



English Language and European Union: A Corpus-Based Study of EU Secondary Legislation

Ornella Guarino

1. Introduction

The European Union is a unique economic and political union shared by 27 different countries; hence its environment is inherently multilingual and multicultural. Indeed, European legislation must be drafted and made available in all EU official languages, although English represents “the most frequently used language both in source texts and target texts [in EU institutions]” (Seracini 2020: 25). English is not only the source language of European legislation, but also the most studied language in Europe. In 2017, 91% of students in primary and secondary schools were studying English (Kuzelewska 2020: 1425). English has therefore long been considered the *lingua franca* of the European Union.

However, it is true that after the withdrawal of the United Kingdom from the European Union, 60 million native English speakers have left the EU (Kuzelewska 2020: 1422). Although it has already been confirmed that English will remain an official EU language, this situation prompts the question as to what the future of English will be in the post-Brexit landscape, especially because of the decrease in the number of native English speakers within the EU who contribute to the drafting of EU legislation.

Many scholars have investigated whether Euro-English¹ can be identified as a variety in its own terms. Garzone already considered in 2000 that “a special language that has been described as Euro-legalese is being generated” (Garzone 2000: 7). Other scholars like Seidlhofer state that in order to consider Euro-English as a variety in its own right, European people need to consider themselves as “a relatively stable community of speakers who themselves acknowledge that they speak a variety of their own” (Garzone 2013: 83). Still other scholars like Balič state that “Euro-English must be regarded as EU jargon due to its technical, administrative or legal nature and not as a separate non-standard form of English for EU institutional settings” (Balič 2016: 131).

Based on the above considerations, it is possible to state that there is a difference between the English legal language used in EU legislation and the British legal language. It is therefore interesting to wonder whether the former will take on increasingly different characteristics from the British legal English. This represents a question which seems to be very topical nowadays.

Ornella Guarino, Università di Palermo, ornella.guarino01@unipa.it

¹ Euro-English is widely considered as an emerging variety of English used within European institutions as a means of communication between speakers of other European languages (Jenkins, Modiano, Seidlhofer 2001: 14).

This study aims therefore at conducting a corpus-based analysis of the English language that is used within the European institutions – the so-called Euro-English – in order to explore specific linguistic features that characterize it and that help to distinguish it from British legal English (cf. Tiersma 1999; Garzone 2000; Biel, Jenkins, Modiano, Seidlhofer 2006; Jacometti, Pozzo 2008; Bhatia 1993, 2008; Mattila 2016; Bolton, Davis 2017; Biernacka, Jopek 2018; Mac Giolla Chríost, Bonotti 2018). First, a brief introduction of the complex linguistic landscape of the European Union is provided, by assessing the historical process of multilingualism and the EU egalitarian language policy. Subsequently, the article goes on to analyze a corpus of EU secondary legislation (directives, regulations, decisions, recommendations, and opinions) drawn up in English – namely the *de facto lingua franca* of the EU – and elaborated within the EU institutions in the last decades (1992-2022) to explore specific linguistic features of Euro-English. The selected corpus of EU legislation is compared to a corpus of UK legislation containing all UK legislation – excluding the legislation originating from the EU – drafted during the same time frame. Such a corpus-based analysis is carried out at the lexical, morphosyntactic and textual levels and the data is produced using the WordSmith Tools 8.0 software².

2. Multilingualism and the EU Egalitarian Language Policy

In order to describe the principle of multilingualism and the EU egalitarian language policy in Europe, it is first necessary to refer to a distinction made by Van Els between ‘institutional’ and ‘non-institutional’ language policy in the EU. In particular, the ‘institutional’ one refers to the language policy determining the use of languages in and between the EU institutions, as well as the use of languages outside the EU and the languages used between the EU and the Member States. On the contrary, the ‘non-institutional’ one refers to the languages used by citizens in individual Member States (Van Els 2006: 205-216; Kuzelewska 2014: 152). Specifically, this article takes into consideration the ‘institutional’ language policy, as the analysis has been carried out on a corpus of EU secondary legislation.

The European Union has currently 24 official languages and, despite Brexit, to this date English still represents the *lingua franca* of the European institutions. The EU has adopted 24 official languages based on the principle of multilingualism³ and on Regulation 1/1958⁴, the latter determining the languages to be used by the European Economic Community and establishing that each Member State has the right to have its official language recognized as an official language of the EU⁵.

Before 1958, the importance of all EU national languages had not been a priority. The Treaty of Paris of 1951 signed by Belgium, France, Italy, Luxembourg, the Nether-

² WordSmith Tools is an integrated suite of programs for looking at how words behave in texts. (Scott 2020).

³ According to the Charter of Fundamental Rights of the European Union, the EU is obliged to respect linguistic diversity, to prohibit discrimination on account of language and to provide for the citizen's right to communicate with the institutions in one of the languages of the Treaties.

⁴ Regulation 1/1958 has been subsequently amended by Regulation 517/2013.

⁵ Regulation 1/1958 determining the languages to be used by the European Economic Community.

lands, and West Germany and founding the European Coal and Steel Community (ECSC) came into force in 1952, had given implicit importance to French by stating that,

*Le présent Traité, rédigé en un seul exemplaire, sera déposé dans les archives du Gouvernement de la République Française, qui en remettra une copie certifiée conforme à chacun des gouvernements des autres Etats signataires*⁶.

[The present Treaty, drawn up in a single copy, will be deposited in the archives of the Government of the French Republic, which will deliver a certified copy to each of the governments of the other signatory States.] [my translation]

Nevertheless, over the next few years the importance of national languages increased and was reinforced by the principle of equal recognition of all four languages (Dutch, French, German and Italian) included both in the *Protocole sur le régime linguistique de la Communauté européenne du charbon et de l'acier*⁷ and in Article 1 of Regulation No. 1/1958. Such regulation further established in Article 2 that,

Documents which a Member State or a person subject to the jurisdiction of a Member State sends to institutions of the Community may be drafted in any one of the official languages selected by the sender. The reply shall be drafted in the same language (EEC Council 1958).

Furthermore, Regulation No. 1/1958 established in Article 4 that, "Regulations and other documents of general application shall be drafted in the four official languages" (EEC Council 1958)⁸.

Thanks to the subsequent enlargements⁹ that allowed other countries to join the EU, there was an evolution of the official language regime of the European Economic Community (EEC). Theoretically, nowadays all current 24 official languages have an equal status (Kuzelewska 2021: 1418). Also, based on the above the European Union's approach has increasingly evolved over time towards an acceptance and promotion of linguistic diversity. However, in practice only three of the official languages are more frequently used within the European institutions. Such languages are known as 'the big three', namely French, German and English – the latter being the most used of all. For instance, while in the European Parliament all 24 official languages are working languages, in the European Commission only English, French, and German are working languages. Also, in the case of the European Central Bank, English is the only working language.

The fact that some languages are more used for internal communication than others seems to be contrasting the principle of equality among languages. Before the accession of the UK to the EU, French had the role of dominant language within the European institutions. However, after 1995 English became the mainly used lan-

⁶ Treaty of Paris (1951). Article 100.

⁷ The *Protocole sur le régime linguistique de la Communauté européenne du charbon et de l'acier* was published in 1952 by determining the language regime of the Community and by concluding that the four languages of the signatory states (French, German, Italian, Dutch) are granted equal official recognition within the ECSC (Mac Giolla Chríost, Bonotti 2018).

⁸ Ibid.

⁹ Particularly, the UK joined the EEC during the 1973 enlargement together with Ireland and Denmark.

guage in the primary texts of the European Commission, and a significant body of EU documents is only available in English (Kuźelewska 2021: 1419; Biel, Biernacka, Jopek-Bosiacka 2018: 251). For this reason, Pym believes that “one of the solutions may be a restructuring of the effective hierarchy of languages, abandoning the illusions of equality” (Pym 2000: 8). Having a single procedural language might be convenient in terms of budgetary constraints and practical considerations, although it would be inconsistent with the EU’s multilingualism policy (Biernacka, Jopek-Bosiacka 2018: 250). On the other hand, Kuźelewska gives credit to the EU egalitarian language policy by stating that it “is a unique example of fully democratic language policy implemented in international organization” (Kuźelewska 2021: 1420). Additionally, it should be noted that the EU promotes and encourages linguistic diversity and plurilingual education. European citizens speak more than 60 national, regional and minority languages (Kuźelewska : 1420), thus constituting an important linguistic and cultural heritage.

Over the past years, the question whether English should be recognized as the only *lingua franca* of the European Union has generated a very delicate debate. What is certain is that English remains an official language of the EU despite Brexit and, most importantly, it remains its *de facto lingua franca* despite the EU’s attempts to preserve linguistic equality and diversity. English is therefore the mainly used language within EU institutions, and as stated in the previous section, it is acknowledged by many scholars that the legal English that is used within the EU institutions differs from British legal English. In the next section, specific linguistic features of the so-called Euro-English are therefore taken into analysis.

3. Corpus-Based Analysis: Comparison of Corpora

The present section aims at exploring specific linguistic features that help to distinguish Euro-English from the British legal English (cf. Garzone 2000; Tiersma 1999; Mattila 2016; Biel, Biernacka, Jopek-Bosiacka 2018; Jenkins, Modiano, Seidlhofer 2006; Seracini 2020). The analysis is carried out with a corpus-based approach (Biber, Conrad, Reppen 1998) and, specifically, on two purposely-built comparable corpora: The EU Corpus contains EU secondary legislation for a total number of 661,512 tokens, while the UK Corpus contains all UK legislation – excluding the legislation originating from the EU – for a total number of 615,678 tokens. The legislation included in both corpora was drawn up over the time frame 1992-2022. In order to build both corpora, four successive periods in the 30-year span have been identified and an equal number of legislative acts has been included in each period of each corpus to ensure balance between the two corpora.

Specifically, the comparison has focused on linguistic features characterizing legal language at the lexical (terminology, collocations, loanwords, archaisms), morphosyntactic (modal verb morphology, impersonal structures) and textual levels (anaphoric references, formulaic expressions). Regarding the specific terminology that is analyzed in Section 3.1.2, a comparison is made between the use of specific words contained

in the EU Corpus and the definitions of such terms provided by the Cambridge Dictionary with the purpose of comparing the meaning of specific terms used in the EU Corpus and the meaning of the same words according to Standard English (SE). Data were produced by using the WordSmith Tools 8.0 software and, although they are limited, they already provide interesting results in terms of linguistic features that help to identify Euro-English. Also, the findings of this study can be of support for further future research on the evolution of Euro English as a variety in its own right.

3.1 Lexical analysis

The most salient results of this study were found at the lexical level. According to Mattila, “Euro-English contains a number of terms that do not exist in common law English and many terms that exist in common-law English but that are used with a more or less distinct continental meaning” (Mattila 2016: 349). Already existing words are therefore sometimes used with new or different meanings. This is mainly because European legislation is drafted in such a way that it can be translated and therefore applied in all Member States (Robinson 2008: 1). Relevant EU collocations are analyzed in Section 3.1.1; subsequently, specific EU terminology is analyzed in Section 3.1.2; Section 3.1.3 analyzes loanwords by considering borrowings both from Latin and French; finally, Section 3.1.4 analyzes another typical feature of legal language, namely archaisms.

3.1.1 EU Collocations

The most relevant collocations were found in the EU Corpus by examining the word list obtained with the WordSmith Tools 8.0 software. Both the EU Corpus and the UK Corpus are compared and the most relevant results in terms of frequency are shown in Table 1. Specifically, in order to carry out the analysis, the Concord tool was used to verify the collocations of specific words. Table 1 shows the collocations that have been detected. As it can be noticed, they have specific meanings related to European concepts that have been developed in the last decades within the European institutions.

EU collocations	EU Corpus 661,512 tokens	UK Corpus 615,678 tokens
Member State(s)	4242 0.64%	17 <0.01%
European Union	2851 0.43%	25 <0.01%
Official Journal (of the European Communities)	1455 0.22%	0
Third country(ies)	737 0.11%	0
European community(ies)	628 0.10%	19 <0.01%
Single/internal market	143 0.02%	18 <0.01%
Free movement	64 <0.01%	0
Candidate state/country	14 <0.01%	0
Acceding state/country	10 <0.01%	0

Table 1, EU collocations

The collocations shown in Table 1 are more frequent in the EU corpus than in the UK corpus. The reason is that such collocations refer to specific European institutions or definitions – as in the case of ‘Member State(s); ‘European Community; ‘European Union or Official Journal (of the European Communities)’ – and others that are not included in Table 1, such as ‘European Parliament’ or ‘European Commission’. Other collocations have other meanings, such as ‘Third country(ies); which is used in the EU Corpus to refer to those countries that do not belong to the European Union. Also, it can sometimes be found together with ‘nationals’, as to include those subjects who are not citizens of the EU. The collocation ‘single/internal market’ refers to the EU Single Market including all 27 Member States of the EU. Another relevant collocation is ‘free movement’, which in the EU Corpus refers to the right of the EU citizens “to move and reside freely within the territory of the EU Member States”¹⁰. Finally, with regard to the last two collocations ‘candidate state/country’ and ‘accessing state/country’, they are less frequent in the EU Corpus than the other collocations, but they are still relevant at the lexical level as they refer to specific European concepts. Indeed, ‘candidate state/country’ refers to a “country still negotiating to join the EU” (European Commission, Directorate-General for Translation 2011: 12), whereas ‘accessing state/country’ refers to a “country about to join the EU”

3.1.2 EU terminology

At the lexical level, interesting results were also found with regard to some terms that are used with a meaning that differs from that of British legal English. This occurs as some specific terms that are used in EU legislation stem from the various national legal systems of the EU Member States, and they are therefore deeply culture-bound. Consequently, because such terms are used in EU legislation with different meanings and often refer to specific Europe-bound concepts, they can generate serious misinterpretations with regard to their correct meaning (Jacometti, Pozzo 2006: 18).

The WordList tool of the Word-Smith Tools 8.0 was used to analyze

Analyzed terms	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
Case	515 0.08%	739 0.10%
Do/done	309 0.04%	309 0.04%
Harmonisation/ harmonization/harmonised	169 <0.03%	1 <0.01%
Important	116 0.02%	1 <0.01%
Proportionality	74 0.01%	0
Coherent/coherence	47 <0.01%	1 <0.01%
Eventual/eventually	40 <0.01%	0
Foresee	26 <0.01%	2 <0.01%
Normally	18 <0.01%	1 <0.01%
Mainstreaming	12 <0.01%	0
Legislator	7 <0.01%	0

Table 2, Specific EU terminology

¹⁰ European Commission. Migration and Home Affairs. Retrieved July 19, 2022, from https://ec.europa.eu/home-affairs/pages/glossary/right-free-movement_en

specific EU terms and to compare the frequency and, therefore, the meanings of such words in the two corpora under analysis. The terms that have been analyzed are shown in Table 2.

Each of the terms taken into analysis deserves an in-depth explanation:

- The term 'case' is often used in both corpora, but in the EU Corpus it is often used together with the preposition 'in' as a synonym with 'if', as in the following sentence example taken from the EU Corpus: "In case the SMP operator provides equivalent access at the MDF, the NRA may decide to set a shorter period".
- The verb 'do' is often used in the EU Corpus as a synonym with 'make', 'elaborate' or 'perform'. For instance, the verb 'do' is often found at the bottom of the EU legislation in sentences such as: "Done at Brussels" or "Done at Strasbourg" - in which the verb refers to the legislative act in question - as in the following example taken from the Council Directive (EU) 2020/2020 of 7 December 2020 amending Directive 2006/112/EC as regards temporary measures in relation to value added tax applicable to COVID-19 vaccines and in vitro diagnostic medical devices in response to the COVID-19 pandemic: "Done at Brussels, 7 December 2020." As it can be noticed, 'done' is used instead of 'made'. On the contrary, in the UK Corpus the verb 'make' is used to refer to the drafting of legislative acts, as in the following example taken from The Coronavirus Act 2020 (Residential Tenancies: Protection from Eviction) (Amendment) (England) (No. 2) Regulations 2021: "This is the original version (as it was originally made)".
- The general meaning of 'harmonisation' is, according to the Cambridge Dictionary, "the act of making different people, plans, situations, etc. suitable for each other"¹¹. However, within EU institutions, 'harmonisation' is used to refer to the process of making identical rules and standards within the internal market, as in the following example sentence taken from the EU Corpus: "Directive 2014/31/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of non-automatic weighing instruments".
- The adjective 'important' means "of great value, meaning, or effect"¹². However, in the EU Corpus it can also be used as a synonym with 'big' or 'large', as in the following sentence example taken from the EU Corpus: "The replacement of copper by fibre up to an intermediary distribution point represents an important investment entailing some risk".
- With regard to 'proportionality', the latter has acquired a precise meaning within the European context as it refers to the principle of proportionality laid down in Article 5 of the Treaty on European Union which states that, "Under the principle of proportionality, the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties"¹³.

¹¹ Cambridge Dictionary (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/harmonization?q=harmonisation>

¹² Cambridge Dictionary (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/important>

¹³ Consolidated version of the Treaty on European Union - Title I: Common Provisions - Article 5 (ex-Article 5 TEC).

- According to the Cambridge Dictionary, 'coherent' means "logical and well organized; easy to understand and clear"¹⁴. In the EU Corpus, however, 'coherent' and 'coherence' are used as synonyms with 'in accordance with' or 'in compliance with', as in the following sentence example taken from the EU Corpus: "The Commission shall assess the implementation and effectiveness of external-cost charging for CO2 emissions, as well as its coherence with Directive 2003/87/EC of the European Parliament and of the Council and Council Directive 2003/96/EC."
- Although 'eventual' and 'eventually' are not often used in both corpora – in the UK Corpus they are not even present – it is possible to notice that both terms have acquired a particular meaning within the EU Corpus. According to the Cambridge Dictionary, 'eventual' means "happening at a later time or as a result at the end"¹⁵, and 'eventually' means "in the end, especially after a long time or a lot of effort, problems, etc."¹⁶. Nevertheless, both terms are used in the EU Corpus as a synonym with 'possibly' or 'in case', as in the following example sentence present in the EU Corpus: "Any recommendation or agreement with professional operators which is aimed [...] at preventing, limiting or imposing specific conditions on the placing on the market or the eventual use of food or feed on account of a serious risk to human health requiring rapid action."
- 'Foresee' means "to know about something before it happens"¹⁷. However, in the EU Corpus 'foresee' has acquired a different meaning - probably stemming from other European languages – which is 'to set out', 'to provide' or 'to include', as in the example sentence present in the EU Corpus: "Therefore, requirements foreseen in this Directive are not relevant for Member States which do not have any ports at which ships falling under the scope of this Directive normally can call."
- The adverb 'normally' means "usually or regularly"¹⁸. Nevertheless, in the EU Corpus it is often used to refer to something that should happen or that is expected to, as in the following example sentence present in the EU Corpus: "The deployment of FTTH will normally entail considerable risks, given its high deployment costs per household [...]."
- The general meaning of 'mainstreaming' is "the process of making something start to be considered normal"¹⁹. However, in the EU Corpus it is used to refer to the act of "taking into account all EU policies" (European Commission, Directorate-General for Translation 2011: 12). Also, sometimes 'mainstreaming' is used to-

¹⁴ Oxford Dictionary. Available at: <https://www.oxfordlearnersdictionaries.com/definition/english/coherent?q=coherent>

¹⁵ Cambridge Dictionary. (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/eventual>

¹⁶ Cambridge Dictionary. (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/eventually>

¹⁷ Cambridge Dictionary. (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/foresee>

¹⁸ Cambridge Dictionary. (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/normally>

¹⁹ Cambridge Dictionary. (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/dictionary/english/mainstreaming>

gether with the collocate 'gender' in position L1, as in the following example taken from the EU Corpus: "In accordance with the principle of gender mainstreaming, this report shall, inter alia, provide an assessment of the impact of the measures taken on women and men", by referring indeed to the principle of gender mainstreaming, which is the policy-making approach that takes into consideration both women's and men's interests and concerns, and aims at designing better policies.

- Finally, according to the Cambridge Dictionary, the term 'legislator' refers to "a member of a group of people who together have the power to make laws"²⁰. However, in the EU Corpus 'legislator' is used to refer to a body or more than one body, such as the European Council or the European Parliament. An example taken from the EU Corpus: "The importance placed by the Union legislator on the protection provided for consumers in Council Directive 93/13/EEC (18) and in Directives 2008/48/EC and 2014/17/EU, means that [...]."

As it can be noticed, most of the words taken into analysis are used with a meaning that differs from that of British legal English. As demonstrated, their meanings differ from those explained in the Cambridge Dictionary. Also, another observation can be made: Some of the words suggest that in the cases of 'important', 'eventually' or 'coherent/coherence', there is strong interference from other official languages. The interference is probably due to the contact from other languages – such as French, Italian, and Spanish – that are spoken by the lawyers and experts involved in the EU legislative drafting as many of them are not native English speakers (Seracini 2020: 37).

3.1.3 Loanwords

Additional interesting results of the lexical analysis concern the borrowings from Latin and French, namely all the loanwords and the Latinisms – which are typical of legal language (Tiersma 1999: 5; Mattila 2016: 246) – present in the two corpora taken into analysis. The most interesting results concerning Latinisms²¹ were found using the WordList tool of the Wordsmith Tools 8.0 and are shown in Table 3.

Latinisms	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
Annex(es)	937 0.14%	43 <0.01%
Minimum	186 0.03%	21 <0.01%
e.g. (exempli gratia)	69 0.01%	0
Medium	66 <0.01%	1 <0.01%
Inter alia	64 <0.01%	1 <0.01%
Via	49 <0.01%	0
i.e. (id est)	34 <0.01%	0
Interim	24 <0.01%	55 0.01%
Mutatis mutandis	23 <0.01%	0
Ad hoc	17 <0.01%	0
Consensus	14 <0.01%	0

Table 3, Latinisms

²⁰ Cambridge Dictionary. (n.d.). Retrieved July 19, 2022, from <https://dictionary.cambridge.org/it/dizionario/inglese/legislator>

²¹ In Table 3, only the most recurrent Latinisms are shown.

Loanwords	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
Acquis	64 <0.01%	0
Surveillance	32 <0.01%	7 <0.01%
Cabotage	21 <0.01%	0
Dossier(s)	42 <0.01%	0
Force majeure	5 <0.01%	0

Table 4, Loanwords from French

frequent in the UK Corpus, whereas they can be considered as a common feature of Euro-English.

With regard to the borrowings from French, the results are shown in Table 4. The terms included in the table are the most recurrent ones.

As it can be noticed, only *surveillance* is present in the UK Corpus and, on the whole, loanwords from French are almost completely unused in the UK Corpus. Instead, they are used in the EU Corpus. Specifically, *acquis* is an interesting loanword which not only comes from French, but it has also acquired a totally unique meaning in the European context, namely that of indicating the body of EU law (European Commission, Directorate-General for Translation 2011: 12).

Archaisms	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
Pursuant to	400 0.06%	81 0.01%
Thereof	289 0.04%	37 <0.01%
Thereby	45 <0.01%	2 <0.01%
Hereby	36 <0.01%	21 <0.01%
Therein	33 <0.01%	16 <0.01%
Thereafter	31 <0.01%	7 <0.01%
Thereto	27 <0.01%	11 <0.01%
Hereinafter	25 <0.01%	5 <0.01%
Forthwith	22 <0.01%	15 <0.01%
Aforesaid	5 <0.01%	12 <0.01%
Aforementioned	5 <0.01%	0

Table 5, Archaisms

According to the data collected, Latinisms are more frequent in the EU Corpus, despite the recommendation of the 'English Style Guide' stating that "Latin should be used sparingly as even common phrases are often misused or misunderstood" (European Commission 2016: 51). The only exception is represented by *interim*, which is more frequent in the UK Corpus. On the whole, however, it is possible to state that Latinisms are not very

3.1.4 Archaisms

Another noteworthy linguistic feature is the presence of archaisms in the EU Corpus. According to Hu and Lu, "Archaism nearly occurs in the everyday use of modern English, [...]. It is a language that is not current or that is used only within a few specific forms" (Hu, Lu 2017: 798). Also, archaism "refers to English from AD 450 to 1150. Archaism and modern English are very different in terms of pronunciation, spelling, vocabulary and grammar" (Hu, Lu 2017: 798). The results concerning archaisms in the EU Corpus and in the UK Corpus are obtained using the WordList tool of the WordSmith Tools 8.0 and are shown in Table 5.

Crystal and Davy state that archaisms come from old English, and that

most of them “take the form of adverbials to which some prepositions are affixed” (1969: 207). Indeed, as it can be noticed, most of the archaisms taken into consideration are there-prefixed archaisms, e.g., ‘thereof’ or ‘thereby’, and here-prefixed archaisms, e.g., ‘hereby’ or ‘hereinafter’. Other archaisms include adjectives, e.g., ‘aforesaid’ and ‘aforementioned’, or the adverb ‘pursuant to’. On the whole, archaisms are more used in the EU Corpus, thus constituting an additional characteristic of the English language used within the EU Corpus and, also, making the legislation less ‘plain’. Indeed, according to Tiersma, archaisms can make text less clear and cause problems of understanding as they are not used in modern English usage, which is the reason why it is usually recommended to replace them with modern words complying with plain English in legal writing (1999: 114).

3.2 Morpho-syntactic analysis

With regard to the morpho-syntactic analysis, two main features are taken into analysis in this study, namely the modal verb morphology (Section 3.2.1) and the impersonal structures (Section 3.2.2), as they have proven to be relevant morpho-syntactic features that help to distinguish Euro-English from British legal English.

3.2.1 Modal verb morphology

Modals are so called as they express modal meanings such as possibility, volition, and obligation and are classified according to the type of meaning they express (Palmer 2014: 2). Particularly, there are two categories that have a place in ordinary language, which are the ‘epistemic’ and the ‘deontic’ modalities. Epistemic modals are “essentially making a judgment about the truth of the proposition” (Palmer 2014: 6), whereas deontic modals are “concerned with influencing actions, states or events” and are “used to express what is obligatory, permitted, or forbidden” (Palmer 2014: 6).

Modals are particularly frequent in legal discourse, as the latter deals with obligations and permissions (Seracini 2020: 11). They are also characterized by polysemy, ambiguity and vagueness, so they shall always be used with caution in European legislation (Garzone 2013: 79). Indeed, as Šarčević states, serious misinterpretations can occur while translating modals in other European languages, thus threatening legal harmonization within the European context (2007: 35). Also, the Directorate-General for Translation at the European Union highlights how modals represent a challenge for translators, as stated in the booklet ‘Lawmaking in the EU Multilingual Environment’,

The excessive use of terms ‘shall’ and ‘will’ caused difficulties in a number of languages where they were translated using future tense although the languages concerned should and could have used the present tense in a prescribing sense. ‘Should’ also caused translation difficulties for some languages, especially when it is used in preambles where target languages would rather use the subjunctive mood, because of the normative aspects of provisions in preambles (European Commission, Directorate-General for Translation 2010: 94).

‘Shall’ represents indeed the most important modal in legal discourse, as historically it has been the most used in legislative texts. In the last decades, the use of ‘shall’ in legal discourse has also attracted the attention of institutional bodies, linguists, and philosophers of law who have given way to important debates and reflections on its use (Garzone 2013: 69, 79). As a matter of fact, in the 1970s the Plain English Movement started to promote clarity and simplicity in legal language (Mattila 2016: 245). Such a movement arose in the United States and strongly criticized the use of ‘shall’ because of its ambiguous nature. Some linguists have even suggested giving up on shall completely and to use ‘must’ instead (Kimble 1992: 69). Particularly, Asprey suggests to abandon ‘shall’ and to use the followings instead (Asprey 1992: 79):

- Must’ for the imperative shall, as in the case of obligation or duties to impose, or directions to make;
- ‘Will’ for the simple future;
- The present tense for everything else, as in the case of statements of facts, legal results or agreement.

Modals and semi-modals	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
Shall	5792 0.88%	898 0.14%
Should	3049 0.46%	73 0.01%
May	1579 0.24%	2484 0.32%
Can	432 0.07%	67 0.01%
Will	375 0.06%	100 0.02%
Would	294 0.04%	316 0.04%
Must	236 0.04%	923 0.14%
Could	181 0.03%	43 <0.01%
Ought to	0	11 <0.01%

Table 6, Modals and semi-modals

As Garzone demonstrated, the use of ‘shall’ has indeed decreased in the UK legislative writing (Garzone 2013: 71). However, within the EU institutions ‘shall’ is still frequently used, although drafters are recommended to be careful when using it as “[t]he use of verbs in legislation, contracts and the like often gives rise to problems, especially when it comes to the verb ‘shall’, since it is not used in this way in everyday speech” (European Commission 2016: 58). The data collected using the WordSmith Tools 8.0, and particularly by using the WordList tool, have demonstrated indeed the highly frequent use of ‘shall’ in the EU Corpus. The most relevant results in terms of frequency are shown in Table 6.

As it can be noticed, in the EU Corpus ‘shall’ maintains the role of “most important word in the world of legal drafting” (Kimble 1992: 61) and remains the most used modal, whereas ‘must’ is used much less frequently. On the contrary, in the UK Corpus the use of ‘shall’ decreases by giving way to ‘may’, whereas ‘must’ is much more frequently used. According to the data collected, it is not possible to determine when ‘shall’ began to be less used in the UK legislation; however, it is true that in EU legislation such a modal continues to be widely used. Indeed, according to the word list of EU legislation elaborated in 2022 – which has been obtained by using the WordList tool of the WordSmith Tools 8.0 – ‘shall’ is the 19th most used word, as shown in Table 7.

It is therefore possible to assume that the EU is still reluctant to change its drafting style regarding the use of 'shall', contrary to the established trend of the UK and other English-speaking countries. As the English Style Guide clearly acknowledges that the general trend of the English-speaking countries is to use 'must' instead of 'shall' (European Commission 2016: 58), the persistent use of 'shall' seems therefore to represent a linguistic feature of Euro-English.

3.2.2 Impersonal structures

At the morpho-syntactic level, another linguistic feature is represented by the impersonal structures as they are important structures in legal language expressing what is obligatory, permitted, or forbidden, just like modals do. The impersonal style represents indeed one of the main characteristics of legal discourse (Mattila 2016: 73). In Table 8 the 'it is + adjective' structure is taken into analysis as it appears to be an impersonal structure that is particularly present in the EU Corpus. The results obtained by using the Concord tool of the WordSmith Tools 8.0 are shown in Table 8.

The booklet 'How to Write Clearly' recommends to try to name the agent²² as much as possible. Despite such recommendation, impersonal structures are more frequent in the EU Corpus than in the UK Corpus – with the only exception of 'it is reasonable' – thus preventing the agents from being identified.

Impersonal structures it is + adjective	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
It is necessary	89 0.01%	14 <0.01%
It is appropriate	78 0.01%	21 <0.01%
It is important	29 <0.01%	0
It is essential	13 <0.01%	0
It is possible	12 <0.01%	0
It is reasonable	9 <0.01%	12 <0.01%

Table 8, Impersonal structures

N	Word	Frequency	%
1	#	15449	7.08%
2	THE	15289	7.01%
3	OF	10858	4.98%
4	AND	6727	3.08%
5	TO	5204	2.39%
6	IN	4883	2.24%
7	OR	2827	1.30%
8	A	2681	1.23%
9	FOR	2455	1.13%
10	DATA	1576	0.72%
11	ON	1556	0.71%
12	IS	1532	0.70%
13	WITH	1515	0.69%
14	THAT	1414	0.65%
15	BY	1353	0.62%
16	BE	1351	0.62%
17	UNION	1300	0.60%
18	AS	1294	0.59%
19	SHALL	1290	0.59%

Table 7, First 19 words of the Word List
EU Corpus 2022

²² According to the booklet *How to Write Clearly*, the agent is the person, organisation or thing that is carrying out the action (European Commission, Directorate-General for Translation 2011: 9).

3.3 Textual analysis

At the textual level, there are specific features characterizing legal discourse such as anaphoric references (Section 3.3.1) and formulaic expressions (Section 3.3.2), that are taken into analysis in this section as their use in the EU Corpus has proven to be particularly relevant for the purposes of this study.

3.3.1 Anaphoric references

Anaphoric references refer to linguistic forms through which *“il parlante fa riferimento ad un referente al quale egli, nel suo discorso, ha già fatto riferimento con un’espressione antecedente”* (Conte 1999: 19) [the speaker refers to a referent to whom he has already referred with an earlier expression in his own speech] [my translation]. In order to analyze the presence of anaphoric references in the EU Corpus and the UK Corpus, specific words and expressions functioning as anaphoric expressions have been analyzed with the WordSmith Tools 8.0 software by using the WordList tool. The results are shown below in Table 9. In particular, the anaphoric references shown in Table 9 are the most recurrent ones in the EU Corpus.

Anaphoric reference	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
Whereas	433 0.07%	2 <0.01%
Having regard to	371 0.06%	20 <0.01%
According to the/ according to article	114 0.02%	9 <0.01%

Table 9, Anaphoric references

‘Whereas’, ‘having regard to’ and ‘according to the/according to article’ are all used in recitals of the legislative texts contained in the EU Corpus. For instance, in Directive 97/9/EC of the European Parliament and of the Council of 3 March 1997, each section in the preamble begins with the word ‘whereas’, as in the following examples: “(1) Whereas on 10 May 1993 the Council adopted [...]”; “(2) Whereas Directive 93/22/EEC lays down prudential rules [...]”; “(3) Whereas, however, no system of supervision can provide complete protection [...]”; “(4) Whereas the protection of investors and the maintenance of confidence in the financial system [...]”; “(25) Whereas, in conclusion, a minimum degree of harmonization of investor-compensation arrangements is necessary [...]”

In the UK Corpus, the anaphoric references under analysis are not widely used. In the case of ‘whereas’, only two occurrences have been found. Also, in both cases ‘whereas’ is not used as an anaphoric expression, as in the following example taken from The Companies Act 1985 (Electronic Communications) Order 2000, which states as follows:

Whereas the Secretary of State considers that the authorisation of the use of electronic communications by this Order for any purpose is such that the extent (if any) to which records of things done for that purpose will be available will be no less satisfactory in cases where use is made of electronic communications than in other cases (The Companies Act 1985 (Electronic Communications) Order 2000).

In the following statements of the Act under consideration, ‘whereas’ is not further used. The use of specific anaphoric expressions therefore represents a linguistic feature of Euro-English.

3.3.2 Formulaic expressions

Formulaic expressions represent another feature of legal discourse which deserves to be taken into consideration for the purposes of this analysis. Formulaic expressions are formulas, phrases or expressions that are repeated continuously in the text. Specific formulaic expressions that are frequently used in legal discourse have indeed been taken into analysis by using the ConcordTool of the WordSmith Tools 8.0 software. The results obtained are shown in Table 10.

What emerges from the table above is that all of the formulaic expressions taken into account occur more often in the EU Corpus than in the UK Corpus, with the only exception of ‘for the purposes (of) (paragraph)’. Moreover, ‘in line with’ never occurs in the UK Corpus, which suggests that the use of such formula within the EU corpus is due to interference from the languages of non-native English speakers. With the exception of ‘for the purposes (of) (paragraph)’, it can be stated that all the formulaic expressions shown in Table 10 are typical features of Euro-English.

Formulaic expressions	EU Corpus 661,512 tokens	UK Corpus 658,678 tokens
In accordance with	1694 0.26%	423 0.06%
For the purposes (of) (paragraph)	632 0.10%	1419 0.22%
Without prejudice to	298 0.05%	38 <0.01%
In compliance with	121 0.02%	4 <0.01%
In/within the framework of	112 0.02%	6 <0.01%
In line with	74 0.01%	0
In conformity with [FR/ES]	40 <0.01%	4 <0.01%
Due to the	25 <0.01%	5 <0.01%

Table 10, Formulaic expressions

4. Concluding remarks

As Tiersma states, “legal languages are inevitably products of the history of the nation or state in which they are used, as well as peculiar developments of the legal system in question” (Tiersma 2008: 7). The case of Euro-English is therefore a very peculiar one as it represents a complex and unique situation in which it is demonstrated that the cultures of all Member States are mixed with each other giving rise to a specific type of language that is influenced by all European cultures and languages. Indeed, Euro-English does not stem from the single culture of a single country, but rather from the cultures of all the Member States (Jacometti, Pozzo 2006: 18-24). The result is that the legislation produced by the European institutions is drafted in a singular way with the use of specific terms and expressions. This has been proved by the exploratory corpus-based analysis carried out in this article.

In particular, the analysis has shown that Euro-English has specific linguistic features at different levels. The most salient results have been proved to exist at the lexical level, in which collocations referring to specific European institutions or concepts refer to the European context. At the lexical level, data have also proved that there are specific terms that are used with a meaning that differs from that of British legal English. Also, some of the words taken into analysis are used with the meaning of similar words of other official European languages. This suggests that there is interference coming from other official European languages as most of the drafters of the European legislation are not native English speakers. Other important results have been found at the morphosyntactic level, in which the use of certain modals, particularly the verb 'shall', has proved to be typical of Euro-English. Also, the use of the impersonal structures 'it is + adjective' appears to be particularly present in the EU Corpus, thus implying another linguistic feature of Euro-English. Finally, at the textual level, specific anaphoric references and formulaic expressions have also proven to be particularly present in the EU Corpus.

On the whole, since Euro-English already has its own linguistic features – most importantly at the lexical level – the results of this study suggest that further linguistic features of Euro-English are likely to develop in the future, especially after Brexit. It is however difficult nowadays to predict the effects of Brexit on Euro-English. The latter may indeed develop even further by taking on increasingly distinctive characteristics. However, what is certain to this day is that Euro-English already possesses specific vocabulary of its own and specific linguistic features at the different levels taken into analysis in this study. With time, it will be possible to observe how Euro-English will evolve, whether it will continue to deviate from British legal English or not, and, most importantly, to what extent.

References

- Asprey Michele M. (1992). "Shall Must Go". *Scribes Journal of Legal Writing*, 3, 79-84.
- Balič Tina (2016). "Attitudes towards Euro-English in a European Union Institution". *ELOPE: English Language Overseas Perspectives and Enquiries*, 13(2), 131-152.
- Bhatia Vijay K. (1993). *Analysing Genre: Language Use in Professional Settings*. London: Routledge.
- Bhatia Vijay K. (2008). *Language, Culture and the Law: The Formulation of Legal Concepts Across Systems and Cultures*. Bern: Peter Lang.
- Biber Douglas, Conrad Susan, Reppen Randi (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biel Łucja, Biernacka Agnieszka, Jopek-Bosiacka Anna (2018). "Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations". In: Silvia Marino, et al. *Language and Law*. Switzerland: Springer, 249-274.
- Bolton Kingsley, Davis Daniel R. (2017). "Brexit and the Future of English in Europe". *World Englishes*, 36(3), 302-312.
- Charter of Fundamental Rights of the European Union (2012/C 326/02). Bruxelles: Official Journal of the European Union. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT>.
- Mac Giolla Chríost Diarmait, Bonotti Matteo (2018). *Brexit, Language Policy and Linguistic Diversity*. Switzerland: Springer.
- Consolidated version of the Treaty on European. Bruxelles: Official Journal of the European Union. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12016M/TXT>.
- Conte Maria-Elisabeth (1999). "Deissi Testuale e Anafora." In: Ugo Cardinale. *Condizioni di Coerenza. Ricerche di Linguistica Testuale. Atti del seminario*, Alessandria: Edizioni dell'Orso, 11-28.
- Crystal David (2017). "The Future of New Euro-Englishes". *World Englishes*. 36(3), 330-335.
- Crystal David, Davy Derek (1969). *Investigating English Style*. United States: Indiana University Press.
- EEC Council (1958.) Regulation No 1 determining the languages to be used by the European Economic Community. Bruxelles: Official Journal of the European Communities. Available at: [https://eur-lex.europa.eu/eli/reg/1958/1\(1\)/oj](https://eur-lex.europa.eu/eli/reg/1958/1(1)/oj). Accessed 21 Feb. 2023.
- Carter Ronald (2011). *English Grammar Today: An A-Z of Spoken and Written Grammar*. Cambridge: Cambridge University Press.
- European Commission (2016). *English Style Guide. A Handbook for Authors and Translators in the European Commission*. Bruxelles: Publications Office. Available at: https://commission.europa.eu/system/files/2023-01/styleguide_english_dgt_en.pdf. Accessed 21 Feb. 2023.
- European Commission, Directorate-General for Translation (2011). *How to Write Clearly*. Bruxelles: Publications Office. Available at: <https://data.europa.eu/doi/10.2782/54318>. Accessed 21 Feb. 2023.
- European Commission. Directorate-General for Translation (2010). *Lawmaking in the EU multilingual environment*. Bruxelles: Publications Office. Available at: https://termcoord.eu/wp-content/uploads/2013/08/Study_on_lawmaking_in_the_EU_multilingual_environment.pdf. Accessed 21 Feb. 2023.

- Garzone Giuliana (2000). "Legal Translation and Functional Approaches: A Contradiction in Terms?" In: *Actes du Colloque International "La traduction juridique. Histoire, Théorie(s) et Pratique"*, École de Traduction et d'Interprétation - Université de Genève, 1-14. Available at: <http://www.tradulex.com/en/translators/Legal-Genève2000>. Accessed 21 Feb. 2023.
- Garzone Giuliana (2013). "Variation in the Use of Modality in Legislative Texts: Focus on Shall". *Journal of Pragmatics*, 57, 68-81.
- Hu Jialei, Lu Huixia (2017). "The Application and E-C Translation Methods of Common Archaisms in Business Contract". *Theory and Practice in Language Studies*, 7(9), 798-803.
- Jacometti Valentina, Pozzo Barbara (Eds.) (2006). *Le politiche linguistiche delle istituzioni comunitarie dopo l'allargamento: redazione, traduzione e interpretazione degli atti giuridici comunitari e il loro impatto sull'armonizzazione del diritto europeo*. Milano: Giuffrè Editore.
- Jenkins Jennifer, Modiano Marko, Seidlhofer Barbara (2001). "Euro English". *English Today*, 17(4), 13-19.
- Jenkins Jennifer, Modiano Marko, Seidlhofer Barbara (2006). "Euro-English". *Perspectives on an Emerging Variety on the Mainland of Europe, from Commentators in Sweden, Austria and England*, 3, 45-54.
- Kimble Joseph (1992). "The Many Misuses of Shall". *Scribes Journal of Legal Writing*, 3, 61-78.
- Kuźelewska Elżbieta (2014). "Unity in Diversity. The Language Policy of the European Union". *Studies in Logic, Grammar and Rhetoric*, 38(51), 151-165.
- Kuźelewska Elżbieta (2021). "Quo Vadis English? The Post-Brexit Position of English as a Working Language of the EU". *International Journal for the Semiotics of Law*, 34(5), 1417-1432.
- Mattila Heikki ES (2016). *Comparative Legal Linguistics: Language of Law, Latin and Modern Lingua Francas*. Farnham: Ashgate.
- Modiano Marko (2006). "Euro-Englishes". In: Braji Kachru, Yamuna Kachru and Cecil L. Nelson (eds.) *The Handbook of World Englishes*. United States: Blackwell Publishing, 223-239.
- Modiano Marko (2017). "English in a Post-Brexit European Union". *World Englishes*, 36(3), 313-327.
- Mollin Sandra (2006). *Euro-English: Assessing Variety Status*. Tübingen: Gunter Narr Verlag.
- Palmer Frank Robert (2014). *Modality and the English Modals*. New York and London: Routledge.
- Pozzo Barbara (2008). *Europa e Linguaggi Giuridici*. Milano: Giuffrè Editore.
- Pym Anthony (2000). "The European Union and its future languages: Questions for language policies and translation theories". *Across languages and cultures*, 1(1), 1-17.
- Robinson William (2008). "Drafting of EU Legislation: A View from the European Commission". *Rivista di Diritto Pubblico Italiano, Comunitario e Comparato*, 1-15. Available at: <https://www.federalismi.it/AppOpenFilePDF.cfm?artid=9008&dpath=document&dfile=08012008032419.pdf&content=Drafting%2Bof%2BEU%2Blegislation%3A%2Ba%2Bview%2Bfrom%2Bthe%2BEuropean%2BCommission%2B%2D%2Bdottrina%2B%2D%2Bdottrin-a%2B%2D%2B>. Accessed 21 Feb. 2023.
- Šarčević Susan (2007). "Making Multilingualism Work in the Enlarged European Union." In: Krzysztof Kredens, Stanisław Goźdz-Roszkowski. *Language and the Law: International Outlooks*. Frankfurt am Main: Peter Lang, 35-56.

- Šarčević Susan (2015). *Language and Culture in EU Law. Multidisciplinary Perspectives*. Farnham: Ashgate.
- Scott Mike (2020). WordSmith Tools version 8, Lexical Analysis Software.
- Seracini Francesca L. (2020). *The Translation of the European Union Legislation: A Corpus-based Study of Norms and Modality*. Milano: LED Edizioni.
- Tiersma Peter M. (1999). *Legal Language*. Chicago and London: University of Chicago Press.
- Tiersma Peter M. (2008) "The Nature of Legal Language." In: Gibbons John, Turel M. Teresa (eds). *Dimensions of Forensic Linguistics*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 7-25.
- Treaty of Paris (1951). Available at: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:11951K:EN:PDF>. Accessed 21 Feb. 2023.
- Van Els Theo (2006). "The European Union, its Institutions and its Languages: Some Language Political Observations". In: Richard Baldauf, Robert Kaplan (eds). *Language Planning and Policy in Europe. The Czech Republic, The European Union and Northern Ireland*. United Kingdom: Multilingual Matters, 202-251.