

A CORPUS-BASED STUDY OF DEONTIC MODALITY IN ENGLISH EUROLECT

Annalisa Sandrelli

(Università degli Studi Internazionali di Roma - UNINT, Italy)

Abstract

The law-making environment of the European Union is a very specific language contact setting in which English has been used as the main drafting language since the 2004 round of enlargement. This situation has produced a specific variety of legislative English, namely English Eurolect. Although its most striking characteristic is the presence of specific EU-related vocabulary, there are also notable morpho-syntactic and textual traits, such as differences in modal usage. The English section of the *Eurolect Observatory Multilingual Corpus* (EOMC) includes an A corpus of EU directives, a B corpus of matching UK national transposition measures and a recently added C corpus of UK domestic legislation that is unrelated to the EU setting. The comparison between A and B has already highlighted significant differences in relation to obligation, permission and prohibition modals (Sandrelli 2018). The present study, based on a mixture of quantitative data (obtained via *Wordsmith Tools* 6.0) and qualitative observations, aims to obtain a more fine-grained view of such differences by adding the analysis of the C corpus of domestic legislation.

1. Introduction

EU legislation is the outcome of extensive negotiations involving various stakeholders in Brussels and in the Member States, via a complex multilingual legal co-drafting process which produces equally authentic versions of every text in all the EU languages. As Wagner *et al.* (2002: 70) point out, “[...] in most cases the authors are unidentifiable: the texts are collectively produced with disparate input from various sources, in the process of consensus formation and political compromise”. While in legal terms there is no ‘original’ source language version as such, in actual practice the text is drafted in one language, altered several times in an iterative process, translated into all the other languages and revised.

This unique law-making process taking place in a multicultural and multilingual scenario has produced specific EU legislative varieties, known as *Eurolects*. In 2013 the Eurolect Observatory was set up at Università degli Studi Internazionali-UNINT in Rome with the aim of describing the key features of Eurolects in several EU languages

through corpus data¹. More specifically, the project is focused on one of the types of binding legal instruments adopted by the EU, namely directives. The reason for this choice is that, while regulations and decisions are directly applicable, directives require transposition into national law before they can be implemented in each Member State; this mechanism creates a close link between directives and the related national transposition measures that is very interesting from a linguistic point of view as well as a legislative one. Indeed, Robertson (2010: 149) points out: “There is thus a direct relationship between EU language and national legal language. They are separate kinds of discourse, and legally and linguistically distinctive, but closely linked and intertwined”.

During the first phase of the project (2013-2016), an A corpus of EU directives covering a 10-year time span (1999-2008) was collected in all the languages involved in the Observatory. It was decided to include all the directives adopted over that period (i.e. those adopted by the European Parliament and Council and those adopted by the Commission), as they all require national transposition to become applicable in the Member States². Likewise, it was decided to include both basic acts and amending acts, which contain text that must be inserted into older acts. Although the drafting style of amending acts is influenced by that of earlier acts, it was decided to include them in the corpus too, as they make up a significant proportion of the directives adopted each year and our aim was to obtain as full a picture of the directives adopted over the ten-year period as possible. Then, a B corpus of all the related national transposition measures (NTMs) adopted by the various Member States was compiled for all the languages involved in the project³. More specifically, “[...] only the laws that were directly promulgated by the national Parliament or were delegated to the Council of Ministers were taken into account as first choice. Consequently, both local norms and ministerial regulations were not included [...]” (Tomatis 2018: 36). As regards the time span of legislation included in the B corpus, it stretches up to 2013, as the transposition process often requires several years.

The resulting Eurolect Observatory Multilingual Corpus (EOMC) is a comparable corpus of EU directives and matching national transposition laws in 11 languages. The Observatory team adopted a common research template that distinguished between EU-rooted phenomena, contact-induced features, and intra-linguistic variability at lexical, morpho-syntactic and textual levels (Mori 2018b). By comparing directives with their matching NTMs via corpus linguistics techniques, it was possible to identify several Eurolect-specific traits, i.e. features that were either not present or were used differently in the domestic varieties of legislative language; the results of this phase of the project have been published in an edited volume (see Mori 2018a). However, as the

¹ Initially, the project included English, French, German, Italian, Maltese and Spanish, but the range of EU languages was expanded by adding Dutch, Finnish, Greek, Latvian, and Polish as soon as partners in other universities were identified; Hungarian has recently been added too. For more information on project partners, research activities and publications, see <https://www.unint.eu/it/ricerca/progetti-direcra/1219-eurolect-observatory-project.html>.

² I wish to thank one of the anonymous reviewers for pointing out that only the directives adopted by the European Parliament and the Council undergo the full EU legislative process and have been checked by the lawyer-linguists of the three EU institutions; it might be interesting to investigate language patterns in the two types of directives in a future study.

³ Information was retrieved from the Eur-Lex portal: <https://eur-lex.europa.eu/homepage.html?locale=en>.

purpose of the national laws in the B corpus is to transpose the provisions contained in EU directives, they cannot be considered as a ‘pure’ product of the legislative drafting tradition of each Member State. Therefore, in the second phase (2017-2020) of the project a C corpus of domestic legislation that was originated entirely domestically (i.e. bearing no relation to the EU setting) was added for English, German, Greek, Italian, Polish and Spanish (at the time of writing). By comparing patterns identified in the three components (A, B and C), it should be possible to obtain a more fine-grained view of Eurolect-specific usage versus the domestic legislative variety of the language.

As regards the English section of the EOMC, one of the aspects in which interesting differences emerged from the comparison between the A corpus and B corpus is usage patterns in deontic modality (Sandrelli 2018: 79-83), one of the key linguistic features of prescriptive legal texts. Thus, the aim of the present contribution is to refine the analysis by adding a comparison with the C corpus of domestic legislation. Our hypothesis is that there is a Eurolect-specific way of using obligation, permission and prohibition modals and semi-modals in EU directives, and that differences will be identified not only between the corpus of directives and the two corpora of UK legislation, but also between the corpus of NTMs (B) and the corpus of domestic laws with no relation to the EU setting (C). The paper begins with a literature review on deontic modality in English legislative language (Section 2); then, my data and methodology are briefly presented in Section 3, the results are illustrated in Section 4, and my conclusions in Section 5.

2. Deontic modality in legal language: a brief overview

The frequency of deontic modals and semi-modals is relatively high in legislation, because “[t]he fundamental functions of law, that is, to impose duties and confer power [...] are realised through deontic modality – modals and related patterns which convey obligation and permission” (Biel 2017: 157). More specifically, the duty-imposing function of laws is expressed via deontic obligation modals, while the power-conferring function is conveyed by deontic permission modals. Traditionally, the most widely used obligation modal is *shall*, which conveys both futurity and obligation. The same nuances are also expressed by the other deontic modals used in prescriptive texts:

One of the major characteristics of deontic modality is precisely the fact that, unlike epistemic modality, it is intrinsically connected with ‘futurity’ [...]. This obviously applies not only to *shall*, but also to the other modal auxiliaries to be found in prescriptive legal texts, notably (in terms of frequency) *may*, as well as *must*, *should* and *can*. In short, when prescribing what people may or may not do, the law cannot but regulate their behaviour or situation *prospectively* with respect to the present moment (Williams 2005: 87).

Over the last few decades, the use of modals in legislation has undergone profound changes in English-speaking countries, as a result of the Plain Language (or Plain English) movement which originated in the US in the 1970s, and then spread to the UK and other English-speaking countries (Williams 2007)⁴. The general aims of this modern-

⁴ Of course, this is not to say that the importance of writing clearly and avoiding *officialese* was never taken into account before then (see for example Ernest Gowers, *Plain Words*, 1948).

ization drive are to make legal texts more accessible to ordinary citizens by removing several features of *legalese*. One of them is the modal *shall*, considered a superfluous archaism to be replaced with “[...] *must* or the semi-modal *is/are to* construction (as in *There is to be a body corporate*) or the present simple” (Williams 2011: 40).

Williams (2005) compiled the World Data Corpus (about 145,000 words), made up of 36 prescriptive texts from the UK, EU, US, UN and ILO, Australia, Canada, South Africa, Ireland, and New Zealand; all the legislation was adopted between the 1990s and the first few years of the 21st century. The most widely used modal in the corpus was *shall*, with a frequency of 11,200 per million words, although the most common prescriptive verbal construction was actually the present simple (21,500 per million words)⁵. However, Williams also found examples of *shall-free* legislation adopted in Australia and South Africa and hypothesized the existence of a North-South divide in this respect, namely that English-speaking countries in the southern hemisphere had fully embraced the *shall-free* policy, unlike those in the northern hemisphere and unlike the major international organisations.

In the UK the first steps towards the modernization of legislative drafting were taken with the Revenue and Customs Tax Law Rewrite Project, which began in 1997 and lasted 13 years; moreover, both the Office of the Parliamentary Counsel in Westminster and the Office of the Scottish Parliamentary Counsel in Edinburgh were involved in drawing up drafting guidelines to improve the linguistic quality of legislation. Williams (2007, 2008) analysed the text of the 2007 Income Tax Act and a number of Acts passed by the Scottish Parliament between 2000 and 2007 to identify relevant style changes. In relation to deontic modality, Williams (2007: 108-109) found that in the Income Tax Act “[...] a considerable portion of the text is also free of the controversial modal auxiliary that so typifies traditional legal texts, i.e. *shall*”; *shall* was generally replaced by *must*, the *be to* construction and the present tense. As regards Scotland, two sets of Scottish Acts were compared, the first one including all the Acts passed in 2000 and the second one most of the Acts passed in 2006-2007. In the later set of laws, the frequency of *shall* had decreased by almost 80%, while the use of *must* had more than doubled. *Shall* was retained especially in those parts of the documents that amended older Acts: “Drafters often adopt the style of the Act being amended while making textual changes and so they continue to use *shall* in these circumstances even if they use *must* or other formulations elsewhere” (Williams 2007: 114).

Williams (2013) continued his investigation of prescriptive verbal constructions in legislation by collecting five subcorpora of UK legislative texts passed in 1970, 1980, 1990, 2000 and 2010, and a corpus of EU legally binding texts adopted over the same period. This diachronic study shows that in the UK the frequency of *shall* went from 12,700 (per million words) in 1970 to 200 in 2010; by contrast, in the EU its frequency remained practically unaltered until the turn of the 21st century and then dropped slightly.

Garzone (2013a) also focused her analysis on the UK context. Firstly, she argued that it is reductive to consider *shall* as a pointless archaism; *shall* can have a performative (constitutive) function as well as a deontic one, and the two have an element in

⁵ All the frequency values identified by the various scholars cited in this section have been normalized to 1 million words for consistency and to make them directly comparable with our own data.

common, namely “the guarantee that the relevant action will be accomplished, either by means of the imposition of an obligation the fulfilment of which is guaranteed, or through the immediate realization of the action predicated” (Garzone 2013a: 79). Secondly, Garzone carried out a diachronic study on modal frequency by collecting four subcorpora of UK General Acts of about 1 million words each (1973-74, 1989-90, 2005-06 and 2010-11). She found that the use of modals remained more or less unaltered until the turn of the 1990s; then, between 1989 and 2005 there was a five-fold decrease in the frequency of *shall*, followed by a further five-fold decrease in the following five years up to 2010-11. The decrease in *shall* was accompanied by a parallel increase in the use of *must* (over 17 times more frequent in 2010-2011 in comparison with 1973-1974) and an increase in the use of *is to/are to* (3.5 times more frequent over the same period of time). In order to determine the timing of such changes more accurately, Garzone also analysed a corpus of all the Finance Acts passed between 1989 and 2012 (over 3.7 million words); as this kind of Act is passed once a year, it can be used as an indicator of the ‘switch’. Her analysis shows that, although the use of *shall* decreased constantly, the decline became especially marked after 2005, until its frequency became almost 0 in the 2012 Finance Act. It can be concluded that the UK drafting reforms have certainly had an effect in relation to the use of modals. However, Garzone (2013a: 79) also argues that “the suppression and replacement of *shall* in legislative drafting is not as unproblematic as it is often presented to be in the relevant literature. Firstly, not all substitutes have exactly the same meaning as *shall* [...]. Secondly, if the meaning of *shall* has been shown to be characterized by a degree of fuzziness, it is also true that most *shall* substitutes are not totally free from fuzziness or pragmatic ambiguity either [...]”.

In another study, Garzone (2013b) investigated deontic modals in three corpora, one of UK General Acts (just over 1.1 million words), another one of US Federal Acts (same size) and an EU corpus of directives and regulations (1 million words). She found further empirical evidence of the sharp decrease in the use of *shall* in the UK, while neither the US nor the EU seemed to have been affected by the changes in the use of modals; in parallel, the other modal forms replacing *shall* were found to have much higher frequency in the UK corpus than in the other two. Indeed, her study “clearly indicates a situation where the use of *shall* in legislative discourse is very uneven in different legal systems, with Britain using it very sparingly, and the other two jurisdictions considered utilizing it much more generously [...]” (*ibid.*: 113).

In relation to the EU, the existing literature does seem to confirm resistance to change in relation to the use of modals. Foley (2001) compiled the EULEG corpus, comprising one Treaty, four regulations, four directives and two decisions (about 160,000 words in total). Foley studied the distribution of modals across the three main textual sections, namely preambles, enacting terms and annexes, and found that *shall* is especially frequent in the enacting terms, while *must* is more common in the preambles. As regards *shall*, he also investigated usage in a small sample of occurrences from the enacting terms; more specifically, he studied the frequency of the modal with a human subject or an inanimate one, and whether it appears in active or passive sentences. Only 40% of the occurrences analysed involved a human subject; therefore, Foley (*ibid.*: 192) argues, “[I]f one adopts relatively strict criteria for agency, e.g., that the logical and grammatical subject coincide and the subject be human, 60% of the occurrences of

shall analyzed are unmotivated [...]. If the passives are included as agents recoverable from context, then some 45% of the occurrences of *shall* in this data are unmotivated [...].”

The EUROFOG corpus was compiled by Biel (2017) to study both Polish and English Eurolect. The English section of the corpus is made up of the JRC Acquis in English (53.8 million words)⁶, which includes a subcorpus of regulations (15.8 million), a subcorpus of directives (just over 8.7 million) and a subcorpus of miscellaneous instruments, including decisions, recommendations, opinions, and so on (Biel 2014). Biel compared the frequency of modals in regulations and directives. Starting with obligation modals, *shall* was found to account for two thirds of all obligation modals in the JRC Acquis corpus, once again confirming its popularity in the EU setting. Moreover, its frequency was higher in directives (10,626 per million words vs. 8,879, respectively); the frequency of *must* was also higher in directives. Biel explains this result by correlating it with the fact that regulations are immediately applicable and are addressed directly to their addressees, while directives impose obligations on Member States, which will then need to transpose them into national legislation. In other words, directives tend to feature obligation modals more than regulations because “[d]irectives, which are addressed to the Member States, indirectly define rights and obligations via the intermediation of the Member States [...]. They impose an obligation on the Member States to ensure that an objective will be attained. Directives are worded in a more general way to leave discretionary power to the Member States in transposition” (Biel 2014: 15). Similarly, permission modals (less frequent than obligation modals in the overall corpus) are 1.5 times more frequent in directives than regulations, because the former must allow for different circumstances in the Member States. In directives the most frequent permission modal was, of course, *may*, followed by *can*, *could*, *might* and *need not*, but it must be highlighted that *may* on its own accounted for over 78% of all permission modals identified in the corpus and had a frequency of 3,245 per million words.

Anselmi and Seracini (2015) compared EU directives with UK national transposition measures, like the present contribution. They collected a corpus of consumer law, ENG CoL, which includes 16 EU directives on consumer law issued between 1984 and 2014 (about 133,000 words) and the 16 UK regulations that transposed them into UK law (about 165,000 words). A 100,000-word reference corpus of UK regulations on related topics (health and safety, fair trading, student loans, and so on) adopted over the same period of time was also collected, “[...] to establish whether the differences that emerged from the analysis were the result of a tendency of the British drafters to adapt to the national language conventions or whether they were ascribable to other constraints linked to the implementation process” (*ibid.*: 43-44). In relation to the use of modals, they found that *shall* was the most common obligation modal in all the corpora. However, while in the older texts (1984-2000) there was a similar frequency of use across the three subcorpora, over the 2000-2014 period (which overlaps with the present study) there was a marked drop in frequency in the NTMs and in the reference corpus of UK

⁶ The JRC Acquis is a corpus of the “*Acquis Communautaire*” legislative texts in 22 languages (i.e. the total body of EU law applicable in the EU Member States), compiled and made available to the research community by the Joint Research Centre of the EU. See Biel (2017) and <https://ec.europa.eu/jrc/en/language-technologies/jrc-acquis> for more information.

regulations. More specifically, the frequency of *shall* was 12,190 per million words in directives, 2,840 in the transposition measures and 4,420 in the reference corpus.

Finally, a recent study compared the English and the Italian versions of the same EU directives, regulations and decisions on consumer law with a focus on translation-al patterns for modality; as well as the ENGLEX and ITALEX components, the EURO-CoL corpus also includes two monolingual reference corpora, UK-LAW and LEGITALIA. Seracini's (2020) study found that *shall* was the most frequent modal in all types of legislation, but especially in decisions and directives, and that it "[...] occurs almost exclusively in the enacting terms" (*ibid.*: 78). Moreover, the frequency of *shall* in the monolingual reference corpus of UK legislation was far lower than in the EURO-CoL corpus. At the same time, *must* and the semi modal *is/are to* have a much lower frequency than *shall* in EU law, while they are very frequent in the UK-LAW corpus. As regards *may*, its frequency was slightly higher in directives than in decisions and regulations and, just like *shall*, it was much more commonly found in the enacting terms than in other sections of each text; *can*, on the other hand, is mostly used in the preambles, generally with a non-prescriptive meaning.

This brief literature review has shown that in recent years there has been a significant change in modal usage in the legislation of most English-speaking countries; there is a widespread tendency to avoid *shall* and to replace it with other modals or prescriptive verbal constructions (in the UK this trend became especially marked after 2005). By contrast, in the EU *shall* is still frequently used, especially in directives (more than in other types of EU legislation). The chapter on English Eurolect in the Eurolect Observatory book (Sandrelli 2018) compared the frequency of obligation, prohibition and permission modals in directives and NTMs and found some interesting differences too; therefore, the present paper expands that study by adding data from the C subcorpus of domestic legislation that is unrelated to the EU setting, in order to compare and contrast frequency and usage of such forms across the three subcorpora, as is illustrated in Section 3.

3. Methodology and data

The English section of the *Eurolect Observatory Multilingual Corpus* (EOMC) includes an A subcorpus of EU Directives (1999-2008), a B subcorpus of UK national transposition measures (Acts of Parliament and Statutory Instruments; 2000-2014), and a C corpus of UK domestic legislation (UK Public Acts and Statutory Instruments; 1999-2013)⁷. The corpus of directives is just over 4 million words, while the two corpora of English law run to just over 8 million tokens each.

After collecting the texts and converting them into a machine-readable format, annotation and mark-up were added to each document to record useful information, such as the subject (as it is indicated in the EUR-Lex database), and to mark the different sections in each document (title/ preamble/ disposition/ annex). This makes it possible to select samples of texts on a specific topic (e.g. all the directives on energy), to select

⁷ The time span of the NTMs is slightly later than the directives because the transposition process takes some time (from a few months to two years, in some cases).

a specific section of each document (e.g. only the preambles) or both (e.g. only the preambles of energy directives).

As was mentioned in the Introduction (see Section 1), in the first phase of the project the A and B corpora were investigated. More specifically, the analysis was conducted on the enacting terms of both corpora, as they contain the actual provisions of each law and are therefore potentially more linguistically interesting at all levels (lexical, morpho-syntactic and textual). Indeed, the analysis identified many differences between the language of the directives and the language of the transposition measures, thus proving the existence of an English Eurolect as a legislative variety in its own right (Sandrelli 2018). This is hardly surprising, for a number of reasons. Firstly, in the European Union English is used to express legal concepts which tend to be closer to the civil law tradition than to common law (Pozzo 2012). Secondly, some differences between the language of directives and the language of UK domestic legislation are to be expected on account of the transposition process. Although the recommended method for transposition is copy-out (i.e. using the same wording as the directive), elaboration is actually much more common (Steunenbergh and Voermans 2006), because EU legislation tends to be written in fairly vague and general terms (Robertson 2010), while “common law drafting has traditionally tended towards exhaustiveness by covering every imaginable situation” (Williams 2008: 3). Thus, as Anselmi and Seracini (2015: 41) remark:

The transposition of an EU directive into the British legal system, which may be regarded as intralingual translation according to Jakobson’s model (1959), involves the rewording of a text written in English using the same language in order to adapt it to a new audience. This process, [...] involves mediation and adaptation to a new context.

Thirdly, and perhaps most importantly, since the 2004 round of enlargement, English has become the main drafting language for practical reasons (Pozzo 2012); more specifically, it has been estimated that 80% of European Commission documents are routinely drafted in English (Robinson 2008). However, although 95% of Commission drafters write in English, only 13% have English as their native language (Wagner 2010). It is unclear whether English will maintain its role of institutional *lingua franca* in the post-Brexit EU, but the issue has no bearing on the present study, as the directives in the A corpus were adopted between 1999 and 2008; however, it is bound to have an impact on the evolution of the English Eurolect, which will need to be investigated.

One aspect in which the analysis conducted on the enacting terms of the A and B corpora found interesting differences is usage patterns in deontic modality (Sandrelli 2018). Therefore, in the present study it was decided to expand the analysis to the C corpus, to try and obtain a more sophisticated description of Eurolect-specific usage of modals and semi-modals in directives versus domestic legislation.

Given the importance of the above-mentioned verb forms in prescriptive legislation, there are some guidelines available to EU drafters, lawyer-linguists and translators. The most important reference can be found in the *Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation*, which states:

2.3.1. The choice of verb and tense varies between different types of act and the different languages, and also between the recitals and the enacting terms (see Guidelines 10 and 12).

2.3.2. In the enacting terms of binding acts, other languages, such as French, use the present tense, whilst English generally uses the auxiliary ‘shall’. In both languages, the use of the future tense should be avoided wherever possible. (JPG 2015: 12)

More detailed instructions can be found in the *English Style Guide* of the Directorate-General for Translation (DGT) of the European Commission, aimed at translators rather than drafters. This is what the Guide says (ESG 2020: 54) on the use of modals in the main clauses of the enacting terms of EU legislation:

The enacting terms of EU legislation (articles) can be divided broadly into three categories: imperative, permissive, and declarative. Imperative and permissive provisions can be positive or negative. They require or oblige (imperative) or allow (permissive) someone to do or not to do something. Declarative provisions are those that are implemented directly by virtue of being declared, for example definitions or amendments.”

Table 1 summarizes the recommendations of the Guide on the use of modals depending on the type of provisions in which they appear:

	positive	negative
imperative (conveying command)	<i>shall</i> (obligation or requirement)	<i>shall not</i> (prohibition)
permissive (conveying permission)	<i>may</i> (permission to do)	<i>need not</i> (permission not to do)
declarative	present tense + optional <i>hereby</i> (direct implementation) <i>shall</i> (future implementation)	

Table 1. Guidelines for the use of modals in enacting terms

Interestingly, as regards positive imperative usage, the guide also specifies that “[h]ere, *shall* means the same as *must*. In contrast with EU usage, most English-speaking countries now generally use *must* instead of *shall*. So you may do the same when translating non-EU legislation as long as you do so consistently” (ESG 2020: 54). In other words, DGT is aware of the recent trends in legislative language use in English-speaking countries (see Section 2). Moreover, on modal verbs used to convey permission, the Guide clarifies that, although *may* is the preferred option for positive forms, negative permission must not be expressed via *may not*, “[...] since it could be interpreted as expressing possibility” (ESG 2020: 55). As legislation drafted in English is translated into all the other languages, it is essential to avoid potential ambiguity.

Neither the *English Style Guide* nor the *Joint Practical Guide* are binding; therefore, it will be interesting to see whether the above principles can actually be seen at work in our A corpus. The first step in this study was the extraction of the enacting terms for the three subcorpora; Table 2 presents the data used in the analysis. As can

be seen in the table, the A subcorpus is about one third smaller than B and C, which are of very similar sizes. Interestingly, however, the number of laws in subcorpus C is more than double the number in corpus B; this means that, on average, the texts in the C subcorpus are shorter than those in B.

	Corpus A directives	Corpus B NTMs	Corpus C UK law
documents	660	674	1,429
tokens	1,475,255	4,622,434	4,645,381

Table 2. Corpora used in this study

The analysis focuses on the occurrences of the obligation, permission and prohibition modals and semi-modals, all retrieved via *Wordsmith Tools 6.0*⁸ (Scott 2013). Since the three corpora differ in size, frequencies were normalized to 1 million words, to make results comparable. Section 4 presents our results, with a few selected examples.

4. Analysis and results

Starting with obligation modals and semi-modals (see Table 3), the first aspect that is worthy of note is that their overall frequency is much higher in the enacting terms of directives than in the other two corpora, with 19,076 occurrences per million words vs. almost 9,000 in the NTMs and almost 8,000 in the UK laws with no relation to the EU setting. This is attributable to the nature of directives, whose purpose is to impose obligations on Member States (see Section 2).

	Corpus A (per 1 m)	Corpus B (per 1 m)	Corpus C (per 1 m)
<i>shall</i>	17,687	6,174	4,503
<i>must</i>	753	1,686	2,041
<i>is to/ are to</i>	407	817	1,257
<i>should</i>	177	189	165
<i>has/have to</i>	52	23	20
TOTAL	19,076	8,889	7,986

Table 3. Obligation modals and semi-modals

Shall was by far the most frequent obligation modal in all three subcorpora, but its frequency in the directives was 2.8 times higher than in the NTMs and almost 4

⁸ In the near future it will be possible to investigate other prescriptive verbal structures as well, such as the imperative and present simple forms, as the part-of-speech (POS) tagging of our corpora has recently been completed. POS-tagging is a process whereby all the words in a corpus are assigned a grammatical tag; this makes it possible to retrieve all the occurrences of a given linguistic structure automatically via tools such as *Sketch Engine* (see <https://www.sketchengine.eu/>).

times higher than in the UK laws (C corpus). Indeed, in the corpus of directives *shall* accounts for about 90% of all obligation modals, while in the NTMs the proportion is about 70% and in UK laws it is about 56%. In other words, obligation is almost always expressed via *shall* in directives, while there is more variation in corpus B and corpus C, in which the other alternatives are not uncommon. Indeed, both *must* and *is/are to* are over twice as frequent in NTMs as in corpus A, and almost three times as frequent in corpus C⁹. The fact that *shall* is more frequent in NTMs than in domestic laws could be attributed to the transposition process, especially when the copy-out mechanism is involved. In the C corpus (the legislation drafted in the UK with no relation to the EU) there is a more even distribution of obligation modals across the three most frequent forms (*shall*, *must*, *is/are to*); these results are in line with the recent trends in UK legislation described in Section 2, namely the gradual replacement of *shall* with other obligation modals and semi-modals.

In all three corpora *shall* is very often used in intra-textual references, i.e. sentences in which reference is made to a section of the text itself (article, paragraph, section, schedule, and so on) or to the whole text (this Directive; this Act), and in extra-textual references (another directive or law), in order to affirm the legal validity of the instrument or to introduce amendments to the text. Example 1 features a sentence that establishes the applicability of several articles and of another directive to a specific case (A corpus); a sentence that sets the date for the coming into force of a regulation (B corpus); and a sentence that amends the text of a paragraph in a Schedule (C corpus). In all three corpora there are plenty of examples of such uses of *shall*.

A	B	C
Articles 3(1) and (2), 4(4), 6(1), 12, 17(1), 19, 20, 23, 24, 25, 40 to 52, 70 to 85, 101 to 108, 111(1) and (3), 112, 116 to 118, 122, 123, 125, 126, second subparagraph, and 127 of this Directive as well as Commission Directive 91/356/EEC(6) shall apply , by analogy, to traditional-use registration granted under this chapter.	This regulation, regulations 2, 7, 9 and 10 and Part 1 of Schedule 2 shall come into force on 30th May 2006.	Where the notice is required to be given under section 86B(2)(b) of the 1974 Act it shall also include the information set out in Part 3 of Schedule 3 and the statement in paragraph 4(1) of that Schedule shall be amended as specified in paragraph 13 of that Schedule.

Example 1. Shall used in textual references

However, taking textual amendments as an example, a quick concordancing query showed that while in the corpus of directives the passive “*shall be amended*” is the most common option, in the C corpus the most frequent options are the present simple forms of the verb, in the active (*amends*) and passive voice (*is amended/ are amended*). Once again, the B corpus is between the two poles, as the frequency of *shall be amended* is lower than in directives but higher than in the C corpus, while the opposite is true of the

⁹ The other options considered here (*should*, *has/have to*) have almost negligible frequencies in all three corpora.

frequency of the above-mentioned present simple forms¹⁰. However, it must be noted that the latest edition of the *English Style Guide* (ESG 2020) prescribes the use of the present tense in such cases rather than *shall* (e.g. article 1 *is replaced* by... rather than *shall be replaced*); it would be interesting to collect a corpus of recent amending acts to see whether there is a significant difference in this respect in comparison with our A corpus.

Example 2 shows a sentence from the corpus of directives in which *shall* is used with an inanimate subject in the phrase *shall be subject to*. A quick concordancing query showed that the frequency of *shall be subject to* in the B corpus is about half the frequency of the corpus of directives, and it is even lower in the C corpus (150 occurrences per million words in A, 85 in B and 66 in C). In the B and C corpora the preferred option is the present simple (*is subject to*).

A	B	C
These objectives shall be subject to approval by the Member States or competent authorities and their implementation shall be monitored by them.	(1) This section applies where: (a) by virtue of section 91 or 92, a planning permission is subject to a condition that the development to which the permission relates must be begun before the expiration of a particular period, that development [...]	(3) A statutory instrument containing an order under section 88 which neither amends nor repeals any provision of primary legislation is subject to annulment in pursuance of a resolution of either House of Parliament. (4) If a draft of an instrument containing an order under section 88 [...]

Example 2. *Shall be subject to /is subject to*

Moving on to prohibition modals, the top frequency ranking follows a similar pattern in all three corpora, with the most frequent one being *shall not*, followed by *may not* (see Table 4). However, in the corpus of directives the third most frequent alternative is *cannot*, which is much less frequent in the other two corpora of UK legislation. By contrast, in the B corpus and the C corpus the third option is *must not*, much less frequently used in directives (about half as frequent as in the NTMs).

The fact that *shall not* is the most frequent prohibition modal in directives is in line with the recommendations in the *English Style Guide* (see Table 1); indeed, it accounts for almost 68% of all such forms in directives. What is interesting, however, is that its frequency is almost the same in the national transposition measures (B corpus), where it accounts for about 68% of all prohibition modals, but it is considerably lower in the C corpus, where it covers about 48% of the total. Once again, therefore, it would seem that the corpus of legislation originated in the UK (corpus C) has a more even distribution of the available options. Moreover, it is interesting to note that the semi-modal *is not to/ are not to* is hardly ever used in the enacting terms of directives, while it is not so uncommon in the other two corpora. With the exception of *shall not* (whose frequency

¹⁰ In the near future, it will be possible to carry out a more sophisticated query in the POS-tagged versions of the three corpora, to obtain more accurate frequency information and automatically retrieve other textual alternatives as well, such as the use of the imperative (*amend*).

	Corpus A (per 1 m)	Corpus B (per 1 m)	Corpus C (per 1 m)
<i>shall not</i>	1,115	1,018	583
<i>may not</i>	308	200	332
<i>must not</i>	57	108	130
<i>is not to/ are not to</i>	7	68	93
<i>should not</i>	13	26	17
<i>cannot</i>	136	80	76
<i>could not</i>	13	27	17
TOTAL	1,649	1,527	1,248

Table 4. Prohibition modals and semi-modals

in directives and NTMs is very similar), the frequency of prohibition modals in the B corpus is in many cases an intermediate value between the two poles of A and C.

Finally, let us now move on to permission modals and semi-modals. Our analysis of such forms in A and B revealed that the most frequent permission modal was *may* in both the directives and the NTMs, with all the other forms in Table 5 being much less common. The same trend can be seen in subcorpus C.

	Corpus A (per 1 m)	Corpus B (per 1 m)	Corpus C (per 1 m)
<i>may</i>	3,942	5,854	5,646
<i>can</i>	174	114	26
<i>could</i>	80	91	111
<i>need not</i>	63	39	38
<i>might</i>	66	76	65
<i>is/are not required to</i>	9	15	36
TOTAL	4,334	6,189	5,922

Table 5. Permission modals and semi-modals

May accounts for 91% of all permission modals in the A corpus A, 94.5% in the B corpus and 95.3% in the C corpus; in other words, it is by far the preferred option to convey permission. However, in terms of actual occurrences *may* is much more frequent in subcorpus B and subcorpus C than in A (almost 6,000 occurrences per million words vs. almost 4,000). This may be related to the nature of national laws, which need to specify what the addressees of the law (citizens, businesses, and so on) are allowed or not allowed to do; by contrast, directives only lay down provisions in very general terms, leaving it to the national authorities of EU countries to flesh out the details.

Another interesting result is that (although their frequencies are much lower than that of *may*) both *can* and *need not* seem to be over-represented in subcorpus A in comparison with B and C. While *need not* is the recommended modal to convey negative

permission (i.e. permission not to do something) in the *English Style Guide* (Table 1), the higher frequency of *can* is harder to explain with any certainty. It may be related to the multilingual language contact in which EU directives are drafted (and therefore the influence of other languages in drafting, such as French) or to the high proportion of European Commission drafters who are not native speakers of English, or to other factors. What is certain is that, although the frequency of *can* in directives is not very high in absolute terms, it is almost seven times higher than in the corpus of UK laws.

5. Conclusions

The present study aimed at verifying whether obligation, prohibition and permission modals and semi-modals are used differently in the enacting terms of directives (A corpus), NTMs (B corpus) and UK domestic laws (C corpus). Our results have provided further confirmation of the high frequency of modals in the enacting terms of directives, which can be explained by the fact that they are binding legal instruments but require transposition into national legislation before their provisions become applicable; therefore, their purpose is to illustrate the obligations imposed on Member States and explain what must be done and what is and is not permissible (see Biel 2017 in Section 2). This is clearly shown by the high frequency of the obligation modal *shall*, the prohibition modal *shall not* and the permission modal *need not*, whose frequencies are much higher than in the other two corpora, and by the under-representation of the alternative modal and semi-modal forms identified in the B and C corpora.

More specifically, in our corpus of directives the frequency of *shall* is even higher than in Biel's (2014, 2017) work and in Anselmi and Seracini's (2015) work; while the analyses conducted by these scholars were carried out on the whole text of directives, the present contribution was focused only on the enacting terms, where a higher frequency was indeed to be expected (Foley 2001). *Shall* seems to be preferred in the EU setting for the sake of clarity, since "[...] *shall* has a value in a multilingual context as its function is well understood as normative and it is a clear and unambiguous sign [...]" (Robertson 2010: 156). As many drafters are not native speakers and they are aware that the English text of each directive will be translated into all the other EU languages, they seem to resort to a well-known marker of prescriptive texts (to facilitate translation), unlike their UK counterparts. The studies carried out by Williams (2005, 2007, 2013) and Garzone (2013a, 2013b) demonstrated a gradual decrease in the use of *shall* and a parallel increase in the use of other obligation modals in UK legislation. Such patterns were confirmed by our data; indeed, the patterns that had already been identified in the A-B comparison (Sandrelli 2018) have emerged even more clearly after adding the C corpus to the picture. In both corpora of UK legislation (B and C) there is a wider variety of obligation, prohibition and permission modals than in the corpus of directives. The increase in the use of other obligation modals and semi-modals is especially marked in the C corpus, made up of laws drafted in Westminster with no direct link with EU laws; by contrast, the B corpus has emerged as a 'hybrid' between the two poles of English Eurolect and national legislative English in relation to the use of modals. In other words, modality usage patterns in Eurolect seem to differ considerably from those found in the domestic variety of legislative English in the UK.

The present study has a number of inherent limitations. Firstly, it has a limited scope, as it is focused exclusively on directives and does not take into account other legal instruments such as regulations and decisions; other scholars (Biel 2014, Seracini 2020) have found some differences in modality usage patterns between directives, regulations and decisions, which probably need to be investigated further. Secondly, there is a geographical limitation as well, since for the time being the only variety of English being considered in the Eurolect Observatory project is UK English; an analysis of Irish legislation may add interesting insights to the study of modality in legislative texts.

Thirdly, the study is based on corpora collected over a specific time-span and results cannot be generalized; it would be useful to investigate modality usage patterns in more recent legislation, to check how stable these features are in the English Eurolect and in the UK domestic variety of legislative language. Indeed, the key questions for the future of the English Eurolect are strictly connected with political developments, namely the impact of Brexit on language policies and (above all) language practice in Brussels (see Williams 2017). While it seems likely that for practical reasons English will still be used as the EU's institutional lingua franca, the UK's exit will also produce a decrease in the number of English native speakers in the EU institutions, which is bound to have an impact on Eurolect.

References

- Anselmi S. and F. Seracini 2015. The transposition of EU directives into British legislation as intralingual translation: a corpus-based analysis of the rewriting process. *Textus* 28/2: 39-62.
- Biel Ł. 2014. The textual fit of translated EU law: a corpus-based study of deontic modality. *The Translator* 20/3: 332-355.
- Biel Ł. 2017. *Lost in the Eurofog. The Textual Fit of Translated Law*. Frankfurt: Peter Lang.
- ESG 2020. *The English Style Guide: A Handbook for Authors and Translators Working in English for the European Commission*. European Commission Directorate-General for Translation. European Union. At https://ec.europa.eu/info/sites/info/files/styleguide_english_dgt_en.pdf.
- Foley R. 2001. Going out in style? *Shall* in EU legal English. *UCREL Technical Paper* 13. Special Issue. Proceedings of the Corpus Linguistics 2001 conference, edited by P. Rayson, A. Wilson, T. McEnery, A. Hardie and S. Khoja: 185-195. At <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/foley.pdf>.
- Garzone G. 2013a. Variation in the use of modality in legislative texts: focus on *shall*. *Journal of Pragmatics* 57: 68-81.
- Garzone G. 2013b. Modality and performativity in legislative texts across jurisdictions: the case of *shall*. In F. Poppi and W. Cheng (eds), *The Three Waves of Globalization. Winds of Change in Professional, Institutional and Academic Genres*. Newcastle: Cambridge Scholars Publishing: 94-122.
- JPG 2015. *Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union Legislation*. European Union. At <https://op.europa.eu/slouUS>.
- Mori L. (ed.) 2018a. *Observing Eurolects*. Amsterdam/Philadelphia, John Benjamins.
- Mori L. 2018b. Introduction. The *Eurolect Observatory Project*. In L. Mori (ed.), *Observing Eurolects*. Amsterdam/Philadelphia: John Benjamins: 1-26.

- Pozzo B. 2012. English as a legal lingua franca in the EU multilingual context. In C.J.W. Baaji (ed.), *The Role of Legal Translation in Legal Harmonization*. Alphen aan den Rijn (the Netherlands): Kluwer Law International BV: 183-202.
- Robertson C. 2010. Legislative drafting in English for non-native speakers: some do's and don'ts (with reference to EU legislation). *ESP Across Cultures* 7: 147-163.
- Robinson W. 2008. Drafting of EU legislation: a view from the European Commission. *Federalismi.it. Rivista di diritto pubblico italiano, comunitario e comparato* 1. At <http://www.federalismi.it>.
- Sandrelli A. 2018. Observing Eurolects: the case of English. In L. Mori (ed.), *Observing Eurolects*. Amsterdam/Philadelphia: John Benjamins: 63-92.
- Scott M. 2013. *WordSmith Tools* 6.0. Lexical Analysis Software Limited.
- Seracini F. 2020. *The Translation of European Union Legislation. A Corpus-based Study of Norms and Modality*. Milano: LED.
- Steenenbergh B. and W. Voermans 2006. *The Transposition of EC Directives: A Comparative Study of Instruments, Techniques and Processes in Six Member States*. Leiden: Leiden University and WODC/Ministry of Justice.
- Tomatis M. 2018. The *Eurolect Observatory Multilingual Corpus*. Construction and query tools. In L. Mori (ed.), *Observing Eurolects*. Amsterdam/Philadelphia: John Benjamins: 27-45.
- Wagner E. 2010. Why does the Commission need a clear writing campaign? *Clear writing* [Languages and Translation]. European Commission Directorate General for Translation, September 2010: 4-5.
- Wagner E., S. Bech and J.M. Martínez 2002. *Translating for the European Union Institutions*. Manchester: St Jerome.
- Williams C. 2005. *Tradition and Change in Legal English: Verbal Constructions in Prescriptive Texts*. Bern: Peter Lang.
- Williams C. 2007. Crossovers in legal cultures in Westminster and Edinburgh: some recent changes in the language of the law. *ESP Across Cultures* 4: 101-118.
- Williams C. 2008. Legal English or legal Englishes? Differences in drafting techniques in the English-speaking world. *Rivista di diritto pubblico italiano, comunitario e comparato* 1. At <http://www.federalismi.it>.
- Williams C. 2011. Legal English and plain language: an update. *ESP Across Cultures* 8: 139-151.
- Williams C. 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*. Cambridge: Cambridge University Press: 353-371.
- Williams C. 2017. What future for the English language in a post-Brexit European Union? In E. Triggiani, F. Cherubini, I. Ingravallo, E. Nalin and R. Virzo (eds), *Dialoghi con Ugo Villani*. Bari: Cacucci: 755-759.