REFERENCES

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

Bhatia, V. 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.

Bhatia, V. 2004. "Applied genre analysis: a multi-perspective model". *Ibérica* 4: 3-19.

Biber, D. 1993. "Representativeness in corpus design". *Literary and Linguistic Computing* 8 (4): 243-257.

Bononia Legal Corpus. <http://corpora.dslo.unibo.it/bolc_eng.html> (Accessed 20 May 2010).

Constitutional Reform Act. 2005. <http://www.legislation.gov.uk/ukpga/2005/4/> (Accessed 20 May 2010).

The CorTec corpus. <http://www.fflch.usp.br/dlm/comet/consulta_cortec.html> (Accessed 20 May 2010).

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

Crystal, D. and D. Davy. 1969. *Investigating English Style.* London: Longman.

Danet, B. 1980. "Language in the Legal Process". *Law and Society Review* 14, (3): 445-564.

The HOLJ corpus. <http://www.ltg.ed.ac.uk/SUM/CORPUS/index.html> (Accessed 20 May 2010).

Hunston, Susan. 2002. *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hyland, K. and P. Tse. 2007. "Is there an "Academic Vocabulary"?". *TESOL Quarterly,* 41 (2): 235-253.

Jackson, B. 1985. *Semiotics and Legal Theory.* London: RKP.

JRC-Acquis corpus. <http://langtech.jrc.ec.europa.eu/JRC-Acquis.html> (Accessed 20 May 2010).

Kennedy, G. 1998. *An introduction to corpus linguistics.* New York: Longman.

Kennedy, G. and R. Bolitho. 1984. *English for specific purposes.* London: Mcmillan.

Kurzon, D. 1984. "Themes, Hyperthemes and the Discourse Structure of English Legal Texts." *Text* 4: 1-3.

Maley, Y. 1987. "The Language of Legislation". *Language and Society* 16: 25-48.

Maley, Y. 1994. "The Language of the Law". *Language and the Law.* Ed. J. Gibbons. London: Longman.

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

McEnery, T. and A. Wilson. 1996. *Corpus Linguistics.* Edinburgh: EUP.

Nation, I.S.P. 1990. *Teaching and Learning Vocabulary.* New York, Newbury House.

Nation, I.S.P. 2001. *Learning vocabulary in another language.* Cambridge: CUP.

Orts Llopis, M. A. 2006. *Aproximación al discurso jurídico en inglés: las pólizas de seguro marítimo de Lloyd's*. Madrid: Edisofer.

Orts Llopis, M. A. 2009. "Legal genres in English and Spanish: some attempts of analysis". *Ibérica* 18: 109-130.

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

Räisänen, C. and I. Fortanet, eds. 2008. *ESP in European Higher Education: Integrating language and content*. Amsterdam: John Benjamins.

Rea, C. 2008. *El inglés de las telecomunicaciones: estudio léxico basado en un corpus específico*. Tesis doctoral. <http://www.tesisenred.net/TDR-0611109-134048/index_cs.html> (Accessed 20 May 2010).

Rea, C. 2010. "Getting on with Corpus Compilation: from Theory to Practice". *ESP World*. <http://www.esp-world.info/Articles_27/Camino%20Rea.pdf> (Accessed 20 May 2010).

Read, J. 2007. "Second Language Vocabulary Assessment: Current Practices and New Directions". *International Journal of English Studies* 7 (2): 105-125.

Rossini, R., F. Tamburini, and E. Martelli. 2001. "Words from the Bononia Legal Corpus". *International Journal of Corpus Linguistics* 6 (special issue): 13-34.

Sánchez, A., P. Cantos, R. Sarmiento and J. Simón. 1995. *Cumbre. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y análisis*. Madrid: SGEL.

Sinclair, J. 1991. *Corpus, Concordance and Collocation*. Oxford: OUP.

Sinclair, J. 2005. "Corpus and Text - Basic Principles". *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. Wynne. Oxford. 1-16. <http://ahds.ac.uk/linguistic-corpora/> (Accessed 20 May 2010).

Swales, J. 1985. *Episodes in ESP*. Hemel: Prentice Hall.

Thorndike, E. L. and I. Lorge. 1944. *The teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University.

Tiersma, P. 1999. *Legal Language*. Chicago: UCP.

Tognini-Bonelli, E. 2001. *Corpus linguistics at work*. Amsterdam & Philadelphia: John Benjamins.

Tribunals, Courts and Enforcement Act. 2007. <http://www.legislation.gov.uk/ukpga/2007/15/contents> (Accessed 20 May 2010).

West, M. 1953. *A General Service List of English Words*. London: Longman.

Wynne, M., ed. 2005. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: ASDS Literature, Languages and Linguistics.