The Features of Legal Language in British Legislation: a Diachronic Perspective

Eva Dvořáková

Abstract The aim of this paper is to explore diachronically the features of legal language in British legislation over a period of 200 years to find out whether the features identified by Crystal and Davy in their seminal work are still present in current British legislation and whether any changes can be attributed to plain language efforts. Further, the paper explores lexical diversity and readability. For these purposes, five corpora of British legislation were compiled from 1820, 1870, 1920, 1970, and 2020. The results show that the current language of British legislation is very different from the legal language described by Crystal and Davy fifty years ago. The plainer texts tend to be lexically less diverse. The readability measures provide rather inconclusive results.

Keywords archaisms, British legislation, legal language, lexical diversity, plain language, readability

1 Introduction

When asked to identify typical features of legal language, most people will mention long, complex sentences, accurate terminology, and certain legalisms, such as *shall*, *hereinafter*, or *aforesaid*. However, do these perceptions match the reality?

In their seminal work, *Investigating English Style*, in 1969, Crystal and Davy offered a perspicacious insight into legal language, summarising its main features by analysing two extracts from legal documents (Crystal/Davy 1990). However, when looking at the two extracts, one wonders whether such examples are representative of legal language today, fifty years after the book was published. In other words, plain language movements[1] and other factors may have changed legal language to the point where the description provided by Crystal and Davy may now only partially apply.

The aim of this paper is to examine diachronically a genre[2] of legal English, namely British legislation, to find out whether it possesses the features of legal language described by Crystal and Davy and whether these features have changed over time. The selected features are lexical features that can be analysed through corpora. They are the following: the scarcity of personal pronouns, the use of archaic compound adverbs, pairs of synonymous adjectives, the modal

[1] Some authors disagree with the word *movement*. For example, according to Balmford (2002), plain language "has grown beyond being a movement to become a product, a business, an industry, or a professional service".

[2] This paper uses the word *genre* as understood by Kurzon: "We recognize texts belonging to specific legal genres – contracts, judgments, legislation – through pragmatic means, by way of looking at the purpose of the text." (Kurzon 1997: 125)

Zitiervorschlag / Citation:
Dvořáková, Eva (2022): „The Features of Legal Language in British Legislation: a Diachronic Perspective." *Fachsprache. Journal of Professional and Scientific Communication* 44.3–4: 112–129.

shall, and the word *such* used as a determiner. These features will be analysed through corpora of British legislation from 1820, 1870, 1920, 1970, and 2020.

Over the past decades, the move towards plain language has influenced the way official and legal documents are drafted. For example, Wydick (1978: 738) recommends using familiar words and avoiding lawyerisms. In this way, by restricting the vocabulary, legal language could become lexically less diverse than language that is not subject to such restrictions. Recent technological developments have made it possible to assess lexical diversity relatively easily. The relevant diachronic corpora will be analysed to support or disprove the hypothesis that the introduction of plain language tends to make texts lexically less diverse.

Finally, the avowed aims of plain language exponents are to make official and legal documents more accessible to their readers. The development of so-called readability measures has facilitated the objective assessment of texts in terms of their accessibility to lay readers. Most of the measures take account of sentence length and the number of syllables in individual words; some check the vocabulary against a list of frequently used words. Thus, the third strand of this research aims to analyse our diachronic corpora of British legislation by means of readability measures, to determine whether these objective tools can confirm that over time British legislation has become plainer and therefore more accessible to lay readers.

In 1969, Crystal and Davy selected two extracts from legal documents (an endowment assurance policy and a hire purchase agreement) as the basis for their analysis. Those were private documents from the domain of contract law. Our study is limited to British legislation. Thus, the distinction between the drafting practices in public and private documents may have affected the results. Nevertheless, Crystal and Davy apparently used their two texts to derive general principles applicable to legal language as such, as evidenced by their explanation that they "have chosen two examples that … [they] felt to be reasonably central in a linguistic sense" (Crystal/Davy 1990: 195). This raises the question of whether the texts selected for the purposes of the present study, namely legislation, can be considered sufficiently "central" so as to be representative of legal language as a whole. We believe that legislation is one of the central genres which sets the pace for the whole domain of legal language. Yet the existence of multiple genres within legal language makes any generalisation problematic; the paper therefore does not aspire to cover the whole of legal language, but rather to explore solely British legislation.

While insightful research into several aspects of plain language in legislation has been undertaken over the last decades (for example the use of *shall* was explored by Garzone [2013a, 2013b] and Williams [2006, 2012]), to the best of our knowledge little effort has been made so far to diachronically compare British legislation over a substantial period of time, using corpora and data. We therefore hope to contribute to a better understanding of the practical benefits of plain language.

2 Literature review

In the 1960s Mellinkoff offered, in his seminal work *The Language of the Law*, a comprehensive analysis of legal language. Rather than applauding the peculiarities of this specific type of language, he vigorously criticised certain typical features which tend to hamper comprehension or are simply redundant. With his statement "the language of the law should not be different without a reason" (Mellinkoff 1963: 285), he embraced a pragmatic approach that readers should not be required to struggle in order to understand legal documents. A similar spirit can be found in the books by Butt (2018) and Tiersma (2000).

In 1969, Crystal and Davy offered their dispassionate analysis of legal language in the book *Investigating English Style*. Rather than taking a stance like Mellinkoff or later Butt and Tiersma, they sought to objectively identify the features of legal language, by analysing two extracts from legal documents (an endowment assurance policy and a hire purchase agreement) that they selected as representative of legal language. The authors provided a comprehensive analysis of legal language, starting with the layout and capitalisation, continuing with the vocabulary, and ending with the sentence structure.

Other influential scholars and professionals (e. g., Garner 2001, Kimble 1992, Williams 2015) facilitated the spread of these ideas across the globe and in professional settings. In this way, the plain language requirement has found its way into statutes (e. g., The US Plain Writing Act of 2010) and numerous legislative drafting manuals.[3] Yet "regulatory demand" is not the only driving force; in Australia, plain language tends to be introduced in response to "client demand" (Balmford 2002: 5.4), being seen as a competitive edge for some law firms: "One day, clients everywhere will refuse to pay for legal services unless they are plain" (Balmford 2002: 5.3). This is in stark contrast to the "bang for the buck" argument (Sneddon 2011: 713) explained by Gopen: "Clients who pay such prices, the argument runs, want to see their received value in terms of the degree of difficulty of the product." (Gopen 1987: 345)

The adoption of plain language statutes requires some guidance for judges to determine which texts are plain and compliant. Cheek (2010) offered a comprehensive approach to plain language definitions, comprising the following: (1) numerical or formula-based definitions (using the Flesch readability formula); (2) elements-focused definitions; and (3) outcomes-focused definitions. The formula-based definitions provide only "rough guides, however, scores derived from readability formulas provide quick, easy help" (DuBay 2004: 19). Cheek prefers outcomes-focused definitions which can, however, be rather demanding in terms of their implementation. She further suggested that "plain English is not an absolute but should be appropriate to the intended audience" (Cheek 2010: 9). In fact, consumers cannot be seen as the only addressees of plain language because "clear communication is for all" (Balmford 2002: 5.1), and even judges seem to prefer plain language (Palyga 1999). Nevertheless, Long and Christensen (2011) have shown that briefs written in plain language do not have an impact on the outcome of the case. Conversely, Benson and Kessler have established that "[l]awyers who write in legalese are likely to have their work judged as unpersuasive and substantively weak" (Benson/Kessler 1987: 319).

The advice given by plain language promoters to lawyers is to use familiar words and avoid lawyerisms so as not to "send your reader groping for the dictionary" (Wydick 1978: 738). Can such an approach have an impact on the lexical diversity of legal texts? Cvrček and Chlumská (2015) examined one of the translation universals, namely simplification, in translated texts, trying to find an answer to the question of whether translated texts tend to be lexically less diverse. To do that, the authors sought to develop a reliable method for measuring lexical diversity, given the limitations of the traditional TTR (type-token ratio) approach. They came up with the zTTR method (a modified type-token ratio approach, incorporated into the online tool https://www.korpus.cz/calc/) which compares the TTR values of the examined text with referential values, with the text type and text size also being taken into consideration. This novel method subsequently confirmed their hypothesis that the translation process tends to deprive translated texts of some of their lexical diversity. Although the present study does not

[3] For more details on the most important plain language initiatives, cf. Williams 2015.

deal with translations, the methodology used by Cvrček and Chlumská could help to determine whether plain language texts tend to be lexically poorer.

Since the 1940s, economic interests have led book and newspaper publishers to try to increase their readership by specifically tailoring the texts to the skills of the readers; this led to the development of objective measures of readability, the so-called readability formulas. In his comprehensive overview of readability principles, DuBay (2004) provided valuable insights into numerous readability formulas. The best-known formula is the Flesch formula, based on sentence length and the number of syllables. A different approach was adopted for the Dale-Chall formula which, apart from sentence length, uses a list of 3,000 easy words and detects the "hard" words not included in the list. DuBay mentions a study according to which "the average adult in the U.S. reads at the 7th-grade level" (DuBay 2004: 1) and materials for public should be written at fifth- or sixth-grade reading level. The research into readability thus provides relatively objective criteria for the drafters of official documents who seek to adapt their messages to their readership. Yet as suggested by DuBay, readability measures have their limitations. This has been aptly demonstrated by Šlerka and Smolík (2010) who applied various readability formulas to a variety of Czech texts. Among other things, they showed that a children's book scored as the most difficult text from their corpus.

3 Methodology

3.1 Corpora

When designing the research, the obvious question arises as to how long the time span to be investigated should be. Since the key results should elucidate how the language of modern British legislation has changed in comparison with the legal language of the time when *Investigating English Style* by Crystal and Davy was published (1969), the key time periods are 1970 and 2020. In this way, it is possible to see how legal language has developed over the past 50 years and whether the plain language campaigns have had any measurable impact on the language of British legislation. In addition, however, it would be interesting to see the bigger picture and find out whether some of the changes to legal language would have occurred anyway. Consequently, it was decided to cover a longer time span, namely 200 years.

The first part of the research required the compilation of corpora of British legislation. It was decided to create 5 corpora from the following years: 1820, 1870, 1920, 1970, and 2020. The selected statutes were public general acts, rather than regulations or statutory instruments. Delegated legislation was excluded because the drafting style might conform to different drafting rules than primary legislation. Primary legislation is drafted by the Office of Parliamentary Counsel, whereas "delegated legislation is instructed and drafted within each Government Department" (Xanthaki 2013a: 58). Another criterion was that no amendments were to be included. The drafters of amendments tend to adopt the same language as in the original statutes, so the amendments are not necessarily representative of the specific period of time. The next criterion was purely technical – the format of the statutes was PDF, not scanned documents which would not be analysable through corpus managers. This applied primarily to the oldest legislation. Furthermore, in order to be included in the corpus, a statute had to contain at least 3 paragraphs of text – again, this seemed to be particularly relevant in the earliest stages when the statutes were often rather short. The last criterion was that statutes were used in their original "as enacted" form, not including any amendments from later peri-

ods. The question then arose as to how large the corpora had to be and whether they should be of comparable sizes. Given the substantial differences in length – new legislation tends to be much longer than old legislation – it was decided to use 100 statutes in total, 20 from each period. The whole corpus contains 974,780 tokens. The earliest sub-corpora account for about 10 % each, while the largest subcorpus represents 40 % of the whole corpus size. This composition is presumably more likely to yield representative results than a corpus which would contain only very few statutes from 2020. Overall, the results are presented in instances per million (ipm) which ensures that the results are comparable across corpora of different sizes.

3.2 Research questions

The overall aim of this paper is to show how British legislation has changed over the last 200 years with respect to certain legal language features identified by Crystal and Davy (1990), whether the trend towards plainer English tends to deprive legal texts of their lexical diversity, and whether the readability scores can confirm that British legislation has become more accessible to ordinary citizens. The corpus analysis was conducted via Sketch Engine (https://www.sketchengine.eu/).

Crystal and Davy presented a comprehensive description of legal language, discussing the layout, vocabulary, and sentence structure. In this paper we decided to focus only on vocabulary for several reasons. Firstly, a study of this size cannot offer a detailed comparison of all the features identified by Crystal and Davy (1990). Secondly, certain features may be inherently linked to the specific genre that Crystal and Davy analysed; e. g., the layout of an endowment assurance policy may be incomparable with the layout of an act of Parliament. Thirdly, the efforts of plain language exponents have largely targeted the lexis of legal documents; consequently, a diachronic comparison of the lexical changes can explain whether these efforts have yielded any tangible benefits. Fourthly, our methodology is corpus-based, and not all the features mentioned by Crystal and Davy are analysable via corpora.

Among other things, Crystal and Davy have noted the "extreme scarcity" (Crystal/Davy 1990: 202) of pronouns and anaphora, resulting in substantial repetitions of lexical items. Firstly, the analysis should show whether there is a noticeable trend regarding the use of personal pronouns over the years. Secondly, the paper seeks to find out whether British legislation has lost some of its archaic features. Crystal and Davy provided three examples of archaisms: the ending *-eth* (as in *witnesseth*), the compound adverbs of the type *hereinbefore*, and the word *aforesaid* (Crystal/Davy 1990: 206 f.). The combination of a corpus-based and corpus-driven approach concentrated on these three types of archaisms and their development over time. Thirdly, coordinated synonymous adjectives were explored, using a computer query to automatically generate coordinated items which were subsequently manually checked to exclude the items which are not synonymous (e. g., *public and private*). The last part of the analysis examines the frequencies of *shall* and *such*. In the case of *such*, the query has to be formulated so as to yield the results for use as a determiner ("unaccompanied by the indefinite article" [Crystal/Davy 1990: 206]). In Sketch Engine, the tag NN covers singular or mass nouns. Ideally, the query should not cover mass nouns (as in *such information*), but only the cases where *such* is combined immediately with a countable singular noun. Nevertheless, this drawback should not substantially distort the results, because the main focus is on the relative frequencies over time, and if mass nouns are included in all time periods, we should still be able to see the trend.

The second part of the research aims to analyse the texts in terms of their readability. For that purpose, we selected 2 readability formulas (Flesch Reading Ease Test and Dale-Chall formula) that can be used on the website https://readabilityformulas.com/. The Flesch Reading Ease Test concentrates on sentence length and word length, while the Dale-Chall formula compares sentence length and the "hard" words which are not included in the list of 3,000 "easy" words. As the website allows texts of a maximum length of 3,000 words, we used 5 random samples from each corpus.

The third part of the research explores lexical diversity. The Czech National Corpus has developed an online tool called the corpus calculator (https://www.korpus.cz/calc/) which makes it possible to determine the lexical diversity of a text, based on the type/token ratio, taking into consideration the length and type of a text. After entering the values for the number of types and the number of tokens, the tool creates a bell curve with two vertical lines: the expected value (in the middle) and the observed value. The position of the observed value indicates whether the text is lexically more or less diverse than expected for the specific type and length of text. The aim is to find out whether the most recent corpus is lexically less diverse than the oldest corpus.

4 Results

4.1 Personal pronouns

With regard to personal pronouns, Figure 1 confirms the "scarcity" of personal pronouns. If general English corpora are used as reference corpora (BNC and English Web 2020), it becomes obvious that personal pronouns are used substantially more in general English than in the legislation corpus. Over time, there is a noticeable downward trend, with the 1820 corpus containing more than twice as many personal pronouns per million (ipm = instances per million) as the 2020 corpus.

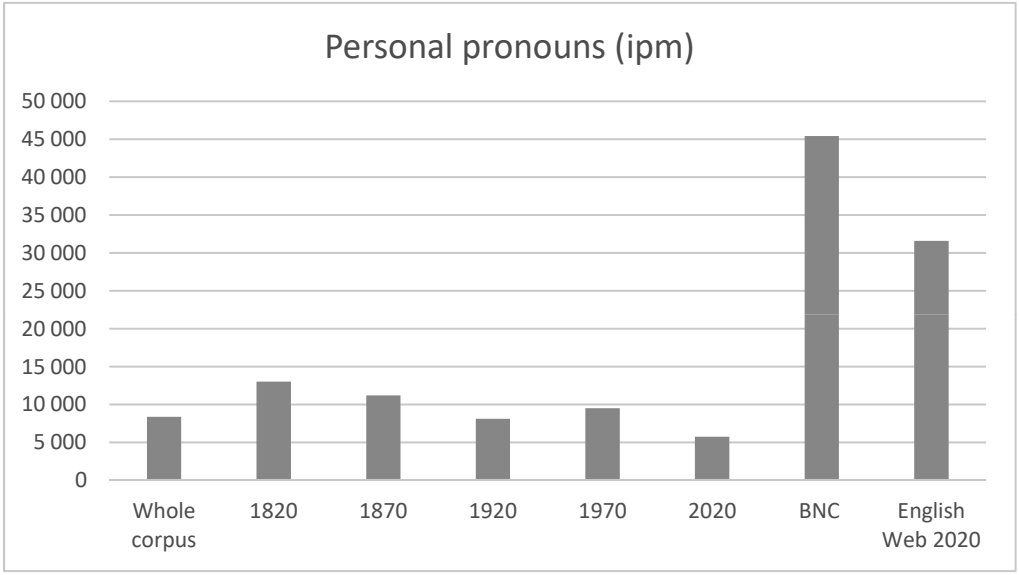

Figure 1: Use of personal pronouns in legislation and in general English

Overall, this paper confirms the finding made by Crystal and Davy (1990) that personal pronouns tend to be eschewed as a species in legal language. Furthermore, from the diachronic perspective we have noted a marked trend towards the use of fewer personal pronouns in legislation over time.

4.2 Archaisms

The first group of archaisms (the ending *-eth)* yielded no results. The ending *-th* was slightly more prevalent, covering the word forms *hath* and *doth*, all of them being from 1820 (ipm 19).

The second group of archaisms (Figure 2), derived from *here-/there-/where* + preposition, is relatively large, composed of 43 compound words. In the literature, these words are referred to as textual mapping adverbials (Gotti 2012: 56), compound adverbs or deictic pro-forms (Quirk et al. 1985: 438, 487), some of them as conjunctive adverbs (Halliday/Hasan 1976: 230), or simply as *here-* and *there-* words (Butt 2018: 286, 638). The "here" part of the word has a pronominal function (for example, *thereof* means 'of that document') rather than a locative meaning; in fact, it denotes only proximity or distance. Interestingly, although it has been established that personal pronouns "seem to be eschewed as a species" in legal texts (Crystal/Davy 1990: 202), the *here-/there-/where-* words (with a pronominal function) have flourished as characteristic features of legal texts for a long time. Over the last few decades, they have been fiercely targeted by the plain language exponents (Tiersma 2000: 94, Butt 2013: 236, Garner 2001: 401).

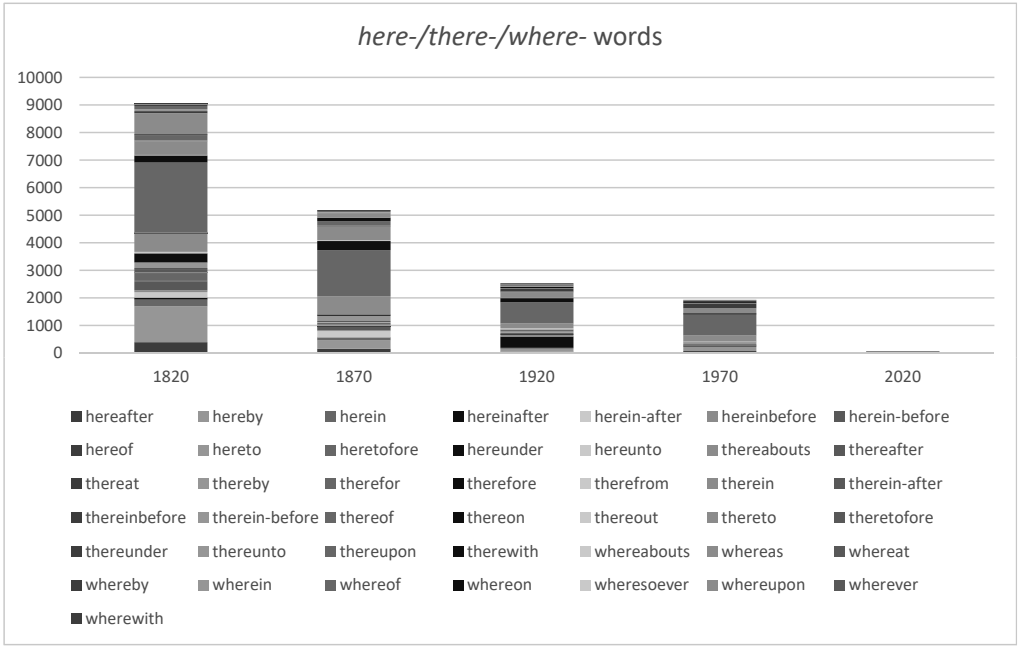

Figure 2: Use of here-/there-/where-*words in legislation from 1820 to 2020*

The distribution of the words across all 5 corpora is uneven; while the 1820 corpus totals over 9,000 *here-/there-/where*-words (ipm), the 2020 corpus contains only 54 (ipm) of them. The

chart indicates that every 50 years the numbers of archaic words halved, with the exception of the corpora from 1920 and 1970 where the changes were less distinct, and the 2020 corpus where the numbers plummeted from 1,929 *here-/there-/where-* words to 54 (ipm). Certain *here-/there-/where-* words were extremely frequent in the earliest corpora, and thus have seen the greatest declines. For example, the words *hereby, therein, thereof,* and *whereas* all decreased from very high numbers in 1820 to almost zero in 2020.

Around 1969, when Crystal and Davy's book was published, the *here-/there-/where-* words still had a relatively high frequency in our corpora. It is therefore hardly surprising that this feature was identified by those authors as characteristic of legal language. Yet our data show that even back then the drop from the earliest corpus (1820) had been very steep, from around 9000 ipm to 2000 ipm (in 1970), suggesting that the archaic *here-/there-/where-* words were already doomed at the time when *Investigating English Style* was published. Every fifty years the ipm of the *here-/there-/where-* words substantially dropped. The 2020 corpus is virtually free of compound adverbs and these words can thus no longer be regarded as a typical feature of British legislation. Does it mean that once these archaic words have been removed, the text is easier for non-experts? Masson and Waldron (1994) have undertaken a study to examine which changes make a legal text (a contract) more accessible to non-experts. They had four versions of a legal document: (1) the original, (2) with archaisms removed, (3) with plainer language – shorter sentences, simpler syntax, etc. and (4) with special terminology replaced by easier words or explained. After checking the comprehension by the subjects, the authors found that solely the removal of archaisms did not significantly improve comprehension. It seems, therefore, that the mere removal of archaisms in legislation does not guarantee greater comprehension by non-experts.

The prefix *afore-* was used only in the word *aforesaid* (Figure 3) which peaked in 1870 (over 3,000 ipm) and then steadily decreased to 0 in 2020.

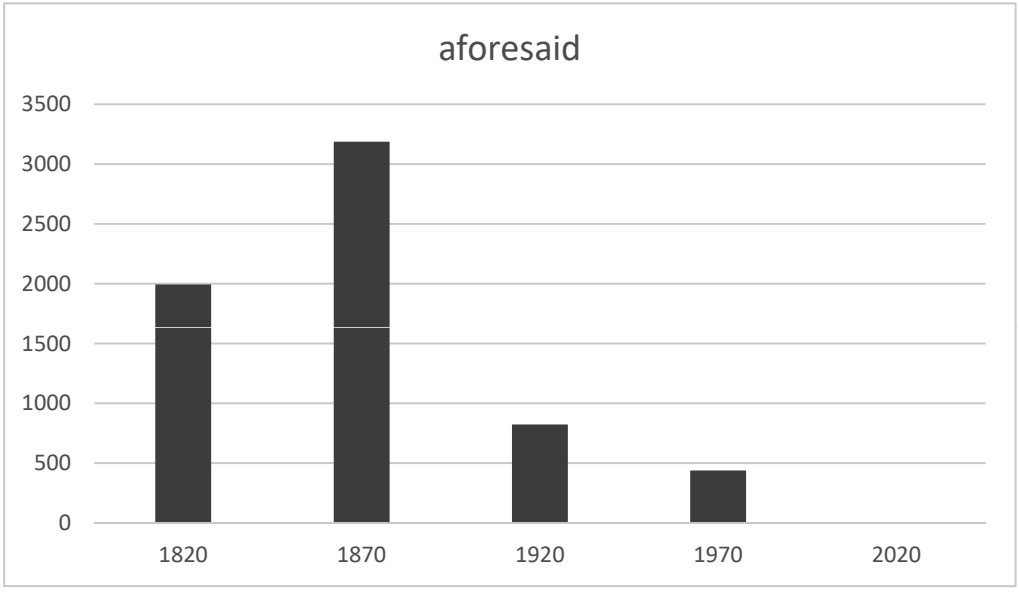

Figure 3: Use of aforesaid *in legislation from 1820 to 2020*

The word *aforesaid* was used as an adjective with the meaning 'stated before'. According to Tiersma (2000: 90), this word had been introduced into legal English from Latin (*predictus*) or Law French (*le dit*) and can be easily replaced by *the* or *this*. A typical collocation from our 1820 corpus is *the aforesaid act* (7 occurrences). The 2020 corpus contains no instances of *aforesaid* and the comparable collocations with *act* contain the words *that* or *the*.

4.3 Coordinated adjectives (binomials)

Figure 4 lists the synonymous coordinated adjectives from our corpora. The 1820 corpus contains as many as 16 distinct pairs, and every successive time period has seen a substantial reduction. The 2020 corpus has only one pair.

Coordinated adjectives	**1820**	separate and distinct	**1870**	due and proper
		valid and effectual		valid and effectual
		null and void		good and effectual
		due and payable		absolute and unfettered
		final and conclusive		usual and ordinary
		due and owing		valid and binding
		good and perfect		due and payable
		requisite and necessary		final and conclusive
		full and complete		
		just and equitable		
		further and additional		
		answerable and responsible		
		good and lawful		
		good and effectual		
		willing and desirous		
		due and unpaid		
	1920	efficient and economical	**1970**	due and unpaid
		final and conclusive		due and outstanding
		fair and equitable		final and conclusive
		suitable and convenient		
		valid and effectual	**2020**	efficient and effective
		frivolous and vexatious		

Figure 4: Use of coordinated adjectives in legislation from 1820 to 2020

In terms of meaning, these expressions have largely become lexicalised and can be treated as a single idiomatic unit. Although the binomials are traditionally not included under standard word-formation processes, they serve to enrich legal vocabulary through the "lexicalisation of syntactic phrases" (Kastovsky 2006: 209). The motivation behind their use was varied: they

could cover synonymous words of different origin (Anglo-Saxon, Latin, French) and thus facilitate communication in a multilingual society, they could serve to achieve precision by specifying the vague first item through the second item (Dobrić Basaneže 2018: 204), they could have had different meanings in the past which later converged, as in *null and void* (Mellinkoff 1963: 358), or they could serve to emphasise the ritual and ceremonial aspects of legal instruments (Mellinkoff 1963: 92). However, as Wydick argued, they are a "lawyer's tautology – a needless string of words with the same or nearly the same meaning" (Wydick 1978: 734) and as such they are discouraged in modern legal language. Semantically, these expressions often denoted *validity* and *debt*: e. g., *final and conclusive*, and *due and payable*.

4.4 Shall

The iconic[4] word *shall* (Figure 5) is a clear example of a very frequent word in legislation which became virtually extinct in the 2020 corpus. With regard to the dynamic situation around *shall*, Williams speaks of a "modal revolution" (Williams 2012: 363) and many experts on legislative drafting recommend that this "chameleon-hued word" (Garner 2001: 939) be abandoned altogether.[5] Others make the case for "a disciplined use of *shall*" (Adams 2014: 12). For example, according to the "American rule" (Garner 2001: 940), *shall* should only be used where it means "has a duty to". Other scholars maintain that "the replacement of *shall* with other forms […] does not always bring a real improvement" (Garzone 2013a: 115). Interestingly, Mellinkoff, the forefather of the plain language movement, did not specifically target *shall*. He even used it in his improved re-drafted version of a lease provision (Mellinkoff 1963: 388).

The word *shall* was particularly targeted by the plain language exponents because of its ambiguity. It can impose an obligation, grant a right, impose a condition precedent/subsequent, state a fact or assumption, and denote futurity (Garner 2001: 940, Butt 2018: 565). Others see the "degree of duty" as problematic (Kimble 1992: 61). According to Garzone, polysemy is a "distinctive property of all modals" (Garzone 2013a: 99) and thus every substitute modal for *shall* has the potential for ambiguity as well. Furthermore, Garzone challenged the view that *shall* should be used only when an obligation is imposed on a person. She argued that in addition to the deontic meaning, *shall* has a performative/constitutive value (e. g., "The declaration *shall apply to …*") (Garzone 2013a: 99), thus casting doubt on the "has a duty" test. Kurzon brought the perspective of speech acts, exploring the difference between the phrases "The Director *shall* give to the Committee …" and "The Director *has the duty* to give to the Committee …". Kurzon contends that the former case is a speech act serving as a command, while the latter ("has the duty") "is the result of an order having been given. It is in fact a description of a state of affairs" (Kurzon 1986: 22). In other words, from the perspective of speech acts, the uses of *shall* and *has a duty* are not identical.

As can be seen in Figure 5, the current British statutes are practically free of *shall*, although 50 years earlier, the ipm was still very high (almost 10,000). If we consider the entire 200-year period, the steady downward trend is apparent throughout (with a minor deviation in the 1920 corpus where the numbers are slightly higher than in the 1870 corpus), and the greatest decrease occurred between 1970 and 2020.

---

4 According to Kimble, "*shall* is the most important word in the world of legal drafting […] *shall* is the most misused word in the legal vocabulary" (Kimble 1992: 61).

5 "*Shall* has had its day" (Butt 2018: 565); "My suggestion is to abandon *shall* altogether" (Asprey 1992: 79).

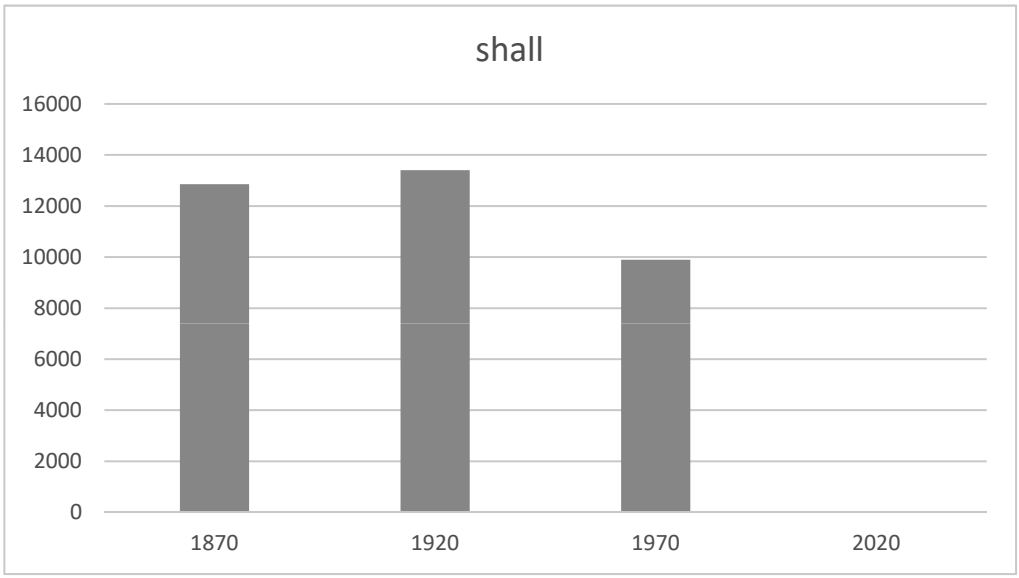

Figure 5: Use of shall *in legislation from 1820 to 2020*

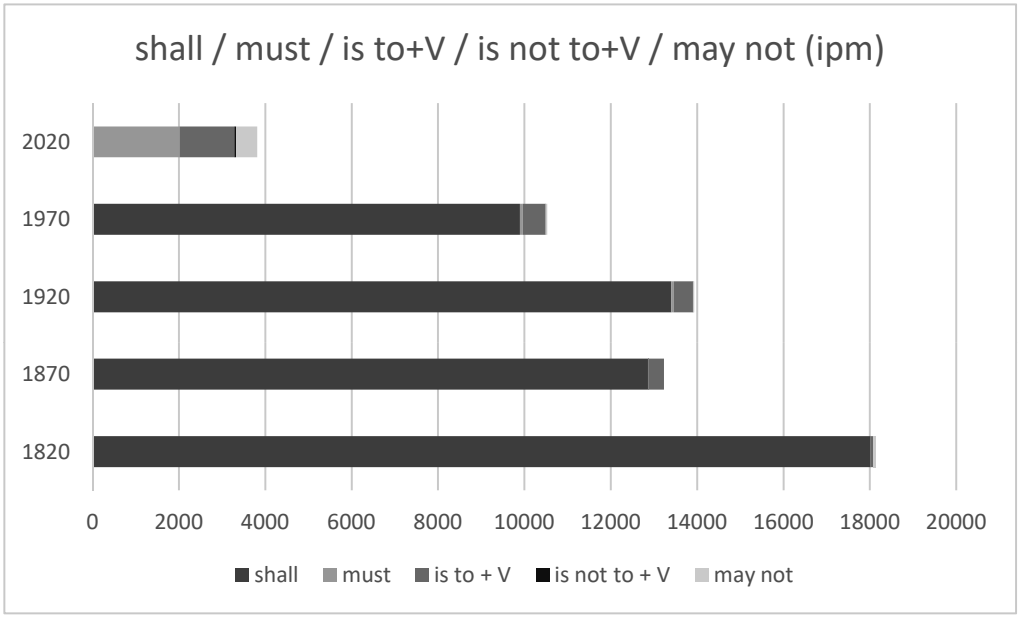

Figure 6: Use of alternatives to shall *in legislation from 1820 to 2020*

Such an enormous reduction in *shall* raises the question of whether it has been substituted with some other words, or whether it was truly redundant and could be easily removed without any substitution. The *Drafting Techniques Group Paper 19* (2008) provided a list of different uses of *shall* and the substitution strategies in various contexts. The authors suggested the following alternatives: *must* for imposing obligations, *there is to be* for the creation of statutory

bodies, and the present tense in provisions about application or effect, amendments, repeals, and other common provisions. Williams proposed the substitutes according to the "hierarchy of normative intensity" (Williams 2012: 366): *must* for maximum strength, followed by *is to / are to* and the present tense. In our corpora, we examined the use of the most straightforward alternatives to *shall*: *must, is to, may not*. It would be expected that fewer occurrences of *shall* mean substantially more occurrences of these alternatives. As Figure 6 shows, this is not the case.

While there is a considerable increase in *must, is to* + verb, and *may not* in 2020, the rows are by no means comparable in terms of their length. Thus, it is necessary to look at other substitution strategies. In many civil law countries legislation is drafted in the present tense. According to Xanthaki, "legislation does not need to repeat that its text is compulsory: irrespective of the use of an imperative form, legislation is inherently compulsory" (Xanthaki 2013b: 104). Consequently, with the decreasing use of *shall* over time, we expect to find an increased use of the present tense. Arguably the best way to determine the use of the present tense is to focus on the 3rd person singular (with an *-s*). The corpus analysis yielded the following results.

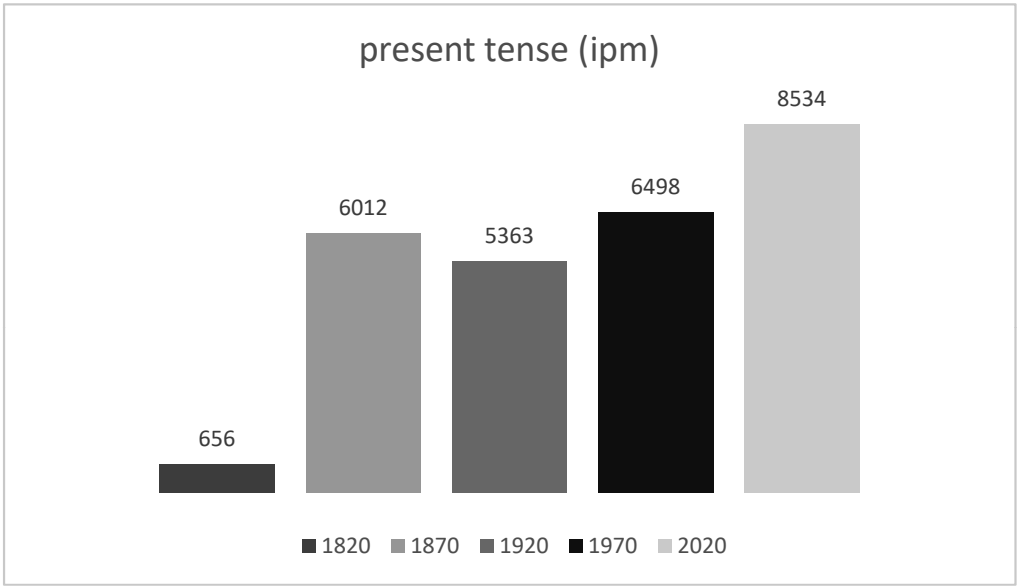

Figure 7: Use of the present tense (3rd person singular) in legislation from 1820 to 2020

As can be seen in Figure 7, in the 1820 corpus the present tense (the third person singular) was very rare, and within 50 years it had become ten times more frequent. In fact, the slight deviation in 1920 (a decreased use of the present tense) corresponds to the deviation in Figure 6 where the 1920 corpus shows an increased use of *shall*. Similar findings over a 50-year period were identified by Garzone: "In theoretical terms, the simple present is the best substitute for *shall*" (Garzone 2013a: 109). Yet Garzone (2013b: 77) gives an example of a provision in which the replacement of *shall* with the present tense changes the meaning: "There *shall* be a body corporate …" is different from "There is a body corporate …" in that the latter case implies that the body is already in existence. Overall, Garzone (2013b: 79) argues that the replacement of

shall with other forms poses difficulties because the substitutes can have a slightly different meaning from *shall* and can be fuzzy and ambiguous as well.

4.5 Such

Figure 8 shows the use of the word *such* as a determiner (without an indefinite article). It peaked in the 1870 corpus, and is probably heading towards extinction in British legislation. Although the values in the 2020 corpus are still significant (498 ipm), compared to the initial values in the 1820 corpus (4814 ipm), this represents a tenfold drop.

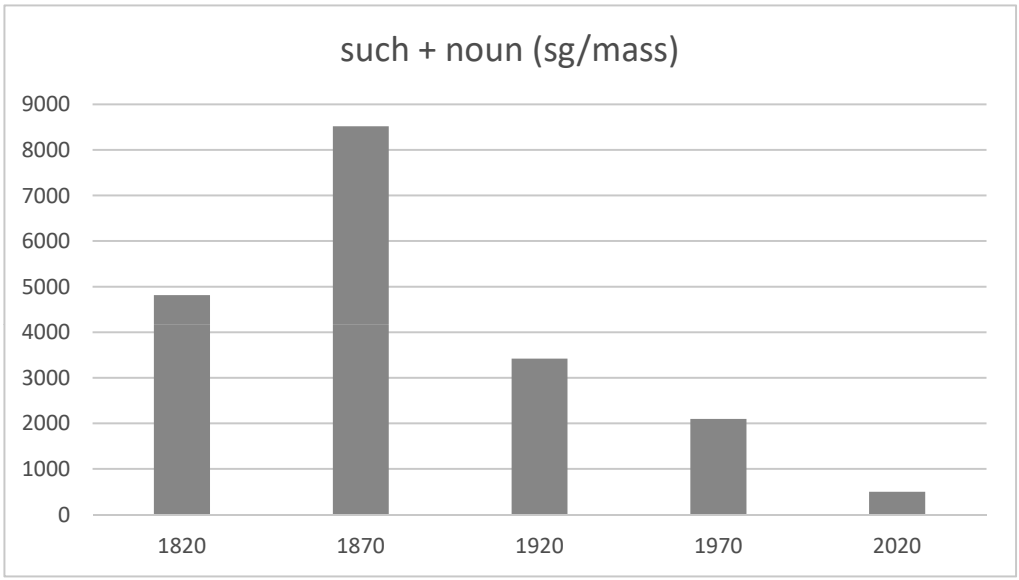

Figure 8: Use of such *without an indefinite article from 1820 to 2020*

The substantial decline shows that the "precision of reference" which was suggested by Crystal and Davy as a possible reason for using *such* was a myth, and precision could be easily achieved by other means, for example the definite article.

4.6 Readability

The readability scores for the Dale-Chall formula and the Flesch Reading Ease formula (5 samples for each corpus) were rather inconclusive and failed to prove that the current British legislation is more readable than the old statutes. Overall, the readers of British legislation need tertiary education (Grades 13–15, 16+) to understand the texts (18 samples out of 25) which indicates that the texts cannot be classified as readable. The clearest texts seem to be those from the oldest corpora.

The two readability formulas thus failed to confirm the hypothesis that the newer legislative texts, which are arguably plainer in terms of their lexis, are more readable. Several factors could account for this:

(1) Contemporary society deals with more complex and technical issues than in the 1820s, and the law needs to reflect this. In this way, while certain "glue words" (Wydick 1978: 729) have been removed from legislation, the "working words" and terminology may have become more sophisticated and technical. Likewise, the complexity and interrelatedness of legal concepts may require more complex sentence structures.
(2) While in recent decades legislative drafters have focused on the vocabulary of the legal texts to make it more compliant with plain language principles, sentence structure may have been slightly neglected. Thus, it could be argued that the mere substitution of individual words does not necessarily ensure greater readability. For example, Garzone maintains that the replacement of *shall* "is an easily accomplished change within the context of a much more complex and long overdue process of simplification of legal texts, while more profound improvements at the level of syntax and text organization are certainly more difficult to achieve" (Garzone 2013b: 79). In addition, there could be some other factors that present difficulties for the layperson. Azuelos-Atias (2018) mentioned three problematic areas: technical vocabulary, syntactic complexity, and intertextuality (implied background knowledge) which requires "detective work" (Azuelos-Atias 2018: 106) on the part of the lay person.
(3) Readability formulas have their limitations – generally they take into account the length of sentences and the length of words, but a long sentence (if properly structured) need not necessarily pose substantial difficulties for the reader.

From the synchronic perspective, all British legislation as a whole seems to be accessible primarily to a university-educated readership. So, if an average adult reads at the 7th-grade level (DuBay 2004: 1), British legislation is largely inaccessible to him/her. This begs the question of addressees: whom are the statutes drafted for? Even experts do not seem clear on this issue, and over time opinions have changed to a certain degree. For example, 70 years ago Driedger believed that the addressees of legislation were professionals and experts, not ordinary citizens: "It must not be supposed, however, that statutes can be written so that everyone can understand them" (Driedger 1949: 295). The more recent approach seems to take the view that if everyone is presumed to know the law, then the law should be written in a form accessible to everyone. Thus, many promoters of plain legal language (e. g., Tiersma, Butt, Garner) subscribe to Mellinkoff's view that "the language of the law should not be different without a reason" (Mellinkoff 1963: 285). But the fundamental question is whether legislation *can* indeed be written for the average adult who reads at a 7th-grade level. After all, a legislative text must meet many criteria that other types of text do not have to fulfil.

4.7 Lexical diversity

Figure 9 shows the lexical diversity of the oldest (1820) and the most recent (2020) corpus, with the expected value (the vertical line in the middle) and the observed value (the vertical line on the left). Both corpora are lexically less diverse than expected for this type of text (as evidenced by the shorter line which appears on the left rather than on the right). From the diachronic perspective, the most recent corpus (2020) is lexically less diverse than the 1820 corpus, with the observed value being about one third of the expected value.

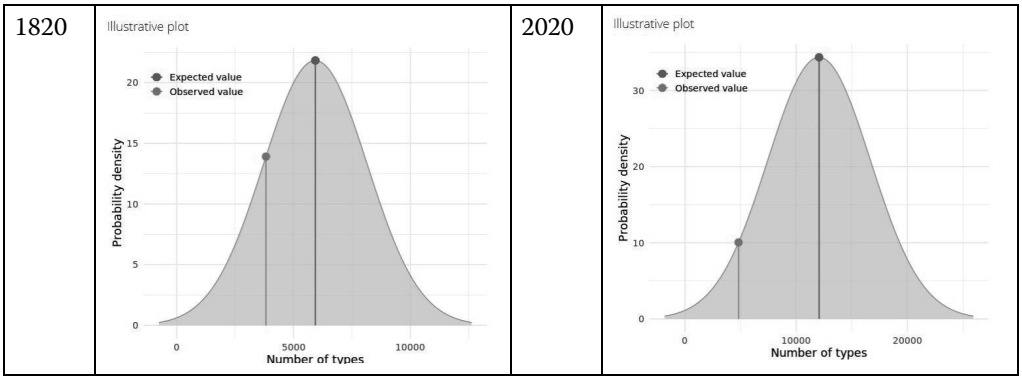

Figure 9: Lexical diversity of the 1820 and the 2020 corpora

The *Drafting Guidance* of the Office of the Parliamentary Counsel mentions the advice given by Sir Ernest Gowers[6] on the choice of words in legislation: "use the most familiar words" (Office of the Parliamentary Counsel 2020: 5). By limiting the choice of words to the most familiar words and by consciously avoiding traditional and archaic words, the lexical diversity of a text can be affected. This was confirmed in this paper, as the most recent corpus (2020) is clearly lexically less diverse than the older corpus. Nevertheless, it would require further research to find out whether plain language generally results in a lexically less diverse text, as there could be some other factors at play.

5 Impact of plain language

Most of the substantial changes that have moulded the language of British legislation into its present form seem to be part of a long trend stretching over the past 200 years. Thus, already at the time when Crystal and Davy were writing their seminal book, some of the features identified by them as typical of legal language were already doomed and on their way to extinction (at least as far as British legislation is concerned), probably due to consistent language planning.[7] However, the greatest change for many of the analysed features was made between 1970 and 2020. The use of the *here-/there-/where-* words fell steadily from 1820 onwards, but the presence of such words in legislation was nevertheless substantial until 1970 (1,929 ipm). Only in the 2020 corpus can we see that this group of words was almost completely eradicated. The same applies to *aforesaid*, whose presence was steadily declining, but some presence of *aforesaid*, albeit minor, was still expected even in 2020. Yet the 2020 corpus does not contain a single instance of *aforesaid*. The word *shall* also confirms that considerable acceleration[8] took place after 1970; in the 1970 corpus, the presence of *shall* was high (9,898 ipm), while in the 2020 corpus there were only 23 instances per million. This is probably no coincidence; since

6 The author of *Plain Words: A Guide to the Use of English*, published in 1948, well before the plain language movements took the form of a concerted effort.

7 See Williams: "Prescriptive engineering", "the changes may have occurred as a result of nurture rather than nature" (Williams 2012: 356).

8 Garzone and Williams do not seem to agree on the moment when the greatest reductions in *shall* occurred; while Williams (2012) places the greatest reductions between 2001 and 2010, Garzone suggests that "the decline may have started before 2001" (Garzone 2013b: 70).

the 1970s plain language exponents have globally exerted a considerable influence on legal language, seeking to remove certain problematic features which could impede comprehension by non-lawyers.

The language of legislation can be substantially influenced from the top, by means of drafting guidelines. In the UK, plain language rules have found their way into the *Drafting Guidance* of the Office of the Parliamentary Counsel. In addition to explicitly referring to certain promoters of plain language (e. g., Asprey), the guide advises on how to use certain traditional legal words. For example, the version from June 2020 expressly recommends avoiding *shall* and archaisms (Office of the Parliamentary Counsel 2020: 4 f.). Since British primary legislation is produced centrally via the Office of Parliamentary Counsel, the *Drafting Guidance* certainly plays an important role in ensuring a high degree of standardisation. In fact, it appears to be a very effective tool for introducing changes to the language of legislation, and thus, compared to other genres of legal language, the language of legislation seems to be rather dynamic.

Another important factor is the entire system of legislative drafting in common law countries where legislation tends to be drafted in one central drafting unit (Stefanou 2016). In the case of the United Kingdom, it is the Office of Parliamentary Counsel. Logically, if all the legislative drafts originate from a single drafting unit, a high degree of standardisation can be relatively easily achieved, and if there is consensus about the introduction of plain language elements into legislation, this change can be quite rapidly implemented in this centralised environment.

6 Conclusion

This paper sought to explore certain lexical features of legal language, as presented by Crystal and Davy in their seminal work *Investigating English Style,* in British legislation from five different periods (1820, 1870, 1920, 1970, and 2020). Our analysis showed that many features identified by Crystal and Davy as typical of legal language (e. g., archaisms, *shall*, the word *such* used as a determiner, pairs of synonymous adjectives) are virtually non-existent in current British legislation. The data from the 2020 corpus show that external factors must be behind the accelerated disappearance of some of the features. These external factors are most likely the plain language efforts which have exerted considerable influence on the language of legislation via the *Drafting Guidance* of the Office of the Parliamentary Counsel.

Because plain language promoters often tend to recommend restricting the choice of vocabulary to the most frequently used words, this paper also explored the lexical diversity of the oldest and the most recent corpus to find out whether the development of legal language has brought about lexically less diverse texts. This hypothesis was confirmed, as the 2020 corpus tends to be lexically less diverse than the 1820 corpus.

The last part of our research concentrated on the readability of British legislation to find out whether the changes made to the language of legislation have made British statutes more accessible to non-lawyers, as far as the readability measures can show. The readability scores from the five samples from each corpus were rather inconclusive (in fact, some scores cautiously suggested that the older texts were more readable), thus the only tentative conclusion that can be drawn in this respect is that (1) the current legislative texts may deal with more technical matters requiring more specialised (and less readable) language; (2) the recent plain language efforts may have focused on vocabulary, neglecting sentence structure and other issues; and (3) readability formulas have their limitations and cannot provide reliable results.

Further research is needed to objectively measure the accessibility of legislation to non-lawyers.

It follows from our analysis that the recent decades have seen dramatic changes in the language of British legislation, which may have affected perceptions about the typical features of legal language. These changes are largely due to language planning efforts. It remains to be seen whether such changes will gradually trickle down to private legal documents as well. Likewise, it will be interesting to see whether changes to the register will gradually affect the perceptions of law students and the general public who often still see legal language as the one presented by Crystal and Davy fifty years ago.

References

Adams, Kenneth A. (2014): "Banishing Shall from Business Contracts: Throwing the Baby out with the Bathwater." *Australian Corporate Lawyer* 24: 12 f.

Asprey, Michele M. (1992): "Shall must go." *The Scribes Journal of Legal Writing* 3: 79–83.

Azuelos-Atias, Sol (2018): "Making Legal Language Clear to Legal Laypersons." *Legal Pragmatics*. Eds. Barbara Kryk-Kastovsky and Dennis Kurzon. Amsterdam: Benjamins. 101–115.

Balmford, Christopher (2002): "Plain Language: Beyond a Movement: Repositioning Clear Communication in the Minds of Decision-makers." *Plainlanguage.gov website:* https://www.plainlanguage.gov/resources/articles/beyond-a-movement/ (31.08.2022).

Benson, Robert W. / Kessler, Joan B. (1987): "Legalese v. Plain English: An Empirical Study of Persuasion and Credibility in Appellate Brief Writing." *Loyola of Los Angeles Law Review* 20.2: 301–322.

Butt, Peter (2013): *Modern Legal Drafting: A Guide to Using Clearer Language.* Cambridge University Press.

Butt, Peter (2018): *Legal Usage. A Modern Style Guide*. LexisNexis Butterworths.

Cheek, Annetta (2010): "Defining Plain Language." *Clarity* 64: 5–15.

Crystal, David / Davy, Derek (1969): *Investigating English style*. London: Routledge.

Crystal, David / Davy, Derek (1990): "The Language of Legal Documents." *Investigating English Style*. Eds. David Crystal / Derek Davy. 12th impres. Harlow: Longman. 193–217.

Cvrček, Václav / Chlumská, Lucie (2015): "Simplification in Translated Czech: A New Approach to Type-token Ratio." *Russian Linguistics* 39.3: 309–325.

Dobrić Basaneže, Katja (2018): "Extended Binomial Expressions in the Language of Contracts." *Phraseology in Legal and Institutional Settings*. Eds. Anne Wagner / Vijay Kumar Bhatia / Stanisław Goźdź-Roszkowski / Gianluca Pontrandolfo. London: Routledge. 203–220.

Drafting Techniques Group (2008): *Shall*. Paper 19 (final).

Driedger, Elmer A. (1949): "Legislative Drafting." *Canadian Bar Review* 27: 291–317.

DuBay, William H. (2004): *The principles of Readability*. ERIC Clearinghouse. https://files.eric.ed.gov/fulltext/ED490073.pdf (04.10.2022).

Garner, Bryan A. (2001): *A Dictionary of Modern Legal Usage*. 2nd ed. New York: Oxford University Press.

Garzone, Giuliana (2013a): "Modality and Performativity in Legislative Texts: The Case of Shall." *The Three Waves of Globalization: Winds of Change in Professional, Institutional and Academic Genres*. Eds. Franca Poppi / Winnie Cheng. Newcastle upon Tyne: Cambridge Scholars Publishing. 94–122.

Garzone, Giuliana (2013b): "Variation in the Use of Modality in Legislative Texts: Focus on Shall." *Journal of Pragmatics* 57: 68–81.

Gopen, George D. (1987): "The State of Legal Writing: Res Ipsa Loquitur." *Michigan Law Review* 86.2: 333–380.

Gotti, Maurizio (2012): "Text and Genre." The Oxford Handbook of Language and Law. Eds. Peter Meijes Tiersma / Lawrence M. Solan. Oxford / New York: Oxford University Press. 52–66.

Halliday, Michael Alexander Kirkwood / Hasan, Ruqaiya (1976): *Cohesion in English*. London: Longman.

Kastovsky, Dieter (2006): "Vocabulary." *A History of the English Language*. Eds. Richard Hogg / David Denison. Cambridge: Cambridge University Press. 199–270.

Kimble, Joseph (1992): "The many Misuses of Shall." *The Scribes Journal of Legal Writing* 3: 61–77.

Kurzon, Dennis (1986): *It is hereby performed … Explorations in Legal Speech Acts.* Amsterdam: Benjamins.

Kurzon, Dennis (1997): "'Legal Language': Varieties, Genres, Registers, Discourses." *International Journal of Applied Linguistics* 7.2: 119–139.

Long, Lance N. / Christensen, William F. (2011): "Does the Readability of Your Brief Affect Your Chance of Winning an Appeal? An Analysis of Readability in Appellate Briefs and its Correlation with Success on Appeal." *The Journal of Appellate Practice and Process* 12.1: 145–162.

Masson, Michael E. J. / Waldron, Mary Anne (1994): "Comprehension of Legal Contracts by Non-experts: Effectiveness of Plain Language Redrafting." *Applied Cognitive Psychology* 8.1: 67–85.

Mellinkoff, David (1963): *The Language of the Law*. Eugene: Wipf and Stock. (June 2020). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/892409/OPC_drafting_guidance_June_2020-1.pdf (03.09.2022).

Palyga, Steve (1999): "Is it Safer to Use Legalese or Plain English? What the Judges Say." *Clarity* 43: 46–51.

Phillips, Alfred (2003): *Lawyers' Language: How and why Legal Language is Different*. London: Routledge.

Quirk, Randolph / Crystal, David / Greenbaum, Sidney / Svartvik, Jan / Leech, Geoffrey N. (1985): *A Comprehensive Grammar of the English Language*. London: Longman.

Šlerka, Josef / Smolík, Filip (2010): "Automatická měřítka čitelnosti pro česky psané texty." *Studie z aplikované lingvistiky / Studies in Applied Linguistics* 1: 33–44.

Sneddon, Karen (2011): "Speaking for the Dead: Voice in Last Wills and Testaments." *St. John's Law Review* 85.2: 683–754.

Stefanou, Constantin (2016): "Comparative Legislative Drafting: Comparing across Legal Systems." *European Journal of Law Reform* 18.2: 123–138.

Tiersma, Peter M. (2000): Legal Language. Chicago: The University of Chicago Press.

Williams, Christopher (2006): "Fuzziness in Legal English: What Shall we Do with Shall?" *Legal Language and the Search for Clarity*. Eds. Anne Wagner / Sophie Cacciaguidi-Fahy. Bern: Lang. 237–264.

Williams, Christopher (2012): "Changes in the Verb Phrase in Legislative Language in English." *The Verb Phrase in English: Investigating Recent Language Change with Corpora*. Ed. Bas Aarts. Cambridge: Cambridge University Press. 353–371.

Williams, Christopher (2015): "Changing with the Times: The Evolution of Plain Language in the Legal Sphere." *Revista Alicantina de Estudios Ingleses* 28: 183–203.

Wydick, Richard C. (1978): "Plain English for Lawyers." *California Law Review* 66.4: 727–765.

Xanthaki, Helen (2013a): "Legislative Drafting: A New Sub-Discipline of Law Is Born." *IALS Student Law Review* 1.1: 57–70.

Xanthaki, Helen (2013b): *Thornton's Legislative Drafting*. London: Bloomsbury.

Eva Dvořáková
Faculty of Law, Charles University in Prague
nám. Curieových 7, 116 40 Staré Město,
Faculty of Arts, Charles University in Prague
nám. J. Palacha 1/2, 116 38 Staré Město,
e-mail: dvorakoe@prf.cuni.cz