ICAME Journal • 47(1) • 2023 • 109–117
DOI: 10.2478/icame-2023-0006

sciendo

ICAME Journal

# The *Corpus of Contemporary English Legal Decisions, 1950–2021* (*CoCELD*): A new tool for analysing recent changes in English legal discourse

**Research Article**

Paula Rodríguez-Puente[1]*, David Hernández-Coalla[2]

[1]*University of Oviedo*
[2]*University of Vigo*

**Received 2 April 2022; accepted 24 June 2022**

**Abstract:** Legal discourse is widely assumed to be resistant to change, and indeed legislative documents are extremely conservative with fixed and formulaic structures. However, recent research has shown that changes can be observed in the lexico-grammatical features of some legal documents when examined diachronically, particularly since the emergence in the 1970s of the Plain Language Movement, which sought to draw attention to the unnecessary complexity of the official language, this including legal discourse. Despite the crucial changes in legal language in recent years, research in that direction is scarce to date, particularly in the British English variety, probably due, in part, to the shortage of specialised corpora that allow this kind of studies. In order to bridge this gap, we have embarked on the compilation of the *Corpus of Contemporary English Legal Decisions, 1950–2021* (CoCELD), a corpus of British judicial decisions produced between 1950 and 2021. In this paper we present the structure and characteristics of CoCELD, as well as the methodology used for its compilation. The new corpus, which was released in February 2022, contains sample texts of roughly 2,500 words for each year from 1950 to 2021, which adds up to more than 730,000 words. The corpus contains files in raw text and with POS-annotation, and is freely available for the research community under signed consent. With CoCELD we hope to contribute with a new, useful resource for linguists with an interest in legal language, from both a synchronic and a diachronic perspective.

**Keywords:** *corpus building • corpus annotation • legal discourse • law reports • Plain Language Movement*

## 1 Introduction

Legal discourse is widely assumed to be resistant to change (Tiersma 1999: 135) and "outside the ravages of time" (Görlach 1999: 145). Indeed, legislative documents do tend to be extremely conservative, with fixed and formulaic structures (see, e.g., Biber and Gray 2019). However, recent research has shown that changes can be observed in the lexico-grammatical features of legal texts when examined diachronically (see, e.g., Rodríguez-Puente 2019, 2020), particularly since the emergence in the 1970s of the Plain Language Movement, which drew attention to the "unnecessary complexity of the official language used by governments, businesses, and other organizations which are in linguistic contact with the public" (Crystal 2019: 401). Most of the activities of the Plain Language Movement, however, focused on materials derived from the law. Hence, since the emergence of these linguistic concerns there have been continued attempts to convince legal drafters to produce documents in a type of language that is "clear and effective for its intended audience" (Butt and Castle 2006: 113). Consequently, most contemporary manuals of legal drafting include recommendations as to what to avoid in order to achieve effective communication, such as the use of the passive voice or the excessive use of nominalisations (see, among others, Garner 2001: 24, 38; Wydick 2005: 27–32; Butt and Castle 2006: 153). This

* *Corresponding author: Paula Rodríguez-Puente, E-mail: rodriguezppaula@uniovi.es*

Open Access. © 2023 Paula Rodríguez-Puente, David Hernández-Coalla, published by Sciendo.
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

has influenced to some extent the way in which legal documents are produced. Williams (2007), for example, notes a considerable reduction in the use of the passive voice in prescriptive legal documents between the 1980s and the 2010s, which may have responded to a need for a more functional and accessible kind of linguistic expression as a means of addressing such demands for plainer language (see also Williams 2004, 2013; Bulatović 2013). Likewise, the use of the modal verb *shall*, one of the most characteristic features of legal English, frequently preferred to *will* to convey obligation, is also declining in UK legislation. This drop is particularly marked from the 1990s onwards, and is also probably influenced by campaigns and drafting recommendations of the Plain Language Movement on the basis that *shall* is itself a token of legalese (see Garzone 2013; Williams 2013), although legislative texts from the European Union seem to have resisted the move away from *shall* (Williams 2013: 362). Outside legal discourse, by contrast, *shall* has been in decline for centuries, in favour of *will* (see, e.g., Rissanen 2000; Gotti 2001, 2003; Leech 2003).

Despite crucial changes in legal discourse over the past few years, research in this area is still scarce, perhaps due in part to the shortage of specialised corpora in this area. In order to bridge this gap, we have embarked on the compilation of the *Corpus of Contemporary English Legal Decisions, 1950–2021* (CoCELD), a corpus of British judicial decisions produced between 1950 and 2021. CoCELD is designed as a new resource, complementary to existing ones, aimed at linguists with a specific interest in English legal language, from both synchronic and diachronic perspectives. In this paper we present the structure and characteristics of CoCELD, the type of texts it contains, as well as the methodology used for its compilation, seeking to illustrate the usefulness of CoCELD in the study of English legal discourse.

The paper is structured as follows. In Section 2 we give an overview of existing specialised corpora of judicial decisions available to the research community, with the aim of justifying the need for the creation of a resource of this kind. In Section 3 we provide a detailed description of the process of compilation of CoCELD, its structure, the type of analyses for which it can be used and also its limitations. Section 4 then provides some concluding remarks and suggestions for further work.

## 2 The place of CoCELD among other legal corpora

The advantages of using any kind of corpus for the analysis of language are manifold, but specialised corpora have proved particularly fruitful in the study of academic and professional language (see Flowerdew 2004; Biel 2010; Pontrandolfo 2019). Contrary to general corpora, with large amounts of data which can be useful in drawing generalisations about language as a whole, specialised corpora are restricted to a particular type of discourse and allow for a more detailed examination of both the linguistic and the extralinguistic features of the texts therein.

Recent research, notably the work of Douglas Biber and associates, has consistently demonstrated the importance of register analysis in the development of languages. As noted by Biber and Gray (2013: 104), “minor differences in register can correspond to meaningful and systematic differences in the patterns of linguistic change”, and in light of this, the analysis of specialised types of discourse has gained crucial importance in recent years. There has been a proliferation of specialised corpora, which have helped to broaden our knowledge of areas such as medical discourse (see, e.g., Taavitsainen et al. 2005; Taavitsainen et al. 2010; Miranda-García et al. 2015; Calle-Martín et al. 2016), news discourse (Ungerer et al. 2000; Lehmann et al. 2006) and different types of scientific discourse (see especially the work of Isabel Moskowich and associates on the *Coruña Corpus of English Scientific Writing*),[1] among many others.

Specialised corpora containing samples of legal documents have also emerged over the past few years, probably in response to a growing interest in legal language which resulted from developments in (applied) linguistics and the social sciences (Bhatia 1987: 227). Comprehensive overviews of specialised corpora are provided by Pontrandolfo (2012) and Fanego et al. (2017). Our focus here is on corpora

1 https://ruc.udc.es/dspace/handle/2183/21846.

which, like CoCELD, contain judicial decisions. In the British Common Law system, judicial decisions, also known as legal decisions or law reports, make up Case Law. Decisions constitute a crucial type of document, in that they set legal precedents, "interpret parliamentary law and fill in the gaps where there is no statute law" (Kearns 2007: 9).

Besides featuring in several multi-genre corpora (see Fanego et al. 2017: 60–61), judicial opinions are included in a number of specialised corpora of American and British English. As far as American English is concerned, the Law School of the Brigham Young University has recently released a series of legal corpora developed by the project *Law and Corpus Linguistics*,[2] which seeks to bring the working methods of corpus linguistics closer to legal scholars and judges. The project website hosts, among others, the *Corpus of Supreme Court Opinions of the United States* (COSCO-US), a 98.4-million-word corpus which contains all opinions in the United States Reports and opinions published by the Supreme Court through the 2017 term.[3] Another notable tool to emerge recently is the *Corpus of US Supreme Court Opinions* (Davies 2017), with ca. 130 million words of legal opinions from the 1790s to the present.

As far as British English is concerned, it is important to mention the Law and Language at the European Court of Justice Project,[4] which aims at providing a new understanding of the development of EU law. The project brings together linguistic theories, anthropological research methods and law in order to examine the production of the multilingual jurisprudence of the Court of Justice of the European Union. One of the goals of the project involved the creation of the *EU Case Law Corpus* (EUCLCORP), a standardised, multidimensional and multilingual corpus containing all judgments of the Court of Justice of the European Union and judgments delivered by constitutional and supreme courts of seven EU member states (see Trklja and McAuliffe 2018). The language subcorpora included in EUCLCORP (currently British English, German and French) constitute valuable resources for cross-linguistic research into judicial language.

Judicial decisions are also represented in three different specialised corpora: the *House of Lords Corpus* (HOLJ; Grover et al. 2004), the *British Law Reports Corpus* (BLaRC; Marín Pérez and Rea Rizzo 2012) and the *Corpus of English Historical Law Reports, 1535–1999* (CHELAR; Rodríguez-Puente et al. 2018). HOLJ contains roughly 188 judgements from the years 2001–2003 delivered in the House of Lords. BLaRC, in turn, contains 8.85 million words of judicial decisions issued between 2008 and 2010 from five different jurisdictions of the British judicial system. Both HOLJ and BLaRC are designed as synchronic corpora. For researchers interested in the diachronic development of the language of legal decisions, CHELAR is available, with roughly half a million words of law reports from 1535 to 1999.

Used in conjunction, HOLJ, BLaRC and CHELAR allow us to trace the development of English legal discourse between 1535 and 2010, albeit discontinuously, since some years of the twenty-first century are not represented. Beyond 2010, however, these corpora do not sample this type of document in the British English variety. For this reason, we saw the need to create a corpus that allows researchers to investigate the evolution of English legal discourse continuously from the time before the implementation of the Plain Language Movement up to 2021. The novelty of CoCELD, then, is that it is the only currently available tool which facilitates the analysis of recent developments in this specialised type of discourse, which is itself remarkable, given the recent influence of the Plain Language Movement. Moreover, CoCELD complements existing resources, so that researchers now have access to a broad overview of the language of judicial decisions over six centuries.

# 3 Building CoCELD

## 3.1 Text selection

The corpus texts were taken from the website of the British and Irish Legal Information Institute (BAILII),[5] which provides free access to British and Irish case law and legislation, European Union case law, Law

2 https://lcl.byu.edu/projects/law-corpus-linguistics-background/.
3 For an overview of this and other corpora of legal English, see https://lawcorpus.byu.edu/.
4 https://llecj.karenmcauliffe.com/.
5 https://www.bailii.org/.

Commission reports, and other law-related British and Irish material.[6] For our purposes, we focused on case law produced within the United Kingdom and the Commonwealth. More specifically, the texts in CoCELD are samples of decisions issued by the Privy Council, the House of Lords, and the United Kingdom Supreme Court. These were preferred to material in other databases because they cover a long and continuous time span. The Privy Council is the highest court of appeal for certain British territories and some Commonwealth countries. In turn, the House of Lords was the highest court of appeal in the United Kingdom until July 2009. From October 2009, the Supreme Court assumed jurisdiction on points of law for all civil cases in the United Kingdom and for criminal cases in England, Wales, and Northern Ireland. It, then, assumed the judicial functions of the House of Lords, which had been exercised by the Law Lords, the twelve judges appointed as members of the House of Lords. With the establishment of the Supreme Court, it was possible to achieve a complete separation between the United Kingdom's senior judges and the Upper House of Parliament, emphasising the independence of the Law Lords and increasing transparency between Parliament and the courts. CoCELD, then, is designed with two subcorpora: one for Privy Council decisions, and the other for the decisions of the House of Lords and the Supreme Court.

In BAILII some of the texts are in PDF format, whereas others are HTML documents. Although text selection was, in principle, random,[7] we gave priority to texts in HTML form, which were relatively simple to process. We downloaded two decisions per year and per court (i.e. four samples per year), that is, a total 288 decisions: 221 in HTML and 67 in PDF.[8]

## 3.2 Text processing and edition

Legal decisions vary considerably in length: some contain three or five pages, others more than twenty. Including whole texts, then, would have resulted in a very imbalanced corpus, biased towards the specific language features of those decisions with an ampler representation. For this reason, following a common practice in corpus compilation (e.g. in ARCHER and CHELAR), in building CoCELD we included samples of around 2,500 words. As with the compilation of CHELAR (see Rodríguez-Puente 2011; Fanego et al. 2017), we removed front matter (i.e. title, names of appellants, judges, etc.) and footnotes, and took the sample text from the beginning up to roughly 2,500 words. We also removed paragraph numbers, which were present in some decisions, since they were irrelevant for linguistic analysis.

The samples extracted were converted to TXT files. This process was far from simple. In principle, the conversion of HTML files was relatively easy. However, we soon realised that some of the texts, particularly the earliest ones, had been converted to HTML using OCR software, often with little revision. This obliged us to carefully revise these texts to identify possible typos (e.g. *1t* for *it* or *1* for *I*) and blank spaces before punctuation, among other issues, guided by our intuitions, as we did not have access to the original documents when revising them. The PDF files were handled in two different ways. Those of good quality were converted to TXT using OCR software, then proof-read. Where the quality was poor, texts were typed manually, then carefully revised for possible typos. At the same time, we removed all non-ASCII characters (e.g. pound symbols, curly quotation marks, etc.) so as to prepare the texts for part-of-speech (POS) annotation at a later stage.

The corpus files were named using the following structure: the year in which the judgment was issued, followed by a number (1 or 2) indicating the earlier (1) or later (2) text from that year, and finally four letters indicating the title of the judgement. For example, the file 2021_2_dunc.txt is the second

6 BAILII allows users to copy, print and distribute legal material published on their website free of change under the copyright restrictions laid out on their copyright notice. See https://www.bailii.org/bailii/copyright.html.

7 An anonymous reviewer enquires about how the random selection of texts took place. The legal decisions in BAILII are arranged by years. To randomly select the texts, we simply opened a file from the first half of the year and another from the second half. If the document contained sufficient textual material (i.e. more than 2,500 words), we downloaded it. If it was too short, we discarded it and selected another one. We also considered format, giving preference to HTML files. When only PDF files were available for a specific year, we tried to discard those with poorer quality, although this was not possible in all cases.

8 Specific information and full reference to the judgments are provided in the corpus documentation.

case from 2021, corresponding to the judgement entitled *Duncan and Jokhan v. Attorney General of Trinidad and Tobago*.

Once the raw-text files had been prepared, we created POS-annotated texts using the CLAWS-7 tagger (Constituent Likelihood Automatic Word-tagging System; see Garside 1987). The addition of other tagging or annotation systems, such as XML or semantic annotation, is, at the moment, outside the scope of our plans. However, we plan to tag CoCELD using the Multidimensional Analysis Tagger (MAT; Nini 2019). This software annotates texts based on the Stanford Tagger,[9] which can be later analysed from a multidimensional functional perspective similar to the method originally designed by Biber (1988).

## 3.3 Structure of CoCELD

Achieving balance in corpus compilation is far from straightforward (for a discussion, see, among others McEnery et al. 2006: 16–18). However, in an attempt to make CoCELD as balanced as possible, as well as to facilitate the analysis of search results, we split our corpus into two subcorpora, divided into seven subperiods of ten years each, except for the last subperiod which comprises the years from 2010 to 2021. The structure of the corpus is set out in Table 1 below. As can be seen, the whole corpus comprises 733,227 words: 366,061 from the Privy Council and 367,166 from the House of Lords and the Supreme Court.

In order to measure the number of types contained in CoCELD, we made a WordList with WordSmith 6 (Scott 2012). This provided a total of 18,498 different entries of which 33.7 per cent (6,231 types) are single-occurrence items or *hapax legomena*. As pointed out repeatedly in the literature, there is no ideal size for a corpus. According to Flowerdew (2004: 18), "a specialized corpus should be of adequate size such that there is a sufficient number of occurrences of a linguistic structure or pattern to validate a hypothesis". For Leech (1991: 27), representativeness in corpus building implies that the findings based on its contents are generalisable to a larger hypothetical corpus, although he also acknowledges that "representativeness must be regarded largely as an act of faith" (Leech 1991: 21).[10] As is well known, the larger the size of a corpus, the fewer new items are encountered. In other words, "while any linguistic corpus increases linearly in tokens in a completely regular or stable shape, its increase in types [...] starts declining the more the corpus grows, as it contributes fewer new types" (Sánchez and Cantos 1997: 261). In this regard, considering the distribution of types and tokens, the samples in CoCELD seem to be representative enough as to detect language evolution, variation, change, and fluctuation.

**Table 1.** Total number of words of CoCELD per period and subcorpus.

| | Privy Council Decisions | | House of Lords and Supreme Court Decisions | | Total words | Total files |
|---|---|---|---|---|---|---|
| Period | Number of words | Number of files | Number of words | Number of files | | |
| 1950–1959 | 50,704 | 20 | 51,066 | 20 | 101,770 | 40 |
| 1960–1969 | 50,848 | 20 | 51,245 | 20 | 102,093 | 40 |
| 1970–1979 | 50,652 | 20 | 50,969 | 20 | 101,621 | 40 |
| 1980–1989 | 50,880 | 20 | 50,827 | 20 | 101,707 | 40 |
| 1990–1999 | 50,998 | 20 | 51,085 | 20 | 102,083 | 40 |
| 2000–2009 | 50,837 | 20 | 50,792 | 20 | 101,629 | 40 |
| 2010–2021 | 61,142 | 24 | 61,182 | 24 | 122,324 | 48 |
| **Total** | **366,061** | **144** | **367,166** | **144** | **733,227** | **288** |

9 https://nlp.stanford.edu/software/tagger.shtml.

10 For discussion on the disparity of views about what representativeness implies in corpus building, see Egbert et al. (2022: 28–64).

## 3.4 Research possibilities of CoCELD

As noted in previous sections, CoCELD is compiled specifically to carry out research, both diachronic and synchronic, on the linguistic features (e.g. pronouns, verbs, nouns, etc.) of judicial decisions on the assumption that those language features serve communicative functions (Biber and Conrad 2019: 2). The ultimate goal is to explore the evolution of English legal discourse in relation to the effects of the Plain Language Movement. In this sense, CoCELD lends itself particularly well to register analyses in terms of Biber and Conrad's (2019: 6–19) terminology and approach. Register description must include three components: 1) the situational context of the texts (i.e. whether they are produced in speech or writing), 2) the lexico-grammatical features of the texts, and 3) the functional relationship between the first two components (Biber and Conrad 2019: 6). This type of analysis assumes that "some linguistic features are common in a register because they are functionally adapted to the communicative purposes and situational contexts of texts from that register" (Biber and Conrad 2019: 2). The function of judicial decisions is to set legal precedents and to interpret parliamentary law. This has a notable effect on the type of lexico-grammatical features which characterise them, because the choices made when producing any type of document are inherently functional. As noted by Rodríguez-Puente (2019), for example, the use of the first person singular pronoun (*I*) is very common in judicial decisions, mostly because these contain "the judges' opinions, including their positions, beliefs and interpretations of laws and precedents [and] an authorial presence in the text confers on them the authority for their decisions and judgments" (2019: 186). In other, more prescriptive, documents, such as acts of parliament, proclamations and statutes, the first person singular pronoun is practically unattested. Given that their function is to establish prescriptive laws and rules, they must be written in an impersonal, objective, rational and fact-based narratorial style, with the minimum degree of involvement and interaction, and hence with no pronominal references to either author or addressee. This and other lexico-grammatical features typically associated with legal discourse, such as the use of compound adverbs (*hereof*, *thereof*), the inclusion of Latin, French and Anglo-Norman vocabulary, a highly nominalised style, and the use of long and complex sentences,[11] can all be explored in CoCELD. However, one clear limitation of CoCELD is that it cannot be used for move analysis (Swales 1981, 1990) or genre analysis (Biber and Conrad 2019: 16–18), where the focus is on the rhetorical organisation of texts, for which full texts, rather than excerpts, are required. While we acknowledge that move analysis has proved useful for the investigation of law-related documents (see, e.g., Groom and Grieve 2019), CoCELD was not specifically designed for this purpose.

Used individually and/or drawing comparisons with other corpora, however, CoCELD can be explored in terms of the four possible research areas which, according to Biel (2010: 4–6), can be applied to the study of legal language: 1) cross-genre variation (that is, differences between legal language and general languages or other specialised languages); 2) genre-internal variation (differences among legal genres); 3) diachronic variation (differences in legal language over time); and, 4) cross-linguistic variations (differences across languages).

## 3.5 Availability of CoCELD

As noted by Flowerdew (2004), it is often the case that analysts are also corpus compilers, and as such are familiar with the underlying socio-cultural dimensions represented in a corpus. Thus they become "compilers-cum-analysts" who can act "as a kind of mediating ethnographic specialist informant to shed light on the corpus data" (Flowerdew 2004: 16). CoCELD is indeed designed with specific purposes in mind, and with the goal of addressing particular research questions concerning the development of the language of judicial decisions over recent decades. However, CoCELD has not been designed to be used exclusively by its compilers. On the contrary, we encourage other researchers to use it towards enriching and broadening our knowledge of the language of judicial decisions. Hence, CoCELD has been made available free of charge, by request to the authors. An open access folder hosted on cloud storage can be accessed under the conditions laid out in the User Agreement.

11 For other features typically associated with legal discourse, see, among others, Tiersma (1999: 51–114), Wydick (2005), Williams (2007: 31–38), and Johnson and Coulthard (2010: 10).

# 4 Conclusion

In this paper we have described the process of compilation and the main features of the *Corpus of Contemporary English Legal Decisions, 1950–2021*, a new, freely available corpus of judicial decisions running from 1950 to 2021. CoCELD is currently the only available corpus which allows for the analysis of recent developments in the language of judicial decisions across a continuous time span in which decisions from the Privy Council, the House of Lords and the Supreme Court are represented. It therefore complements existing resources that include judicial decisions issued in the United Kingdom, expanding the potential for research for those with an interest in legal discourse both from a diachronic and a synchronic perspective. In spite of this, there still remain lacunae as far as the compilation of specialised corpora of legal English is concerned. Further work should be aimed at expanding the number of words and texts of CoCELD with the goal of making it more representative of the text variety it represents. Moreover, as legal discourse is not homogeneous, it would also be interesting to compile parallel corpora representative of other subregisters of the domain of Law. These and other aspects will be the focus of future work in the field.

## References


ARCHER = *A Representative Corpus of Historical English Registers* version X. 1990–1993/2002/2007/2010/2013/2016. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Current member universities are Bamberg, Freiburg, Heidelberg, Helsinki, Lancaster, Leicester, Manchester, Michigan, Northern Arizona, Santiago de Compostela, Southern California, Trier, Uppsala, Zurich.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas and Susan Conrad. 2019. *Register, genre, and style*. Cambridge: Cambridge University Press.

Biber, Douglas and Bethany Gray. 2013. Being specific about historical change: The influence of sub-register. *Journal of English Linguistics* 41 (2):103–134.

Biber, Douglas and Bethany Gray. 2019. Are law reports an “agile” or an “uptight” register? Tracking patterns of historical change in the use of colloquial and complexity features. In T. Fanego and P. Rodríguez-Puente (eds.). *Corpus-based research on variation in English legal discourse*, 149–169. Amsterdam: John Benjamins.

Biel, Łucja. 2010. Corpus-based studies of legal language for translation purposes: Methodological and practical potential. In C. Heine and J. Engberg (eds.). *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*, 1–15. Aarhus: Aarhus School of Business, Aarhus University.

Bhatia, Vijay K. 1987. Language of the law. *Language Teaching* 20 (4): 227–234.

Bulatović, Vesna. 2013. Legal language: The passive voice myth. *ESP Today. Journal of English for Specific Purposes at Tertiary Level* 1 (1): 93–112.

Butt, Peter and Richard Castle. 2006. *Modern legal drafting: A guide to using clearer language*. Cambridge: Cambridge University Press.

Calle-Martín, Javier, David Moreno-Olalla, Laura Esteban-Segura, Miriam Criado-Peña, Juan Camilo Conde-Silvestre, Teresa Marqués-Aguado, Santiago González-Fernández-Corugedo, Graham D. Caie, Jacob Thaisen, Hanna Rutkowska and Jesús Romero-Barranco. 2016. *The Málaga Corpus of Early Modern English Scientific Prose* (MCEMESP). Málaga: University of Málaga. Available from: https://modernmss.uma.es.

Crystal, David. 2019. *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.

Davies, Mark. 2017. *Corpus of US Supreme Court Opinions*. Available online at https://www.english-corpora.org/scotus/.

Egbert, Jesse, Douglas Biber and Bethany Gray. 2022. *Designing and evaluating language corpora: A practical framework for corpus representativeness*. Cambridge: Cambridge University Press.

Fanego, Teresa, Paula Rodríguez-Puente, María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo, Cristina Blanco-García and Iván Tamaredo. 2017. The *Corpus of Historical English Law Reports 1535–1999 (CHELAR)*: A resource for analysing the development of English legal discourse. *ICAME Journal* 41: 53–82.

Flowerdew, Lynne. 2004. The argument for using English specialized corpora to understand academic and professional language. In U. Connor and T. A. Upton (eds.). *Discourse in the professions*. *Perspectives from corpus linguistics*, 11–33. Amsterdam: John Benjamins.

Garner, Bryan A. 2001. *Legal writing in plain English: A text with exercises*. Chicago: The University of Chicago Press.

Garside, Roger. 1987. The CLAWS word-tagging system. In R. Garside, G. Leech and G. Sampson (eds.). *The computational analysis of English: A corpus-based approach*, 30–41. London: Longman.

Garzone, Giuliana. 2013. Variation in the use of modality in legislative texts: Focus on *shall*. *Journal of Pragmatics* 57: 68–81.

Görlach, Manfred. 1999. *Nineteenth-century England: An introduction*. Cambridge: Cambridge University Press.

Gotti, Maurizio. 2001. Semantic and pragmatic values of *shall* and *will* in Early Modern English Statutes. In M. Gotti and M. Dossena (eds.). *Modality in specialized texts*, 89–112. Bern: Peter Lang.

Gotti, Maurizio. 2003. *Shall* and *will* in contemporary English: A comparison with past uses. In R. Facchinetti, M. Krug and F. R. Palmer (eds.). *Modality in contemporary English*, 267–300. Berlin: Mouton de Gruyter.

Groom, Nicholas and Jack Grieve. 2019. The evolution of a legal genre: Rhetorical moves in British patent specifications, 1711 to 1860. In T. Fanego and P. Rodríguez-Puente (eds.). *Corpus-based research on variation in English legal discourse*, 201–234. Amsterdam: John Benjamins.

Grover, Claire, Ben Hachey and Ian Hughson. 2004. The HOLJ corpus: Supporting summarisation of legal texts. In S. Hansen-Shirra, S. Oepen and H. Uszkoreit (eds.). *Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora (LINC-04), Geneva, Switzerland* [no pagination]. Geneva: University of Geneva.

Johnson, Alison and Malcolm Coulthard. 2010. Introduction: Current debates in forensic linguistics. In M. Coulthard and A. Johnson (eds.). *The Routledge handbook of forensic linguistics*, 1–15. London: Routledge.

Kearns, Martin. 2007. *Legal English*. Madrid: Colex.

Leech, Geoffrey. 1991. The state of the art in corpus linguistics. In K. Aijmer and B. Altenberg (eds.). *English corpus linguistics*, 8–28. London: Longman.

Leech, Geoffrey. 2003. Modality on the move: The English modal auxiliaries 1961–1992. In R. Facchinetti, M. Krug and F. R. Palmer (eds.). *Modality in contemporary English*, 223–240. Berlin: Mouton de Gruyter.

Lehmann, Hans Martin, Caren auf dem Keller and Beni Ruef. 2006. Zen Corpus 1.0. In R. Facchinetti and M. Rissanen (eds.). *Corpus-based studies of diachronic English*, 135–155. Bern: Peter Lang.

Marín Pérez, María José and Camino Rea Rizzo. 2012. Structure and design of the British Law Report Corpus (BLRC): A legal corpus of judicial decisions from the UK. *Journal of English Studies* 10: 131–145.

McEnery, Tony, Richard Xiao and Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book*. London: Routledge.

Miranda-García, Antonio, Javier Calle-Martín, David Moreno-Olalla, Laura Esteban-Segura, Juan Camilo Conde Silvestre, Teresa Marqués-Aguado, Graham D. Caie, Santiago González-Fernández-Corugedo and Alejandro Alcaraz-Sintes. 2015. *The Málaga Corpus of Late Middle English Scientific Prose* (MCLMESP). Málaga: University of Málaga. Available from: https://hunter.uma.es.

Nini, Andrea. 2019. The Multi-Dimensional Analysis Tagger. In T. Berber-Sardinha and M. Veirano Pinto (eds.). *Multi-dimensional analysis: Research methods and current issue*s, 67–94. London: Bloomsbury Academic.

Pontrandolfo, Gianluca. 2019. Corpus methods in legal translation studies. In Ł. Biel, J. Engberg, R. Martín-Ruano and V. Sosoni (eds.). *Research methods in legal translation and interpreting. Crossing methodological boundaries*, 13–28. London: Routledge.

Rissanen, Matti. 2000. Standardization and the language of early statutes. In L. Wright (ed.). *The development of standard English 1300–1800*, 117–130. Cambridge: Cambridge University Press.

Rodríguez-Puente, Paula. 2011. Introducing the *Corpus of Historical English Law Reports*: Structure and compilation techniques. *Revista de Lenguas para Fines Específicos* 17: 99–120.

Rodríguez-Puente, Paula. 2019. Interpersonality in legal written discourse: A diachronic analysis of personal pronouns in law reports, 1535-present. In T. Fanego and P. Rodríguez-Puente (eds.). *Corpus-based research on variation in English legal discourse*, 171–199. Amsterdam: John Benjamins.

Rodríguez-Puente, Paula. 2020. Historical legal discourse: British law reports. In E. Friginal and J. A. Hardy (eds.). *The Routledge handbook of corpus approaches to discourse analysis*, 499–517. New York: Routledge.

Rodríguez-Puente, Paula, Teresa Fanego, María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo, Cristina Blanco-García and Iván Tamaredo. 2018. *Corpus of Historical English Law Reports 1535–1999 (CHELAR), v.2*. University of Santiago de Compostela: Research Unit for Variation, Linguistic Change and Grammaticalization, Department of English and German.

Sánchez, Aquilino and Pascual Cantos. 2017. Predictability of word forms (types) and lemmas in linguistic corpora. A case study based on the analysis of the *CUMBRE Corpus*: An 8-million-word corpus of contemporary Spanish. *International Journal of Corpus Linguistics* 2 (2): 259–280.

Swales, John. 1981. *Aspects of article introductions*. Birmingham: University of Ashton.

Swales, John. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taavitsainen Irma, Päivi Pahta and Martti Mäkinen (eds.). 2005. *Middle English Medical Texts*. Amsterdam: John Benjamins.

Taavitsainen Irma, Päivi Pahta, Turo Hiltunen, Martti Mäkinen, Ville Marttila, Maura Ratia, Carla Suhr and Jukka Tyrkkö. 2010. *Early Modern English Medical Texts*. Amsterdam: John Benjamins.

Tiersma, Peter M. 1999. *Legal language*. Chicago, IL: The University of Chicago Press.

Trklja, Aleksandar and Karen McAuliffe. 2018. The European Union case law corpus (EUCLCORP): A multilingual parallel and comparative corpus of EU court judgments. In A.U. Frank, C. Ivanovic, F. Mambrini, M. Passarotti and C. Sporleder (eds.). *Proceedings of the Second Workshop on Corpus-based Research in the Humanities: CRH-2. Vol. 1, Gerastree Proceedings*, 217–226. Vienna, Austria: Second Workshop on Corpus-Based Research in the Humanities.

Ungerer, Friedrich, Kristina Schneider and Birte Bös. 2000. *Rostock Newspaper Corpus*. Rostock: Rostock University.

Williams, Christopher. 2004. Legal English and plain language: An introduction. *ESP Across Cultures* 1: 111–124.

Williams, Christopher. 2007. *Tradition and change in legal English. Verbal constructions in prescriptive texts*. Bern: Peter Lang.

Williams, Christopher. 2013. Changes in the verb phrase in legislative language in English. In B. Aarts, J. Close, G. Leech and S. Wallis (eds.). *The verb phrase in English. Investigating recent language change with corpora*, 353–371. Cambridge: Cambridge University Press.

Wydick, Richard C. 2005. *Plain English for lawyers*. Durham: Carolina Academic Press.