

Alma Mater Studiorum Università di Bologna

DIPARTIMENTO DI INTERPRETAZIONE E TRADUZIONE

Corso di Laurea magistrale Specialized Translation (classe LM - 94)

TESI DI LAUREA

in Terminology and Information Mining

TRACES OF EUROLECT IN SWISS LEGAL ITALIAN? A CORPUS ANALYSIS ON
THE CHEU-LEX CORPUS

CANDIDATO:

Silvia Polito

RELATORE:

Adriano Ferraresi

CORRELATORI:

Silvia Bernardini

Annarita Felici

Anno Accademico 2020/2021

Terzo Appello

Summary

ABSTRACT	3
INTRODUCTION	5
CHAPTER 1	8
1.1. Introduction	8
1.2. Drafting and translational process at the European Union	8
1.2.1. The EU and its multilingual environment.....	8
1.2.2. Drafting process.....	9
1.2.3. Drafting and translational issues	12
1.2.4. Europeanisation process	14
1.3. The languages of the EU: the Eurolects	16
1.3.1. The birth of the term “Eurolect”	16
1.3.2. The Italian Eurolect	17
1.4. Legal Italian	18
1.4.1. Historical background	18
1.4.2. Defining legal Italian.....	18
1.5. Comparative Jurilinguistics	20
1.5.1. Historical background	20
1.5.2. Combining legal and linguistic studies: the birth of Jurilinguistics	21
1.6. Background: The Eurolect Observatory Project	22
1.6.1. Research project.....	23
1.6.2. The creation of the <i>Eurolect Observatory Multilingual Corpus</i> and the research tools	24
1.6.3. The research and the results.....	24
CHAPTER 2	27
2.1. Introduction	27
2.2. Research question	27
2.3. Two different implementation procedures: Italy and Switzerland	28
2.4. The CHEU-Lex corpus	31
2.4.1. Corpus structure.....	31
2.4.2. The creation of the corpus	32
2.4.2.1. Texts selection and downloading.....	33
2.4.2.2. Text cleaning	36
2.4.2.3. Segmentation	40
2.4.2.4. Alignment	45
2.4.2.5. POS tagging and dependency parsing	48
2.4. Methodology	54
2.4.1. The Research Template	61
2.4.2. Lexical analysis	62
2.4.2.1. EU-rooted lexical features	64
2.4.2.2. Intra-linguistic lexical variability	66
2.4.2.3. Latinisms	68
2.4.3. Lexical morphology	69
2.4.4. Verb morphology.....	70
2.4.5. Morphosyntax.....	73
2.4.6.1. EU-rooted syntactical phenomena	77
2.4.6.2. Intra-linguistic syntactic variability	78
2.4.7. Textual level	79
2.4.7.1. EU-rooted textual phenomena.....	79

2.5. General remarks, limitations to this research and future developments.....	80
CHAPTER 3.....	82
3.1. Introduction.....	82
3.2. Lexical analysis.....	82
3.2.1. EU-rooted lexical features	85
3.2.2. Intra-linguistic lexical variability.....	92
3.2.3. Latinisms and Eurolect renderings.....	97
3.3. Lexical morphology.....	100
3.4. Verb morphology	103
3.5. Morphosyntax.....	110
3.6. Syntax.....	115
3.6.1. EU-rooted syntactical phenomena	116
3.6.2. Intra-linguistic syntactic variability	118
3.7. Textual level.....	122
3.7.1. EU-rooted textual phenomena	122
3.8. General remarks, limitations of this study and future developments	124
CONCLUSIONS.....	126
AKNOWLEDGEMENTS	129
BIBLIOGRAPHY	130
SITOGRAPHY.....	133

ABSTRACT

In the multilingual environment of the European Union, laws are drafted in a *lingua franca*, usually English or French, and then translated into the other EU official languages. In this regard, the *Eurolect Observatory Project* discussed the possible existence of legal linguistic varieties called *Eurolects*, i.e. legal varieties born within the European framework and characterized by a set of linguistic features.

Although the Swiss Confederation is not part of the European Union, it is located in the middle of Europe. This led to the creation of the *bilateral agreements* between the Swiss Confederation and the European Union. For this reason, the Swiss legal Italian might present some European linguistic features. Therefore, this thesis discusses whether any relevant influence of the European legal drafting process can be observed in the Italian of Swiss laws of implementation.

To test this hypothesis, data were extracted from the Italian CHEU-Lex sub-corpus, i.e. the Italian corpus of Swiss laws of implementation. The corpus was built with the aim of providing a multilingual resource to investigate the influence of EU drafting and translational process on Swiss legislation. It was built in a joint project led by Professor Annarita Felici (University of Genève) between the University of Bologna and the University of Genève.

Following the same analysis carried out during the *Eurolect Observatory Project*, results obtained from the CHEU-Lex Corpus were compared to those obtained from the two Italian corpora previously analyzed by Mori (2018c): Corpus A, i.e. the corpus of Italian EU directives, and Corpus B, i.e. the corpus of Italian laws of implementation. Results underlined the presence of few elements which could be related to the influence of the European legal drafting process on the Italian of Swiss laws of implementation. However, these results did not provide strong and convincing evidence confirming the presence of traces of Eurolect in Swiss legal Italian. Indeed, some of these features might derive from the content of the laws analyzed and/or from the context (Swiss Confederation) in which they are written.

To better understand if these traits are the result of the influence of the European legal drafting process or if, instead, they are observed in the CHEU-Lex Corpus for reasons other

than European contact, it would be necessary to carry out a further analysis, comparing the CHEU-Lex Corpus with a corpus of Italian Swiss laws unrelated to the EU context. In this sense, if the same features are observed in the corpus of Italian Swiss laws, then it might indicate that these features are typical of the Swiss legal Italian. If, instead, these elements are not observed in the corpus of Italian Swiss laws, or they are observed to a lesser extent, then this might be a further clue confirming the possible influence of the of EU drafting and translational process on Swiss legislation.

INTRODUCTION

The multilingual framework in which European laws are drafted is extremely peculiar. Indeed, when drafting laws at the EU level, these must be as understandable and accessible as possible to every European citizen, regardless of their native language. To achieve this aim, texts are drafted in a chosen *lingua franca* (which usually happens to be English) and then translated into all the 24 official languages of the European Union. Also, each law must have the same legal force of its translated versions, which means that, in theory, there should not be an “original version” and its “translated version”, as this would somehow subordinate the translated versions to the original one. However, this aim is extremely difficult, if not utopic, to achieve, as it would imply the simultaneous drafting of the legislative texts in 24 different languages. Therefore, the procedure mentioned above was adopted.

It is in this peculiar scenario that the *Eurolect Observatory Project*¹ was born. It observed and discussed the existence of legal linguistic varieties called Eurolects for some of the official languages of the European Union. These linguistic varieties are legal varieties born within the European framework and are characterized by a set of linguistic features which differentiate them from their domestic varieties. This thesis is based on the analysis carried out by Mori (2018c) during the *Eurolect Observatory Project*, where the Italian corpus of European directives (Corpus A) was compared with the corpus of their Italian laws of implementation (Corpus B). In order to answer the main research question of the *Eurolect Observatory Project*, i.e. “Does a Eurolect exist in all or any of the eleven languages here considered?” (Mori, 2018b: 12), a corpus-based and corpus-driven analysis was carried out. Eventually, results “[...] confirmed the existence of an Italian Eurolect” (Mori, 2018c: 199).

Recently, the *Eurolect Observatory Project* has broadened its research to include the Swiss Confederation. Despite Switzerland not being part of the European Union, it is located in the middle of Europe. In this scenario, the European Union and the Swiss Confederation are bound to each other, both from a political and an economic point of view. This setting led to the drafting of the so-called bilateral agreements. These bilateral agreements are the basis for the construction of the CHEU-Lex Corpus² (5,266.714 tokens), a parallel and comparable corpus created with the aim of providing a richly annotated multilingual resource to investigate the influence of the European legal drafting and translation practices on Swiss

¹ <https://www.unint.eu/en/research/research-projects/33-page/490-eurolect-observatory-project.html>

² <https://transius.unige.ch/en/research/chEU-lex/>

legislation. The CHEU-Lex Corpus was built in a joint project between the University of Bologna and the University of Genève in which also the author of this thesis took part. It was led by Professor Annarita Felici (UniGe) and coordinated by Professor Adriano Ferraresi (UniBo). The project aimed at building the CHEU-Lex Corpus *ex novo*, providing the Swiss legal texts and, more in general, the setup on which this study is based. The corpus is made up of three sub-corpora, i.e. the Italian sub-corpus (1,762,261 tokens), the French sub-corpus (1,985,041 tokens) and the German sub-corpus (1,519,412 tokens); each of these sub-corpora comprises the bilateral agreements entered between the Swiss Confederation and the European Union from 1972 to 2017 (called *agreements*), and the Swiss federal legislation representing the reception of these agreements (called *laws*).

Against this background, this thesis compares the Italian of Swiss laws of implementation represented by the CHEU-Lex Italian sub-corpus, both to Mori's corpus of European directives (Corpus A) and to the corpus of their Italian laws of implementation (Corpus B). The aim is that of observing any relevant feature that could be linked to the possible influence of the European legal drafting process.

For the purpose of this study, data were extracted from the Italian component of the CHEU-Lex Corpus. These data were then compared to the results obtained by Mori (2018c). To properly contextualize the study, the thesis also discusses the environment in which these varieties of Italian are produced and previous studies that have paved the way for this kind of research.

The first chapter of this thesis introduces the background in which this research was born; particular attention is devoted to the European drafting and translational process; then, small sections are devoted to the description of the so-called Eurolects, legal Italian and the newly born field of comparative jurilinguistics. Eventually, the *Eurolect Observatory Project* is described.

The second chapter introduces the research question, as well as the different implementation procedures carried out both in Italy and Switzerland. Then, the creation of the CHEU-Lex Corpus, in which the author of this thesis took part in, is set forth. The last section of the second chapter is devoted to the description of the method followed during the research, which closely follows that adopted during the *Eurolect Observatory Project*.

Finally, the third chapter discusses the data analysis; each sub section refers to different linguistic levels. Here, data are visually represented in tables and graphs. Eventually, conclusions are drawn in the last section, highlighting the results of this thesis as well as the limitations and future developments.

CHAPTER 1

1.1. Introduction

This chapter outlines the background on which this thesis is based. In Section 1.1. particular attention is devoted to the law-making process in the European Union, including drafting and translational stages and their relative problems. The description of the so-called *Eurolects*, the European languages which emerged from this unique setting, follows in Section 1.2. Section 1.3. describes legal Italian, followed by an outline of the newly born discipline of Comparative Jurilinguistics in Section 1.4. Eventually, the *Eurolect Observatory Project*, the research project which provided the basis for this thesis, is described in the last Section.

1.2. Drafting and translational process at the European Union

The unique multilingual situation in which European law operates is part of the greater, complex lawmaking process established by the European Treaties. In the earliest days the challenges of the choice of simultaneously drafting legislative texts in a variety of languages were largely ignored, but nowadays this situation is changing. The Court of Justice is constantly dealing with problems concerning inconsistencies among the various language versions of the EU legislation (Graziadei, 2015). There is a strong need for legal harmonization across Europe, which is very hard, if not utopic, to achieve, and the main reason for this is the multicultural environment in which the law needs to be produced (Šarčević, 2015).

1.2.1. The EU and its multilingual environment

The first thing to understand is the unique setting in which laws are produced: the European Union (EU). Defined as “[...] not a state, but a unique partnership between European countries, known as Member States” (Directorate-General for Communication of the European Commission, 2021), the EU is currently made up of 27 Member States. Due to the large number of States belonging to this union, the reality in which laws are drafted and created is multilingual. This characteristic gives rise to difficulties to different extents. For the purpose of this thesis, it is essential to underline the linguistic problems which characterize this situation.

One of the main peculiarities that characterize the European Union is that all official languages³ are afforded equal status. This reflects the *principle of equality* to which every European citizen is subject (Pozzo, 2006). Also, the need for a multilingual environment complies with Article 6 of the *Treaty on European Union* (Eur-Lex, 1992)⁴, which promotes the respect for the national identity of each Member State (Pozzo, 2006). Turchetta (2005) also argued that whilst the EU is an international organization, it is not entirely comparable to other international organizations such as the United Nations, OMS etc. since the EU aims at integrating the Community legal system into the legal systems of each Member State. This requires greater efforts in terms of equality, since the legislation of the European Union has a direct impact on each Member State, and, whether a national law comes into conflict with European law, the latter shall prevail. (Turchetta, 2005).

1.2.2. Drafting process

Taking into consideration the various language contact scenarios which can be found in Western cultures, one of the most peculiar happens to be in the European Union institutions. In this context, it is extremely important to be able to share common principles and values, as well as to agree on financial and economic rules. In such a multilingual environment, issues related to the internationalization of laws are crucial. In the EU law-making process, it is essential to guarantee equality among the Community languages. Also, unlike what happens during an ordinary translation, at the beginning of the EU law-drafting process, there is no source text and no target text, “[...] but rather a primary text, which is amended and changed by several contributors at different stages in different languages” (Mori, 2018b: 6).

In such context, texts are translated from one or two source languages and must be in line with the requirement for parallel versions, following a three-step procedure including drafting, translation and legal revision. It is fundamental to remember that, according to the *principle of equality of the language version*, every language version has the same legal force in case of divergences, which means that each translation has equivalent legal validity by virtue of the authenticity principle.

³ To the present day, European official languages are 24: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish.

⁴ The full text is available online at the Official website of the European Union <https://eur-lex.europa.eu/homepage.html>.

As it is stated by Mori (2018c: 7),

“[...] According to Robertson, EU legal texts are produced by applying the synoptic approach: « each language version of a text has the same number of pages, the same structure in the text, the same numbering and paragraphing, the same sentence length, and the same information is given at the same point in each language version » (2013, p.21)”

The drafting process which is going to be described below is adapted from Robinson (2008), but it can also be checked online on the European Parliament website.⁵

Almost all European acts are drafted by the Commission, which starts the legislative initiative. Indeed, it is the Commission that decides whether it is appropriate to propose legislation. The other institutions can ask the Commission to present a proposal. However, the latter is not obliged to do so.

The Commission counts over twenty-five technical departments, which are also called Directorates-General (DG), and each of them deals with the different sectors of the EU's activities. They are also responsible for preparing the first draft of the legislative acts and implementing acts in their sectors. These drafts are usually produced by technical experts who rarely have specific drafting expertise. Interestingly enough,

“The drafting language is determined by the DG. There is no requirement for a drafter to be a native speaker of the language concerned and in fact that is rarely the case. A draft act passes through all the internal discussion stages within the Commission in just one language, but it must be translated into all the official languages before it can be submitted for adoption by all the Members of the Commission, the College.” (Robinson, 2008: 3)

Once the draft is ready, it is submitted to the other DGs as part of the Inter-Service Consultation, designed to check that the Commission works in an effective and coordinated manner, and which includes consulting the Legal Service and the Legal Revisers Group.

The Legal Service must be consulted about all files that have legal implications. It has a staff of almost 400 people and acts as the Commission's in-house lawyer. It represents the Commission before the European Court of Justice and any other juridical body, and it reports

⁵ For further information, please visit https://www.europarl.europa.eu/infographic/legislative-procedure/index_en.html

directly to the Commission President. It also has 11 teams of lawyers covering the various sectors (Robinson, 2008). In the Inter-Service Consultation, a lawyer specialized in the sector must check the lawfulness of the measures proposed and, above all, the compatibility with the Treaty, other provisions of legislation and international obligations. Separately from the specialized lawyers, two other groups are responsible for the quality of legislation; the first group is responsible for the drafting quality and the second for codification, bringing an existing act and all its amendments together in a whole new act.

The Legal Revisers Group is made up of almost sixty legal revisers, divided into three sub-groups, each of them specialized in various sectors. Each reviser must have a legal qualification and language skills, including at least French and English. At the Inter-Service Consultation stage, revision is done following two parallel circuits: checking lawfulness by the sectoral lawyers on one side and examining form and presentation by legal revisers on the other.

The Secretariat-General is responsible for the operations of the Inter-Service Consultation and the coordination of the Commission's work. It keeps track of the decisions made by the Commission and transmits documents to the other institutions or the Office for Official Publications of the European Union (OPOCE).

Particular attention is paid to translation: all legislation must be translated into all the official languages before adoption (Robinson, 2008). All translations are produced by the Translation Directorate-General (DGT), which counts a staff of around 1,500 permanent translators, plus a network of freelancers. Robinson (2008: 6) claims that

“With its teams of highly skilled and trained linguists who have excellent human and computer back-up, the Translation Directorate-General is able to guarantee that all the language versions of the EU legislation will say the same thing.”

Once all these steps are completed, the proposal is passed to the legislative authority, i.e. the European Parliament and the Council. They act together during the co-decision procedure.⁶ The cooperation between the European Parliament and the Council is a long process, and the proposal might bounce from one institution to the other until it is either approved or rejected.

⁶ However, sometimes the Council may act alone.

The procedure starts with the first reading in Parliament, where the Commission's proposal can either be approved without modification or amended. Then the text is sent to the Council to be read for the first time. The Council can either accept the Parliament's position – in which case the legislative act is adopted – or amend it and return the modified proposal to Parliament for a second reading. During this second reading, the Parliament examines the Council's position and can either approve it – which would lead to the approval of the act - or reject it; if this is the case, the act will not enter into force, and this could be the end of the procedure, unless the Parliament proposes amendments and returns the proposal to Council for a second reading. During this second reading, if the Council approves all of Parliament's amendments, the act is finally adopted; if this is not the case, the intervention of the Conciliation Committee becomes necessary. The latter will try to reach an agreement between the Council and the Parliament on a joint text. If unsuccessful, the act will not enter into force and the procedure is ended; if instead, the joint text is agreed, it is forwarded to the Parliament and the Council for a third reading. During this last reading, neither the Parliament nor the Council can change the wording; they can only accept or reject the text. The act will be adopted only if both the Parliament and the Council approve it.⁷

1.2.3. Drafting and translational issues

After having introduced the drafting procedure, particular attention in this paragraph is addressed to issues related to the multilingual environment of the European Union.

Although this feature is undoubtedly considered a treasure-source of European culture, it is at the same time “[...] source of innumerable problems, when it comes to drafting, translating and interpreting acts produced by the Community institutions in all the various official languages.” (Pozzo, 2006: 4). Needless to say, the EU multilingual nature can be guaranteed only through translation. However, since according to the *principle of equality* all language versions must have the same legal force to ensure equality among the Member States, it was necessary that the “translational process” was not considered part of the legislative process, as it would have questioned the principles of unity and universality of law. In the words of Turchetta (2005: 75-76):

⁷ The guide can be consulted online at https://www.europarl.europa.eu/infographic/legislative-procedure/index_en.html#step1

“La legislazione comunitaria, in quanto legislazione plurilingue, poggia su una ‘finzione giuridica’ particolare, secondo cui, teoricamente, sarebbe possibile dare una formulazione identica a ciascuna delle lingue ufficiali. Tuttavia, se non si vogliono rimettere in discussione le proprietà basilari di unità e universalità della norma, la traduzione non deve essere riconosciuta come una fase del processo legislativo.[...] Per ovviare a questo *impasse*, si fa riferimento ad una stesura contemporanea di ‘testi paralleli’, redatti simultaneamente in più lingue durante una fase di ‘coredazione’: nel linguaggio giuridico si definiscono paralleli i testi legislativi bilingui o multilingui che si configurano come autentici strumenti legali dotati di identica autorevolezza.”

The solution proposed above was a shallow solution which aimed at solving a deeper issue: although at a legal level “translation” must not be part of the drafting process, in practice, legal drafting in the European Union would not have been possible without it. This *camouflage* can also be observed in the document published by the Council of the European Economic Community (1958), i.e. the *Regulation No 1* concerning the languages to be used within the Community, where no reference is made to an *original text* and its *translated* versions, but only to the four official and working languages, which at the time were Dutch, French, German and Italian (new languages were subsequently added; Turchetta, 2005; Felici, 2015). In the same document, the verb “to translate” is carefully substituted with the verb “to draft”. Also, Felici (2015) states that considering all the documents, it is almost impossible not to notice how ‘original’, ‘official’, and ‘working’ are considered on the same level. However, these rules were drafted back in the late 1950s, when parallel drafting with four official languages could still be considered a feasible option (Felici, 2015).

Various issues arose due to the multilingual nature of the EU. One of the main problems occurred when two (translated) versions were not consistent: since all the official languages shall be deemed authentic and no version shall prevail over the others, which one should be considered correct? To solve this problem, the Court of Justice set up a list of criteria to establish which should be the right one. If inconsistencies were found among two or more versions, all the linguistic versions must be taken into consideration to dispel any doubt. Unambiguous translations would have been the preferred versions. If, however, the problem persisted, further interpretation according to various criteria should be carried out:

“Là dove, anche applicando questi principi, la divergenza non viene risolta, la Corte ricorre a un’interpretazione in funzione dell’economia generale, della

volontà dell'autore e, in particolare, della finalità perseguita (Sentenza Erich Stauder v/ville d'Ulm – Sozialamt, 29/69, recepita nel 1969).” (Turchetta, 2005: 77)

Nowadays, the procedure to produce law in the EU is to firstly draft the text base in a commonly agreed language, or *lingua franca*, and then translate it into the other official languages. After the accession of the United Kingdom to the European Economic Community (EEC) and the enlargement of 2004, English has become the unofficial *lingua franca* together with French, although, to the present day, the latter is giving way to English (Felici, 2015). Moreover, the use of English as the main drafting language has almost doubled, and it seems as though this trend is going to go on in the future.

1.2.4. Europeanisation process

To overcome the problems discussed in Section 1.1.3., multiple attempts have been made. In 2016, the European Union published the *Joint practical guide of the European Parliament, the Council and the Commission for persons involved in the drafting of European Union legislation*, establishing a rigorous method to be followed when drafting the EU law. Interestingly enough, unlike the previous *Regulation N° 1*, here references are made both to *translated versions* and *original texts*. This underlies the fact that translation has eventually been recognized as an essential and unavoidable step during the legal drafting process:

“First, the original text must be particularly simple, clear and direct, since any over-complexity or ambiguity, however slight, could result in inaccuracies, approximations or complete mistranslations in one or more of the other Union languages.” (European Union, 2015: 16)

Another example that shows the engagement of the European Union in achieving harmonization is the guide published in 2011, *How to write clearly*. According to the guide, the European Commission staff is in charge of writing several different documents, be they legislative texts, technical reports, press releases etc. No matter the type, “[...] a clear document will be more effective, and more easily and quickly understood.” (European Commission, 2011: 2). Also, one of the main points made extremely clear in the document is the fact that particular attention must be paid to the *reader*. This term refers both to the end-users (since most European Union documents are available online, this includes any

European citizen who needs to read the document), but also to EU insiders (colleagues inside the European Commission or in other institution), outside specialists, and, obviously, revisers and translators. Trying to draft a document while keeping in mind all these future *readers* is quite complicated, but it is an effort that is necessary to achieve the drafting of legislative texts as clear as possible and accessible in all the 24 official languages. These are just a few of the multiple guidelines given in the document. Other hints suggest, for instance, that the structure of the document should be *short* and *simple*, and that ambiguity should be avoided. Hints on verb tense, grammar and morphosyntax are also given. All of them aim at creating a simple and accessible document, avoiding the use of complex and sometimes obscure language which usually characterizes bureaucracy (European Commission, 2011).

This is just but a small part of the attempts made by the European Union to create unity out of (linguistic) diversity. To achieve this goal, different processes of *Europeanization* are at work: first, the creation of a common European Language to draft European law that would be equally foreign to all Member States, but also the creation of a common European legal culture of shared values and standards (Šarcevic, 2015).

To achieve a greater degree of harmonization, comparative law scholars have cooperated to try to create a common terminology in areas of private law, aiming at removing as many barriers to cross-border transactions as possible. As stated by Maja Bratanić and Maja Lončar (2015), this terminology must be easily accessible. This could be made possible through the *Interactive Terminology for Europe* (IATE), i.e. the EU inter-institutional terminology database (EU law - EUR-Lex, 2021), as well as *EuroVoc*, the EU's multilingual thesaurus (EU law - EUR-Lex, 2021), together with a plethora of internal translation and documentation tools (Bratanić and Lončar, 2015). Nonetheless, it seems as though, despite legal language and terminology are less favorable to change due to their highly specialized nature,

“[...] the ‘Europeanization’ of law and its multiplication in an already impressive number of terminologies certainly makes room for uncontrolled variation in spite of the ongoing process of concept and term harmonization mentioned earlier.”
(Bratanić and Lončar, 2015: 217).

Despite the various, ongoing attempts to harmonize and standardize communication within the EU, the birth of a set of specialized European legal languages called Eurolects was almost

unavoidable. Due to their unique nature, these special languages gave rise to different linguistic studies which will be further discussed in the following paragraphs.

1.3. The languages of the EU: the Eurolects

After having introduced the context in which laws are being drafted, it is essential to focus on the means by which European legislation is being created. Besides various tools and methodologies already mentioned in Section 1.1.4. and concerning how to write laws clearly, the main tool to be analyzed is, indeed, language. The fact that legislative texts are being drafted in different languages outside their respective nations and in a completely new setting provides the optimal environment for the birth of new varieties of the European languages.

1.3.1. The birth of the term “Eurolect”

The written language used in the European Union has been given different names, the majority of which bear negative connotations. Some interesting examples are *Eurobabillage* or *Brouillard linguistique européen* in French; *Eurowelsch* or *Eurokaudelwelsh* in German; in English *Eurospeak*, *Eurobabble*, *Europese*, *Eurofog* but also *Dadefinspeaking*⁸ and more recently *Euro-English*⁹; Italian has its versions as well: *Eurocratese*, but also *Comunitarese*, similar to *Burocratese*, which negatively connotes the complex and sometimes obscure language used in bureaucracy (see Mori, 2018; Goffin, 1994; Turchetta, 2005). Although it is legitimate to talk about a new language forged within Europe, Goffin asks whether it is righteous to talk about a *jargon*, or whether it would be more appropriate to talk about a -*lect*:

“Est-il pourtant légitime de parler de jargon [...] ? Reconnaissons que le terme jargon véhicule des contenus qui se diluent et que la terminologie linguistique ne contribue guère à donner une image précise. Le mot jargon a pris des connotations abusivement péjoratives. [...] Peut-on objectivement coller ces étiquettes sur le langage communautaire dont la majorité des termes, qu'ils soient ou non nouvellement forgés ou qu'ils soient utilisés dans un sens spécifique, ont fait

⁸ This one is, interestingly enough, an acronym referring to the official languages before 1984 (da = danois, d = deutsch, e = english, f = français, i = italien, n = néerlandais).

⁹ Euro-English was coined in the late '90s when some scholars hypothesized the existence of a Euro-English variety with continental patterns. This was due to the fact that the widespread use of English as a *lingua franca* gave birth to several non-native varieties, which differed from the original British and American English (Felici, 2015).

l'objet de définitions d'ailleurs parues au Journal officiel [...] ? Ne faudrait-il pas plutôt admettre que ce langage malgré sa spécificité ne se démarque, ni par ses caractéristiques sémantiques ni par ses mécanismes morphologiques et syntagmatiques, ni par ses constructions syntaxiques des autres *lectes*, tels que les technolectes, les chronolectes, les régiolectes, les gynolectes ou les androlectes ?” (Goffin, 1994 : 637-638)

In his article, Goffin concludes by stating that thanks to its nature, its origins, and the way it is put into action, this specific variety of language can be righteously called *Euro-lect* (1994). Also, Turchetta (2005) claims that the EU language can be defined as *Eurolect* since it is a specialized language, and it is possible to identify a bureaucratic variety and a legal variety within it. It is a subcode deeply rooted in the European context. Due to this layered structure, where three dimensions interact with each other (specialized, bureaucratic, and legal), *Eurolect* has been described by Nystedt (1998) as a triple-layered variety. Later, Goffin (1997) argued that it is “[...] a Language for special purposes, not limited to specific lexical items and terms but assuming features at other language levels”.

1.3.2. The Italian Eurolect

Within the *Eurolect Observatory Project* (see Section 1.5.), the term *Eurolect* refers exclusively to the EU legal variety differing from the corresponding national legal one (Mori, 2018b). This means that the term is used to identify the language of law only. In particular, the term is adopted when differences between the EU legal variety and the domestic variety are observed at different linguistic levels, such as lexical or morphosyntactic (see chapter 2).

Traces of the existence of an Italian Eurolect have been indeed observed by Mori in her research (2018c), which was part of the bigger *Eurolect Observatory Project*.

However, it is important to underline the fact that the European Union is not the first institution using legal Italian outside national borders. Chronologically speaking, the first official variety of extra-national Italian is the so-called *Swiss Italian* (Turchetta, 2005). The two varieties of Italian have much in common: for instance, similarly to the Italian Eurolect, Swiss Italian is the result of the need for the establishment of a multilingual legal regime (trilingual in the case of Switzerland and multilingual in the case of EU). Another analogy

with the European Union is that, by virtue of the principle of the trilingual Swiss system, each official language has the same legal force (Turchetta, 2005).

1.4. Legal Italian

Since this thesis aims at comparing two varieties of legal Italian, it is necessary to understand what legal Italian is. Therefore, this paragraph aims at describing this linguistic variety, as well as discussing the characteristics of legal language.

1.4.1. Historical background

Origins of *legal Italian* can be traced back to the Medieval period, when notaries, besides other tasks, had to translate law and documents – which at the time were in Latin, the language of legal proceedings (Mattila, 2006) – to their clients, who often spoke only vernacular. Later, during the XVII and XVIII centuries, legal Italian started to affirm itself in various fields of legal practice (Mori, 2018c). Therefore, it was first shaped starting from Latin but was later affected by European influences, in particular the French legal tradition. Also,

“[...] the Roman law-based system was influenced by the French Code, and new legal words or new meanings for existing words arose. Therefore, quite interestingly, Italian Legal and administrative language originated from a several century-long translational process”. (Mori, 2018c: 200)

Since the Middle Ages, then, contact with different languages has shaped legal Italian, starting from the bilingualism Latin – vernacular to French, as well as the influence of the German conceptual apparatus. More recently, legal Italian has been influenced by French and English within the European multilingual context (Mori, 2018c).

1.4.2. Defining legal Italian

Defining legal Italian is not straightforward. Nonetheless, it is safe to say that legal language does not qualify as a *common language* such as French, Italian or German, for instance. Instead,

“[...] it operates as a functional variant of natural language, with its own domain of use and particular linguistic norms (phraseology, vocabulary, hierarchy of

terms and meanings). Legal language possesses a number of specific features. These are morphosyntactic, semantic and pragmatic.” (Mattila, 2013: 1)

According to Mantovani (2008), legal language is a *diaphasic variety* of the common language – Italian – which comprises two distinct species, namely the language of law and the language of jurists. It is *diaphasic* in the sense that it strongly depends on the extralinguistic context, and, in particular, on the specific field of communication (Adinolfi et al., 2008). More generally, it can be defined as a *sub-code* used by a group of speakers within the *common language*, to fulfil specific communicative needs.

Similarly, Caterina & Rossi (2008) state that legal Italian is a *sub-domain* of the Italian language which, although being highly formal, is intended not only for experts but also for semi-experts and non-experts. In particular, they claim that:

“[...] il linguaggio giuridico mira all’univocità e alla chiarezza referenziale, attraverso la ricercatezza delle scelte lessicali, le complicazioni morfosintattiche, e i particolarismi nell’organizzazione testuale, ma al tempo stesso non si volge mai verso una tecnicizzazione completa, lasciando sempre aperta la strada a fertilizzazioni e contaminazioni con il linguaggio ordinario e con altri domini linguistici specialistici.” (Caterina & Rossi, 2008: 185-186)

One of the main aims of this sub-genre is to produce an extremely clear and unambiguous text, and this is done through the use of specific features such as lexical choices or morphosyntactic structures.

However, legal language is not entirely separate from the ordinary language, but, instead, it has some features in common with it. At a formal level, legal Italian shares morphological and syntactical rules with *common Italian*, as well as the vast majority of the lexicon. Indeed,

“It is clear to see that legal language is based on ordinary language. For that reason, the grammar and – in general- the vocabulary of legal language are the same as in the case of ordinary language.” (Mattila, 2016: 1).

Despite this, legal Italian remains a distinctive variety, not only thanks to its specific terminology but also to the frequency of certain morphosyntactic characteristics, which represents a marker with respect to common language (Adinolfi et al., 2008).

It is also worth mentioning that legal language, despite being itself a *sub-genre*, can be divided into further sub-genres reflecting the various sub-groups in which also law is divided, ranging from the language of legislators to that of judges, administrators, and advocates (Mattila, 2013). This depends on the fact that the language of each sub-group possesses, to some extent, particular characteristics which differentiate it from the languages of the other sub-groups. To give just a few examples, courtroom language is particularly formal and often uses archaic terms or syntax, it has a categorical character, since judges “[...] use unreserved declaration and peremptory orders” (Mattila, 2013), whereas another branch such as criminal law language contains terms which are never used in other branches of legal language.

To conclude, keeping in mind that legal language – and therefore legal Italian – ranges from the language used in institutions to that used in various situations related to the legal community, and that it can also be further divided into sub-genres, it is necessary to define the specific sub-language on which this research is conducted: for the purposes of this thesis, legal Italian is observed by focusing on the Italian legal language used in the Swiss Confederation when implementing the bilateral agreements. It is compared both to the Italian of EU directives and the Italian used to implement the latter.¹⁰

1.5. Comparative Jurilinguistics

To better define the field to which this type of research belongs, it is necessary to talk about comparative jurilinguistics, which can be considered as a branch of linguistics.

1.5.1. Historical background

The modern sense of the term *linguistics* was developed in the first half of 1900 thanks to Ferdinand de Saussure, who established a synchronic study of the language system. After Saussure, different schools emerged, and each of them gave their contribution to the development of this discipline. This made it possible to examine languages from a modern point of view, which was lately applied to the language of the law as well. It is on this basis that research on legal language has been developed, paving the way for legal linguistics (Mattila, 2006).

¹⁰ It is essential to remember that the domestic legal Italian analyzed is the specific variety used in Italy when implementing the EU directives.

Although legal linguistics is a relatively new discipline, legal language has been studied for years from various perspectives. Law is undoubtedly bound to language, and in this sense legal language has existed as long as the law (Mattila, 2006). Not surprisingly, the oldest translated text that has survived until today is the peace treaty in two languages between the Egyptians and the Hittites, dated 1271 B.C. (Mattila, 2006; Mattila, 2013). There followed countless legal translations, but even so, it has been necessary to wait until the last Century to see the birth of this discipline.

1.5.2. Combining legal and linguistic studies: the birth of jurilinguistics

As explained by Mattila (2006), researchers with different backgrounds – linguists and lawyers - have been analyzing legal language for almost a century now. However, due to their different professional background, they tackled the research from different standpoints.

On one side, linguists usually analyzed legal language by applying quantitative methods to their studies, investigating how certain language elements behave in legal language compared to their use in common language (Mattila, 2006). This type of research was implemented through the use of computers and the birth of Computational Linguistics, i.e. the field of study which combines computer science and linguistics with the aim of understanding and generating natural language (Grishman, 1986). Computational Linguistics is also a synonym for Natural Language Processing (NLP) (Basile, 2020).

On the other, lawyers always tackled these studies from the inside: their research is often

“[...] of a diachronic nature and have a close connection with legal history. They try to make the characteristics and the vocabulary of the language of law intelligible to outsiders, taking into account the needs of the legal system.”

(Mattila, 2006: 22)

Nonetheless, the viewpoints of linguists and lawyers can be combined in a specific study, where research optics are defined by the needs of the lawyers, and research methods are provided by linguistics. This kind of study is defined as “jurilinguistics” (Mattila, 2006: 23).

Nowadays, this field of study has increased its relevance since lawyers have to cooperate more and more often with their foreign colleagues. Indeed, thanks to globalization, the use of different languages for international cooperation is becoming more and more common. This context implies that lawyers learn to use foreign legal languages well, and this applies

in particular to foreign legal terminology and other linguistic features which belong to the juridical use of these languages (Mattila, 2006). However, lawyers do not cooperate only with other foreign lawyers, i.e. colleagues from other countries, but they cooperate “[...] with language specialists, terminologists and translators” as well (Mattila, 2006: 27).

Being able to properly handle such competences can obviously take a lot of time, but it can also be speeded up thanks to background information about the legal language(s):

“If lawyers receive general knowledge about the history, vocabulary, basic characteristics and use of foreign legal languages, as well as about their relationship to each other, they may learn the specific use of these languages in their own branch of law more quickly and easily. On the other hand, lawyers [...] often must work with language specialists, terminologists and translators. A general knowledge of legal languages then helps lawyers to better understand the basis of translation activities and terminological work.” (Mattila, 2006: 27)

But, to do so, it is necessary to do research on those languages. This type of research has been particularly profitable in multilingual countries such as Canada or the Swiss Confederation. However, it goes without saying that in the European Union, with its 24 official languages, this type of research has been deemed necessary as well. As stated by Mattila (2006), a lot of work is still to be done to clarify the properties of the various legal languages and their relationship to each other. To tackle this goal, studies combining legal needs and linguistics methods may be very useful. “We could speak about ‘comparative jurilinguistics’” (Mattila, 2006: 27).

1.6. Background: The Eurolect Observatory Project

This thesis is based on *The Eurolect Observatory. Interlingual and intralingual analysis of EU legal varieties* project, which has been launched in December 2013 at Università degli Studi Internazionali di Roma (UNINT). It starts from the hypothesis that

“[...] language contact through translation of EU legislation has resulted in the creation and dissemination of standardized lexical variants, structural features and textual patterns in many EU official and working languages.” (Mori, 2018b:

1)

The project aims at verifying the existence of specific varieties of legal language as a result of the drafting process of EU laws.

1.6.1. Research project

The *Eurolect Observatory Project* focuses on the results of the multilingual law-making process by observing peculiarities of legislative language across Europe. In particular, the aim of the project is to detect any possible evidence of the existence of a European legal variety for at least some of the 24 official languages of the European Union. These European legal varieties are, indeed, the so-called Eurolects (see Section 1.2.). The main research question, which is “Does a Eurolect exist in all or any of the eleven languages here considered?”, can be addressed if specific Eurolectal features emerge thanks to the drafting-translating-revision process.

The project had a duration of 7 years, and it was divided into two different phases: the first one starting in 2013 and ending in 2016, and the second one starting in 2017 and ending in 2020. Led by Laura Mori (Università degli Studi Internazionali di Roma), the project’s goals are the description of the dynamics of linguistic contact-induced phenomena and the provision of reference data to eventually improve the quality of legal texts produced by national institutions.¹¹ To achieve these aims, the *Eurolect Observatory Multilingual Corpus* (EOMC) was created (see Section 1.1.2.).

To confirm the existence of legislative varieties at different levels, a common research template conceived by Mori as a reference was adopted.

The template, which will be further discussed in this thesis (Section 2.4.1.), outlines three different macro-areas of research (EU-rooted phenomena, contact-induced phenomena, and intra-linguistic variability) and selects the expected areas of linguistic interests (Mori, 2018b).

¹¹ For further information please visit <https://www.unint.eu/en/research/research-projects/33-page/490-eurolect-observatory-project.html>.

1.6.2. The creation of the *Eurolect Observatory Multilingual Corpus* and the research tools

The Eurolect Observatory Multilingual Corpus (EOMC) comprises a total of 22 sub-corpora in 11 different languages: Dutch, French, English, German, Italian, Maltese, Polish, Spanish, Finnish, Latvian and Greek. Two different corpora have been compiled for every language: one containing EU directives (called “Corpus A”) and one containing their respective laws of implementation (called “Corpus B”). Due to the various orthographic systems, all texts were encoded in Unicode UTF-8 and were downloaded either as plain text or as HTML.

As mentioned above, all versions of Corpus A are a collection of European Union Directives, published between 1999 and 2008 and available on the official EU law website Eur-Lex.¹² Corpora named Corpus B are a collection of national laws of implementation of their respective EU directives.

Both Corpus A and Corpus B were then processed by means of a script, in order to remove all the markup tags, fill the texts with header and structural information and save the results in a UTF-8 format. Thus, for each language, Corpus B endorsed the same structural markup as Corpus A (Tomatis, 2008).

Eventually, for what concerns queries both on Corpus A and Corpus B, to perform the research the user needs to insert some parameters in the search engines, such as the specific text section or keywords; data are retrieved when at least one parameter is set.

1.6.3. The research and the results

To tackle the main research question, that is whether a legislative EU variety named Eurolect exists for some of the languages given, a mixed method was applied: a qualitative analysis of data was carried out and combined with quantitative analysis. Quantitative data can also be either corpus-based or corpus-driven. Eventually, results discussed in *Observing Eurolects* (Mori, 2018d) showed the existence of nine out of eleven Eurolects diverging from their national varieties.¹³

¹² <http://eur-lex.europa.eu/homepage.html>

¹³ For Latvian and Maltese, the hypothesis of concerning the existence of a national Eurolect variety was not confirmed (Mori, 2018d: 371).

Based on these analyses, languages can be grouped according to differences and similarities regarding the language policy for EU law implementation. (Mori, 2018d). The presence (or absence) of similarities in multiple Eurolects might be due to extra-linguistic conditions, related to the contact scenario where Eurolects develop, that is to say, the already-mentioned drafting-translating-revision process that characterizes EU laws. Indeed, the need for linguistic uniformity across the European languages, aiming at enhancing cross-linguistic comparison, results in a set of Eurolectal features which have spread across the Eurolects (Mori, 2018d). These features are grouped into three macro-categories, also proposed in Mori's Research Template: EU-rooted phenomena, contact-induced phenomena and intralinguistic variability.

EU- rooted phenomena mainly concern the lexical level, and include, among the others: europeisms referring to concepts, institutions policies etc. which are highly contextualized in the EU environment, semantic calques with a specific EU meaning, the production of "international affixes" regarding the lexical morphology, and a higher presence of lexical bundles.

Contact-induced phenomena are prompted by the literal reproduction of the original text which happens to be the presumed source text¹⁴, the latter being the best guarantee of a legal equivalence. Also, before the year 2000 English and French were the two languages used to draft laws almost to the same extent, whereas, after the 2004 enlargement, English became the official *lingua franca* of the EU. Therefore, during the description of this type of phenomenon, great attention has been devoted to the influences of English and French on other languages. Examples of contact-induced phenomena can be loanwords or calques, either structural or semantic. Word order was also affected, together with some English-induced graphic choices (Mori, 2018b).

Intra-linguistic variability embraces differences that might occur between corpus A and corpus B, analyzing therefore the gap between the well-established domestic legislative varieties on one side and legislative varieties developed in the EU context on the other. A corpus-driven analysis was carried out in order to observe lexical variability. Register-related issues are underlined, and, among others, Latinisms can be considered register-raising devices. The difference in the range of tenses is also taken into consideration,

¹⁴ It is important to remember that translation in the EU environment is not a "source text – target text translation", but rather the product of a multilayered process.

whereas on the textual level anaphoric and cataphoric textual references are also considered as a yardstick.

Particular attention must be paid to the Italian analysis since it is the starting point of this thesis. The results obtained, analyzed by Mori (2018c), provide proof of the existence of an Italian Eurolect, characterized by various phenomena, all falling within the three categories mentioned previously. Europeisms, lexical and syntactic calques, longer Eurolect-related bundles and a smaller lexical variation are the main features analyzed. The research showed that, in the case of Italian, the language-contact phenomenon led to the creation of more accessible directives, especially to citizens, heading towards the direction of language simplification. Indeed, legislative texts written in Italian Eurolects better fulfil the requirements of a plain legislative language, being the latter more accessible, while the legislative Italian of national implementation seems to be still more complex.

CHAPTER 2

2.1. Introduction

After having defined the background on which the study in this thesis is based (Chapter 1), this chapter introduces the main research question (Section 2.1), as well as the procedure adopted to implement the European legislation in the Swiss Confederation (Section 2.2). The creation and the structure of the CHEU-Lex Italian sub-corpus follows (Section 2.3). The CHEU-Lex corpus is the result of a joint research project that aimed at creating a trilingual, parallel and comparable corpus of Swiss and European Union legislation, in which the author of this thesis took part. In Section 2.4 the methodology used in this project is described in detail. Eventually, a brief paragraph on general remarks, limitations of this study and future developments is set out (Section 2.5). It is important to underline the fact that the methodology closely follows that of Mori (2018c), in order to compare the results obtained by Mori and those obtained in this study. However, Mori aimed at proving the existence of the Italian Eurolect, whereas the scope of this study is to observe any eventual linguistic influence of the European legal drafting process on the Italian of Swiss laws of implementation.

2.2. Research question

In the light of the previous chapter (Chapter 1), the scope of this thesis is to observe whether in the corpus of Swiss laws of implementation (CHEU-Lex Corpus) traces of the European legal drafting process can be found, and to what extent. To do so, results obtained from the CHEU-Lex Corpus were first compared to those of the corpus of Italian laws of implementation (Corpus B), and then to results obtained from the corpus of the European directives (Corpus A), both already analyzed by Mori in *Observing Eurolects* (2018c). The names of the two corpora will be left unchanged in this thesis. It is important to underline that the Italian of Swiss laws of implementation is represented exclusively by the *laws* sub-section of the Italian CHEU-Lex sub-corpus (henceforth named “CHEU-Lex corpus” for the sake of simplicity). This is because the Italian sub-corpus, as well as the other sub-corpora that compose the whole CHEU-Lex Corpus (i.e. German and French), is made up of two different types of texts: *laws* and *agreements* (further details on the structure of the corpus will be given in Section 2.3.1.). The *agreements* sub-section collects the texts of the bilateral agreements entered between Switzerland and the European Union (Université de Genève,

2021), whereas the *Laws* sub-section comprises the Swiss laws of implementation of the bilateral agreements. In this sense, the *agreements* sub-section can be considered similar to – but not the same as - the corpus of EU directives (Corpus A) analyzed by Mori, whereas the *laws* sub-section is closer to Corpus B, since they both gather national laws of implementation.

First, data obtained from the CHEU-Lex Corpus were compared to those of Corpus B, with the aim of observing any possible linguistic similarity (see Section 2.4.1.) between the two. To ensure the comparability of the data, these were extracted from the CHEU-Lex corpus following the same method used by Mori (2018c). Data extracted were then compared to those of Corpus B obtained by Mori both through the normalized frequency and by means of the Log-Likelihood ratio (Section 2.4.). This ratio detects any statistically significant difference in the use of the analyzed elements between the two corpora. Subsequently, data were compared to those of Corpus A using the same method. This second comparison is meant to provide further evidence, either confirming or disproving the previous results obtained from the first comparison (that between the CHEU-Lex Corpus and Corpus B). Evidence gathered from the data analyses are discussed in the results chapter (Chapter 3) of this thesis.

2.3. Two different implementation procedures: Italy and Switzerland

The implementation of European Legislation in the Italian law system is carried out through two types of translation: first, during the EU law-making process, legislative texts are drafted in one or two source languages¹⁵, and are eventually translated into all the official languages of the EU. This interlingual translation is the first one taking place at a European level. The Italian legal system must then ensure a rapid implementation procedure of EU directives through the European Delegation Bill (*Legge di delegazione Europea*) and therefore the second, intralingual translation is carried out: from the EU legislative Italian into the national legislative Italian (Mori, 2018c).

The reception of European Law in the Swiss Legislation, on the other hand, is slightly different. Although the EU law-making process is the same, the adaptation to the national legislation changes. As stated in *Règles d'or de la reprise du droit de l'UE dans le droit Suisse* (Swiss Federal Chancellery, 2019), the Swiss law of implementation is in German.

¹⁵ Nowadays, English has become the *de facto* European *Lingua Franca*.

This means that the German EU version inspires the German Swiss version, generating an intralingual translation. Subsequently, the German Swiss version needs to be translated into Italian and French as well, a step which leads to a third, interlingual translation.

The following images graphically represent the points mentioned above:

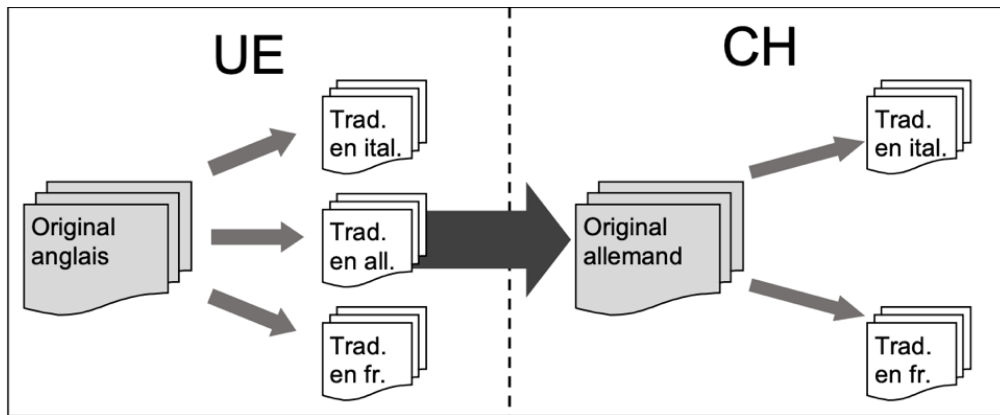


Figure 2.1. The implementation process from the first original draft of legal texts in the EU to the final translated versions in the Swiss Confederation (Adapted from the Swiss Federal Chancellery, 2019).

This picture (Figure 2.1) shows the procedure followed when generating the two final versions in Italian and French translated from German. However, this method implies that these translations will differ from their respective European versions, and this could cause problems in terms of direct comparability, as shown below (Figure 2.2):

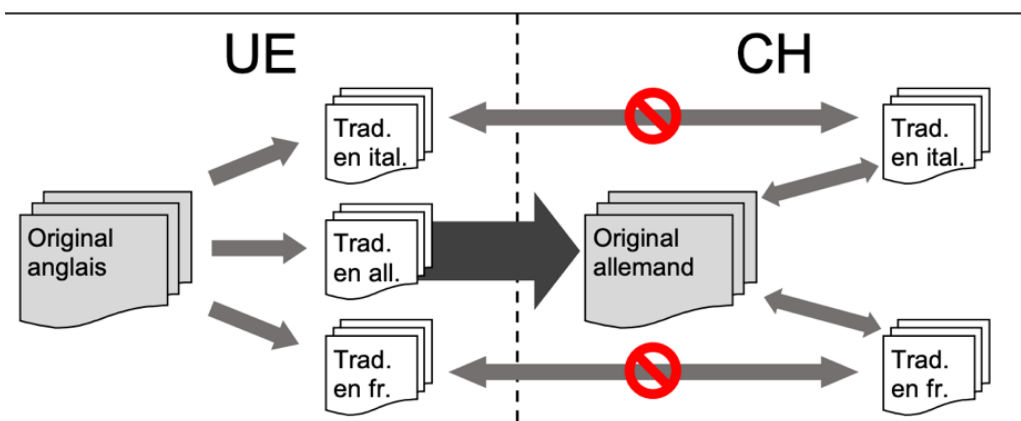


Figure 2.2. Discrepancies between the translated CH versions and their respective UE versions. (Adapted from Swiss Federal Chancellery, 2019).

On the other hand, if these were directly translated from the European Italian and French versions, they might differ from the Swiss German version, as illustrated in the following picture (Figure 2.3.):

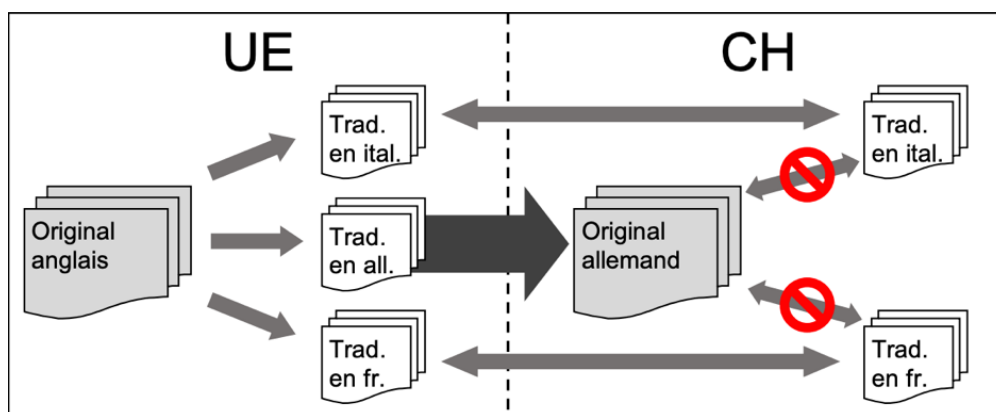


Figure 2.3. The image shows that, even if the procedure of laws implementation changes, and the IT and FR versions are translated directly from their respective EU versions, these might differ from the final German version. (Adapted from Swiss Federal Chancellery, 2019).

The difference between the two implementation procedures is relevant when comparing the CHEU-Lex corpus to Corpus B and must therefore be taken into consideration. Indeed, texts of both corpora are translated versions, but the source languages differ. This implies that some phenomena occurring in corpus B might not occur in CHEU-Lex corpus in the same way. For example, in a translated text it is common to find contact-induced phenomena (section 2.4.) such as calques, i.e. words or phrases borrowed from the source language through a word-per-word translation. Since the source languages are different, calques in the CHEU-Lex corpus might derive from German, whereas calques in Corpus B, if present¹⁶, might derive from English or French.

In *Règles d'or de la reprise du droit de l'UE dans le droit Suisse*, it is clearly stated that, given the complexity of the translation process, great care must be taken when translating legislative texts. Moreover, the Swiss Federal Chancellery (2019) recommends that translators check the European Italian, French, and even English versions as well. Overall, this leads to a procedure where the translation is officially done from a single source language, but, at the same time, several other language versions must be considered in the target language rendition. This characteristic represents the *fil rouge* of the translation in the EU environment.

¹⁶ In chapter 9, Mori shows that calques from English and French are wide more present in Corpus A (2018: 229).

2.4. The CHEU-Lex corpus

In this section, particular attention is devoted to the Swiss corpus CHEU-Lex. It was created as part of a research project led by Professor Annarita Felici¹⁷ with the contribution of a group of students, including the author of this thesis¹⁸, working under the supervision of the project coordinator and Professor Adriano Ferraresi (University of Bologna). Along the lines of the research framework set forth in the *Eurolect Observatory Project* led by Professor Laura Mori¹⁹, the CHEU-Lex corpus aims at providing a richly annotated multilingual resource to investigate the influence of EU law-making and translation practice on the Swiss legislation.

For the following sub-sections regarding the structure and the creation of the CHEU-Lex Corpus, part of the information therein is taken and readapted from the internal documentation shared among the collaborators. Examples are directly copy-pasted from the backup files saved at the end of every phase.

2.4.1. Corpus structure

CHEU-Lex (5,266.714 tokens) is a parallel and comparable corpus of Swiss and European Union legislation. It is divided into three sub-corpora, representing the three official languages of the Swiss Confederation, i.e. French, German and Italian. The German sub-corpus comprises 1,519,412 tokens; the French sub-corpus comprises 1,985,041 tokens; the Italian sub-corpus comprises 1,762,261 tokens. Each corpus contains the official translated versions of the same legislative texts. Overall, the corpus is made up of 444 *agreements* (bilateral agreements entered between Switzerland and the European Union between 1972 and 2017), and 348 *laws* (Swiss federal legislation representing the reception of the agreements). These are equally divided among the three sub-corpora: 148 bilateral agreements and 116 national legal acts, for a total of 264 texts each.

¹⁷ <https://transius.unige.ch/en/research/cheu-lex/>

¹⁸ Antonio Giovanni Contarino (UniBo), Francesco Fericola (UniBo), Silvia Mattiuzzi (UniGe) and Silvia Polito (UniBo).

¹⁹ <https://www.unint.eu/en/research/research-projects/33-page/490-eurolect-observatory-project.html>

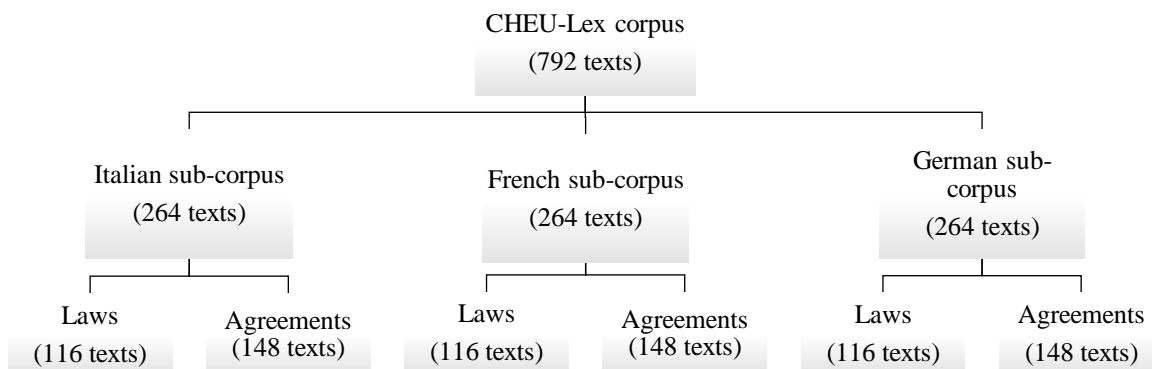


Figure 2.4. The CHEU-Lex corpus structure.

Thanks to its structure, the corpus can be explored from different points of view: monolingual, analyzing laws or agreements in a single language; cross-textual, comparing laws and agreements in the same language; parallel, comparing either laws or agreement in the three languages, etc. (Université de Genève, 2021).

For the purpose of this study, the corpus is analyzed from a monolingual perspective: data were extracted from the *laws* sub-section of the Italian sub-corpus only (see Section 2.1.).

In the following Section, the creation of the corpus is thoroughly described, and the different stages of the process are put forward.

2.4.2. The creation of the corpus

The creation of this corpus began in May 2020 as a collaboration between *Université de Genève* and *Alma Mater Studiorum - Università di Bologna*. Due to the pandemic outbreak which occurred in March of the same year, the work was completely managed remotely.

The creation of the corpus underwent five interconnected phases: text selection and downloading, text cleaning, segmentation, alignment and Part-of-Speech (POS) tagging. Besides the text selection and downloading, carried out by Professor Ferraresi, for each of the remaining phases three out of four students in rotation cooperated to complete the task. During every step, each student had to deal with a sub-corpus, working with one language

only. However, working language could change from step to step, so that each student was supposed to be familiar with all three languages. During every phase it was essential to carry out a job as clean and consistent as possible, since the three sub-corpora would eventually have been merged.

2.4.2.1. Texts selection and downloading

Texts selection and downloading was the first task to be carried out. The task was performed by Professor Adriano Ferraresi by means of a Perl²⁰ script, which downloaded raw texts starting from a list of pre-selected URLs.

The script automatically extracted metadata from the HTML versions and inserted them in the header. These were *id*, *decade_entry*, *date_entry*, *date_signature*, *date_status*, *original_text*, *topic_macro*, *topic_micro* and *type*. The *URL* was added as a reference to the online versions of the texts. The following example shows the metadata structure adopted:

```
<text id="0.192.122.974" decade_entry="1970" date_entry="1 agosto 1971"  
date_signature="24 marzo 1972" date_status="NA" original_text="N"  
topic_macro="0.1 Diritto internazionale pubblico generale" topic_micro="0.19 Relazioni  
diplomatiche e consolari. Missioni speciali. Organizzazioni internazionali. Componento  
dei conflitti. Riconduzione di accordi" type="agreement"  
url="https://www.admin.ch/opc/it/classified-compilation/19720062/index.html">
```

Figure 2.5. An example of the metadata structure.

If one of the metadata was missing, the script automatically added *****NB DATA MISSING***** in place of the metadata missing. During the text cleaning phase, this problem was fixed manually by the students.

Then, the script divided the text into sections according to a standard structure. These sections were defined using a list of XML tags:

²⁰ <https://www.perl.org/>

Tags list	Description
<text>	This tag opens and closes every text. It comprises all the other tags and includes metadata as well.
<title>	It indicates the title of the text.
<title_info>	This tag was created to separate the main title from additional information such as the date of approval of the law etc.
<preamble>	This tag indicates the part of the text introducing the law.
<body>	It is the macro-tag which opens and closes the body of the law. It opens right after the <preamble> tag, if present. It comprises two micro-tags which further divide the text in article title and article text.
<article_title>	This tag is enclosed in the macro-tag <body> and highlights the title of the article.
<article_text>	It usually follows the <article_title> closing tag and refers to the main text.
<annex>	This tag is the macro-tag which marks the annexes, when present. It usually opens after the <body> closing tag. It comprises two more micro-tags which highlight the title of the annex and the text.
<annex_title>	This tag marks the title of the annex.
<annex_text>	It encloses the text of the annex, and it starts right after the <annex_title> closing tag.

Table 2.1. The tags used to divide the texts and their respective meaning.

The following example is taken from one of the agreements of the Italian sub corpus and shows how the tags were used to structure the texts:

```
<text id="0.420.519.121" decade_entry="1990" date_entry="31 luglio 1990"
date_signature="31 luglio 1990" date_status="31 luglio 1990" original_text="Y"
topic_macro="0.4 Scuola – Scienza – Cultura" topic_micro="0.42 Scienza e ricerca"
type="agreement" url="https://www.admin.ch/opc/it/classified-
compilation/19900197/index.html">
<title>
Accordo di cooperazione tra la Confederazione Svizzera e la Comunità economica
europea nel settore della ricerca medica e sanitaria
</title>
<title_info>
Concluso il 31 luglio 1990 Entrato in vigore per scambio di note il 31 luglio 1990
(Stato 31 luglio 1990)
</title_info>
<preamble>
```

La Confederazione Svizzera,
denominata qui di seguito «Svizzera»
e la Comunità economica europea
denominata qui di seguito «Comunità»,
denominate qui di seguito «parti contraenti»,
Considerando che [...]

</preamble>
<body>
<article_title>
Art. 1
</article_title>
<article_text>
Le parti contraenti cooperano per il periodo che va dal 1° gennaio 1988 al 31 dicembre
1991[...]
</article_text>
[...]
<article_title>
Art. 10
</article_title>
<article_text>
Il presente accordo è redatto in duplice copia in lingua tedesca, francese, italiana,
danese, greca, inglese, olandese, portoghese e spagnola, ciascun testo facente
ugualmente fede.
[...]
</article_text>
</body>
<annex>
<annex_title>
Allegato A Temi di ricerca coperti dall'Accordo [...]
</annex_title>
<annex_text>
Obiettivo I.1 – cancro
[...]
</annex_text>
[...]
<annex_title>
Appendice Calendario provvisorio delle spese relative agli obiettivi di ricerca [...]
</annex_title>
<annex_text>
Voce di bilancio 7311 «Ricerca medica e sanitaria» (Stanzamenti d'impegno)
[...]
</annex_text>
</annex>
</text>

Figure 2.6. An example of how tags are used to split texts.

However, due to their structure, some texts such as the *exchange of letters* do not present the whole set of tags. These might only have few of them, such as <text>, <title> and <body>, while the others could be missing. The following example is taken from one of the agreements of the Italian sub-corpus and clarifies what just said:

```

<text [...]>
<title>
Scambio di lettere del 6 novembre 1986 tra la Svizzera e la Comunità europea [...]
</title>
<title_info>
Entrato in vigore il 6 novembre 1986
</title_info>
<body>
<article_text>
Commissione
delle Comunità europee
Bruxelles, 6 novembre 1986
[...]
Paolo Fasella
</article_text>
</body>
</text>

```

Figure 2.7. An example of the structure of the *Exchange of Letters*.

The script was also responsible for removing all HTML tags. Footnotes, as well as their in-text references, were also removed by the script.

Finally, the name of the corpus, i.e. CHEU-Lex, was chosen during this phase. It is made up of CH and EU, which refer to the Swiss Confederation (CH, from the Latin name “Confœderatio Helvetica”) and the European Union (EU).

2.4.2.2. Text cleaning

Text-cleaning was the second phase to be carried out; this operation aimed at obtaining “cleaned” texts starting from the raw versions resulting from the previous phase. It consisted in checking the integrity of texts, as well as the metadata structure and the tags set. This task was performed using text editors such as *Notepad++*²¹ or *BEdit*²² and carried out mostly through *Regular Expressions* (where this was not possible, texts were cleaned “by hand”).

Consistency among the three languages was essential, and it was therefore extremely important to perform the same operations for each language.

The first task was to check the accuracy and the integrity of texts, to be sure that, during the first phase, every part of the text was correctly included in the .txt file by the script. When missing parts were found, these were copy-pasted to the target file.

²¹ <https://notepad-plus-plus.org/>

²² <https://www.barebones.com/products/bbedit/>

Secondly, it was necessary to check the metadata set. As already explained in Section 2.3.2.1, it was not uncommon to find *****NB DATA MISSING***** instead of the appropriate metadata; in this case, information was sourced, when possible, from the webpages and/or the PDF version of the laws. If it was not retrievable, the value “NA” was entered.

Besides metadata and missing text parts, structural tags were another important part to be checked, in order to be sure that each part of the text was correctly tagged according to its function. For instance, some problems occurred when dealing with titles: sometimes these were numbered (usually when belonging to a list), and numbers were divided from the rest of the title, as in the following example:

```
<article_title> Art. 1 </article_title>

<article_title> I. Uso dello spazio aereo svizzero / 1. Principio e
definizioni </article_title>
```

Figure 2.8. An example of numbered titles.

Although this does not represent a major problem, it was conventionally decided to keep the number and the title together, joining them as shown in the following example:

```
<article_title> Art. I. Uso dello spazio aereo svizzero / 1. Principio
e definizioni </article_title>
```

Figure 2.9. The titles merged with their numbers.

Tables also posed a big problem, and it would be necessary to write a whole paragraph concerning this issue: no XML tag was used to highlight the presence of these elements, and these were therefore merged with the body of text by the Perl script. Moreover, several tables only included numbers and names of items, bearing no syntactic function, and being therefore almost useless for linguistic research purposes. This posed a big problem especially during the alignment phase, but this will be analyzed later.

Images were not downloaded by the script, and therefore the abbreviation [IMG] was inserted instead of the real image. However, some inconsistencies occurred among the three language versions, since some documents might contain images, whereas their corresponding versions in the other languages contained the text of the image, and not a .jpg file.: this situation led to pieces of texts being present in one language version and replaced by [IMG] in its translation(s). An example to this is *Law 017*, whose French version displays an image:

–  **Art. 16 Droits de l'assuré dans le système de la primauté des prestations**

¹ Dans les institutions de prévoyance appliquant le système de la primauté des prestations, les droits de l'assuré correspondent à la valeur actuelle des prestations acquises.

² Les prestations acquises sont calculées comme suit:

$$\text{prestations assurées} \times \frac{\text{période d'assurance imputable}}{\text{période d'assurance possible}}$$

³ Les prestations assurées sont fixées par le règlement. Elles sont déterminées par la période d'assurance possible. Les prestations temporaires au sens de l'art. 17, al. 2, peuvent être omises lors du calcul de la valeur actuelle, si elles ne sont pas financées selon le système de capitalisation.

Figure 2.10: The image is taken from the French version of Art. 16 of Law 017²³ and shows the case above mentioned.

Unlike the French version, the Italian and German versions do not have the image, but, instead, the written text is inserted in a table:

–  **Art. 16 Diritti dell'assicurato nel sistema del primato delle prestazioni**

¹ Negli istituti di previdenza gestiti secondo il sistema del primato delle prestazioni, i diritti dell'assicurato corrispondono al valore attuale delle prestazioni acquisite.

² Le prestazioni acquisite sono calcolate come segue:

prestazioni assicurate x	periodo d'assicurazione computabile

	periodo d'assicurazione possibile

³ Le prestazioni assicurate sono fissate dal regolamento. Sono determinate per il periodo d'assicurazione possibile. Le prestazioni temporanee giusta l'articolo 17 capoverso 2 possono essere trascurate nel calcolo del valore attuale, se non sono finanziate secondo il sistema di capitalizzazione.

Figure 2.11: The image is taken from the Italian version of Art. 16 of Law 017²⁴ and shows the case above mentioned.

–  **Art. 16 Ansprüche im Leistungsprimat**

¹ Bei Vorsorgeeinrichtungen im Leistungsprimat entsprechen die Ansprüche der Versicherten dem Barwert der erworbenen Leistungen.

² Die erworbenen Leistungen werden wie folgt berechnet:

versicherte Leistungen x	anrechenbare Versicherungsdauer
	mögliche Versicherungsdauer

³ Die versicherten Leistungen sind im Reglement niedergelegt. Sie bestimmen sich aufgrund der möglichen Versicherungsdauer. Temporäre Leistungen gemäss Artikel 17 Absatz 2 können bei der Barwertbestimmung weggelassen werden, wenn sie nicht nach dem Deckungskapitalverfahren finanziert werden.

Figure 2.12: The image is taken from the French version of Art. 16 of Law 017²⁵ and shows the case above mentioned.

²³ Available online at https://www.fedlex.admin.ch/eli/cc/1994/2386_2386_2386/fr

²⁴ Available online at https://www.fedlex.admin.ch/eli/cc/1994/2386_2386_2386/it

²⁵ Available online at https://www.fedlex.admin.ch/eli/cc/1994/2386_2386_2386/de

This dissimilarity caused a difference in the display of the three texts on SketchEngine, as shown below:

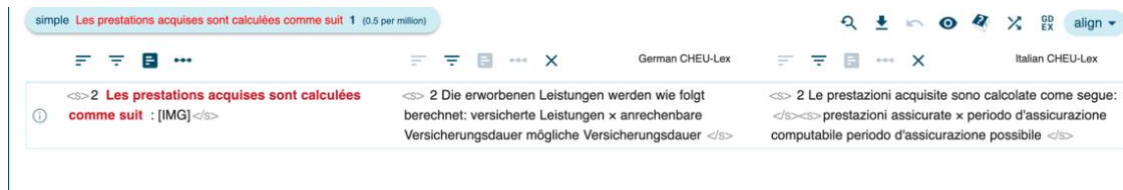


Figure 2.13: The image is a screenshot from the SketchEngine²⁶ interface: after choosing the French sub-corpus, it is necessary to perform the advanced “Parallel Concordance” and look for the sentence “Les prestations acquises sont calculées comme suit” and look for the parallel concordances both in German and Italian. French occurrence displays “[IMG]”, where the other versions have the written text.

As already mentioned in Section 2.3.2.1, footnotes and their in-text references were removed by the script. This did not pose a problem in general, except when the text contained a superscript referring to a deleted footnote, and sometimes followed by ellipses. These cases sometimes posed problems, as shown in the following example:

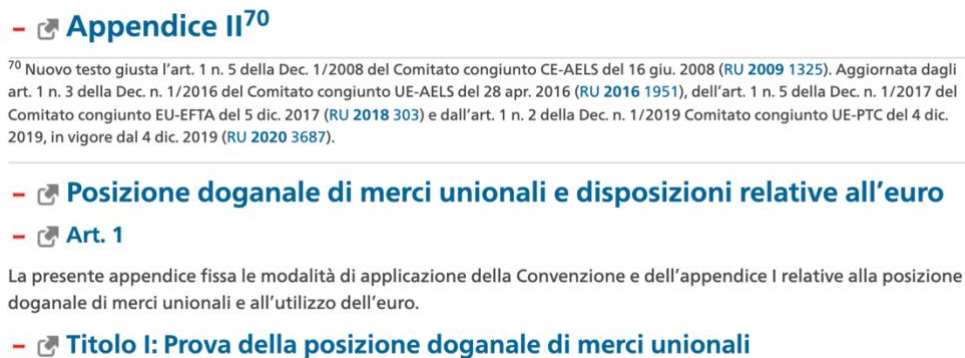


Figure 2.14: The title “Appendice II” contains a superscript referring to a footnote which has been deleted by the script in the downloading phase.

In such cases, both the superscript and the footnote were deleted from the script, and the first raw version of the text displayed an empty space, as shown below:

```
<annex_title>
Appendice II
</annex_title>
<annex_text>

</annex_text>
<annex_title>
Appendice II
```

²⁶ <http://corpora.fti.unige.ch/crystal/#open>


```
Posizione doganale di merci unionali e disposizioni relative all'euro [...]  
</annex_title>
```

Figure 2.15: How the text in figure 14 is displayed after being downloaded from the script.

Therefore, it has been necessary to add the Latin form [OMISSIS], as shown in Figure 2.16:

```
<annex_title>  
Appendice II  
</annex_title>  
<annex_text>  
[OMISSIS]  
</annex_text>  
<annex_title>  
Appendice II  
Posizione doganale di merci unionali e disposizioni relative all'euro [...]  
</annex_title>
```

Figure 2.16: The use of the Latin form [OMISSIS].

Eventually, since texts were converted into XML files, in order to avoid parsing errors XML special characters (&, < and >) were substituted with their escaped characters (&, < and > respectively), i.e. characters which invoke an alternative interpretation of the main character. This happens because in the XML markup language, special characters bear a different meaning than their ordinary one: the “<” symbol, for instance, usually means “less than”, but in the XML language it indicates the opening of a tag. If, however, the text which is being tagged contains the symbol “<” to indicate that something is less than something else, it is necessary to use the escaped character “<” instead of the symbol.

2.4.2.3. Segmentation

Text segmentation aimed at splitting texts into consistent segments (sentences) and was carried out through *InterText Editor*²⁷, a parallel text alignment editor. This application is mainly conceived to align two or more parallel texts at the level of sentences, which means that sentences of parallel texts (i.e. translated versions of the same text) are paired to their respective translation²⁸. To perform this operation, *InterText* automatically pre-splits texts into sentences and delimits them using the <s> tags. Then, the application automatically pre-aligns any given pair of texts according to its XML tags (if provided by the user) and the integrated splitter tool automatically segments the text. To identify these segments, the

²⁷ <https://wanthalf.saga.cz/intertext>

²⁸ Note that this will be the next step, i.e. the alignment phase (section 2.3.2.4.).

program follows a list of pre-set criteria, such as the presence of a strong punctuation mark, which, however, are not fixed. Also, segment tags were required to have an id metadata (the outcome is supposed to be as follows: `<s id="2"> [...] </s>`), which means that the *single level numbering* option was selected: this resulted in the opening tag of a segment also including the number of the sentence to which it refers.

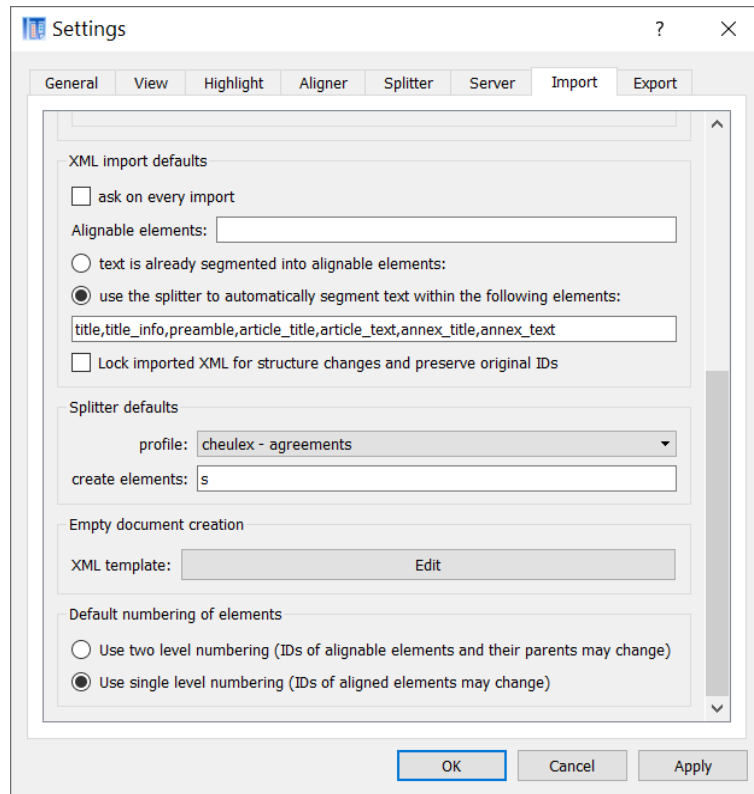


Figure 2.17: The importing option pane of *InterText*. It is possible to choose the XML import defaults as well as the default numbering of elements.

During the segmentation phase, texts were segmented using the sentence splitter tool available on *InterText*. However, since sentences were eventually aligned during the following phase (see Section 2.3.2.4.), they had to be as coherent and consistent as possible in all three languages. Therefore, it was necessary to conventionally define the sentence as “[...] a sequence of words bearing a syntactic function and generally starting with a capital letter and ending with a strong punctuation mark.” (This definition is taken from the internal documentation concerning the segmentation phase). Also, due to the legal character of the texts, as well as the three languages in which they are written, it was necessary to apply some exceptions to this definition.²⁹ To carry out this operation it was necessary to set a series of

²⁹ The grammar of each language was taken into account. Take for instance German: every noun is written with a capital letter, so *InterText* might think it is the beginning of a new sentence. To avoid this as well as other problems related to the single languages, for each language a personalized set of splitting rules was conceived.

segmentation rules which differed from the pre-set list, called *replacement rules*. These were written using Perl-compatible *Regular Expressions* and were tailored for the purposes. These criteria included, among others, splitting the sentences in the presence of a strong punctuation mark such as full stop or exclamation point, as well as at a breakup line. Abbreviations (usually ending with a full stop) represented a major problem throughout all the texts, and a list of exceptions (therefore a list of the most common abbreviations) was created for each language in order to avoid segmentation when an abbreviation appeared. The different structure of the two types of documents (Laws and Agreements) was another problem and made it necessary to establish a set of segmentation rules for each. These sets of rules were almost the same for every language.

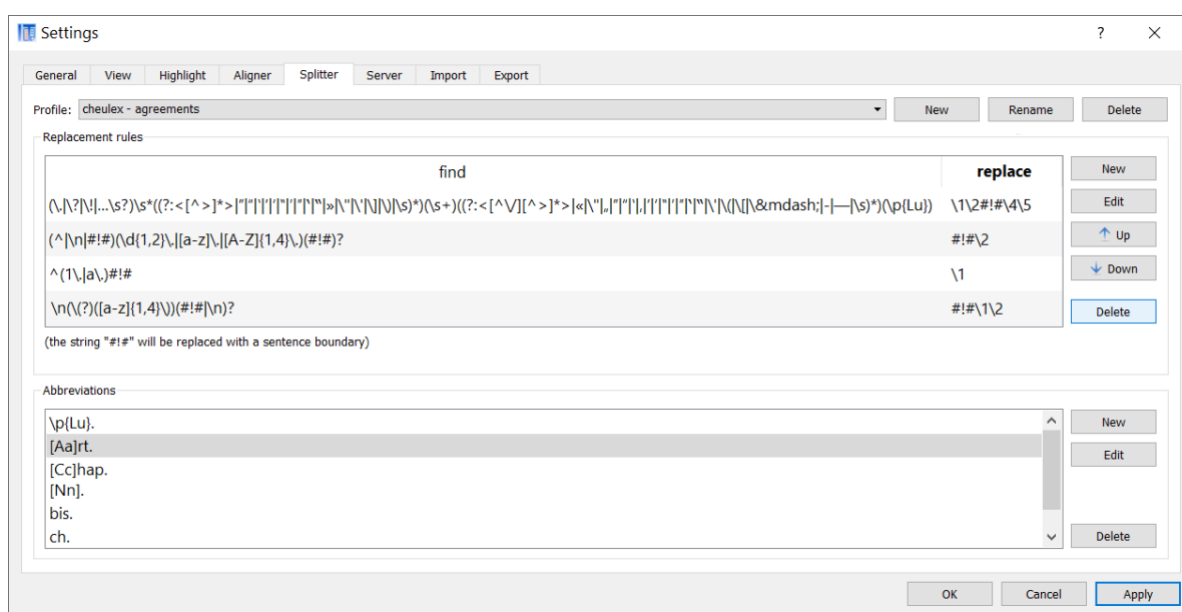


Figure 2.18: The sentence splitter pane of *InterText*. Replacement rules were written using Perl-compatible Regular Expressions. Below, the list of the abbreviations.

During this phase, every student worked with a language only. However, *InterTexts* requires the upload of two parallel texts to work³⁰. Therefore, each student had a language pair to upload on *InterText* (DE-FR, FR-IT and IT-DE), but he or she had to perform segmentation on the assigned language only. For instance, suppose a student was assigned with Italian, he or she had to upload the IT-DE pair and work on the IT text only. The choice of the language pair was made taking into consideration the languages known by the students and avoiding repetitions (no language pair was left out). Nonetheless, having for instance the Italian text on one side of the screen and the German on the other helped to highlight in a more effective

³⁰ Remember that *InterText* is conceived to align parallel texts, and therefore it needs more than one text to function.

way any sort of discrepancy between any two language versions, and any problematic point was written down and discussed with the other colleagues on a shared excel file³¹.

As previously mentioned, some exceptions were necessary: for instance, due to the structure of “Preambles” and “Final Acts”, breaks between sentences with an almost complete meaning were marked by a comma or a semi-column:

-  **Preambolo**

La Confederazione Svizzera,
da un lato,

La Comunità Economica Europea,
dall'altro,

considerando gli stretti rapporti esistenti fra la Confederazione Svizzera e la Comunità;

desiderose di consolidare, in occasione della creazione di un mercato unificato in materia di assicurazione nell'ambito della Comunità, le relazioni economiche esistenti nel settore fra le due Parti e di promuovere, nell'osservanza di eque condizioni di concorrenza, lo sviluppo armonioso di tali rapporti, garantendo la protezione degli assicurati;

risolute ad eliminare a tal fine, su una base di reciprocità e di non discriminazione e con la garanzia delle condizioni giuridiche necessarie in materia di vigilanza, gli ostacoli all'accesso all'attività e all'esercizio dell'assicurazione diretta diversa dall'assicurazione sulla vita e risolte quindi ad introdurre nei reciproci rapporti la libertà di stabilimento in materia;

sottolineando che ciò non pregiudica in alcun modo il loro potere di legiferare nei limiti stabiliti dal diritto internazionale pubblico;

adoperandosi per quanto possibile affinché i rispettivi ordinamenti giuridici interni in materia evolvano in modo reciprocamente compatibile;

constatando che nell'interesse delle rispettive economie occorre sviluppare e approfondire in tal senso i loro rapporti in un settore finora non soggetto a una disciplina convenzionale e contribuire in tal modo al coordinamento del diritto dell'economia fra le due Parti;

dichiarandosi disposte ad esaminare, in base a qualsiasi elemento di valutazione e, in particolare, all'evoluzione del diritto comunitario delle assicurazioni, la possibilità di concludere altri accordi nel settore delle assicurazioni private;

hanno convenuto, per raggiungere i suddetti scopi, di concludere il presente Accordo e a tal fine hanno designato come plenipotenziari:

Figure 2.19: The first part of the Preamble of agreement 035³²: sentences with almost a complete meaning are marked by the use of a semi-column.

Here, the decision was taken to avoid segmentation and keep the sentences together, resulting in a lower level of granularity of segmentation of these elements, when compared to others. Also, some titles and subtitles had to be kept altogether, even if they were made up of two or more sentences. If this was the case, they were manually merged.

Finally, another problem was posed by some of the elements present in tables: due to their content being almost useless to the purpose of the project, the decision was made to leave

³¹ This also helped to keep track of every correction made.

³² Available online at https://www.fedlex.admin.ch/eli/cc/1992/1894_1894_1894/fr

the automatic segmentation unchanged. Only minor corrections were made, such as splitting table data from headings, text body or any other meaningful part of the text.

Together with the automatic segmentation, manual segmentation was carried out when necessary, to ensure the best outcome in terms of consistency. However, despite the thorough approach adopted, some minor errors persisted. For instance, as already mentioned, segmentation discrepancies can be found in tables content, due to the high recurrence of these elements and the variety of their content. Also, suspension marks "..." occurring in titles and lists to signal an abrogation of that part of the document were not always printed by the script, were not always restored during the segmentation phase.

Here is an example of the segmentation file structure taken from the French corpus:

```
<title_info>
  <s id="2">(Développement de l'acquis Schengen)
    Entré en vigueur le 28 mars 2008
    (Etat le 28 mars 2008)</s>
</title_info>

<article_text>
  <s id="3">Commission des Communautés européennes
    Bruxelles, le 5 février 1981
    S. E. M. Pierre Cuenoud
    L'Ambassadeur de Suisse
    auprès des Communautés européennes</s>

  <s id="4">Monsieur l'Ambassadeur,
    J'ai l'honneur d'accuser réception de votre lettre dont la teneur est la
    suivante:</s>

    ...

  <s id="21">Au nom du Conseil des Communautés européennes:
    David Hannay
    G. Giola</s>
</article_text>
```

Figure 2.20: An example of the segmentation file structure.

Eventually, since during the next phases texts were merged into a single document for each language, the last step of the segmentation phase was to change the numbering of the id metadata for each segment. This operation was performed by means of a Python³³ script. A naming convention was adopted: if the sentence belonged to the agreements, the id was <s

³³ <https://www.python.org/>

`id="agrXXX_LL_N">`. If the sentence belonged to the laws, the id was `<s id="ordXXX_LL_N">`. Here, “XXX” corresponds to the number of the document, “LL” corresponds to the language version (de, fr or it) and “N” corresponds to the number of the original segment id. The following are examples are taken from the merged Italian file:

`<s id="ord040_it_1">Ordinanza sull'ammissione al Politecnico federale di Losanna</s>`

`<s id="agr041_it_1">Accordo tra la Confederazione Svizzera e la Comunità europea sul commercio di prodotti agricoli</s>`

2.4.2.4. Alignment

Segment alignment was carried out by means of *Intertext Editor* (see Section 2.3.2.3.). As stated in the previous Section, the program automatically split texts and aligned segments of the two files. A manual evaluation and correction of each text for every single pair of languages (IT-FR, IT-DE and DE-FR) was performed, following the *SketchEngine* guidelines³⁴ for m:n alignment, where “m” and “n” stands for variables. This means that multiple segments in one language are aligned to multiple segments in the other, so it would be possible, for instance, to have a 3:2 alignment (three sentences in a language version are paired with two sentences in another language version). During the previous phases of this project, texts were treated individually. Yet, during this and the next phases, a single document for each language was deemed more appropriate. Texts in the same language were, therefore, merged all together in three main files.

Despite major errors being corrected in the previous phases, some persisted and were corrected manually. For instance, during the translation process, it is common to invert sentence order. This led to a scenario in which the apparently missing segment was actually included in the previous sentences.

³⁴ <https://www.sketchengine.eu/guide/setting-up-parallel-corpora/#tab-id-3>

▶ 8 In Zusammenarbeit mit den Kantonen ordnet der Bundesrat schweizweit Betriebsvergleiche zwischen Spitälern an, insbesondere zu Kosten und medizinischer Ergebnisqualität.	▶ 8 En collaboration avec les cantons, le Conseil fédéral fait procéder à l'échelle nationale à des comparaisons entre hôpitaux – qu'il publie par la suite – en ce qui concerne notamment les coûts et la qualité des résultats médicaux.
▶ Die Spitäler und die Kantone müssen dafür die nötigen Unterlagen liefern.	▶ Les hôpitaux et les cantons doivent livrer les documents requis à cette fin.
▶ Der Bundesrat veröffentlicht die Betriebsvergleiche.	

Figure 2.21: The *InterText* interface: the 1:0 alignment is an error because the last segment is included in the French translation. For such cases, the m:n alignments were used by grouping various segments together. This particular example has been converted into a 3:2 alignment.

Another problem was related to tables. Given the considerable length of some of them, the decision was made to align them in a way that would not affect the research. This resulted in two possible solutions: either aligning line per line or, when this was not possible, merging the whole content of the table in a single segment. In particular, some tables presented terminology listed in alphabetical order; these lists were rather problematic to align, considering that the order of items changes from language to language.

- 📄 1. Professioni sanitarie

Denominazione			Osservazioni
Assistente di cura per animali	Tierheilpraktiker/in	Assistant en soins pour animaux	
Assistente podologo	Podologieassistent/in	Assistant podologue	
Audioprotesista	Hörgeräte-Akustiker/in	Audioprothésiste	

- 📄 1. Professions de la santé

Dénomination			Remarques
Ambulancier	Rettungssanitäter/in	Soccorritore	Y compris direction d'une entreprise d'ambulance et technicien ambulancier.
Assistant en soins et santé communautaire	Fachmann/frau Gesundheit	Operatore sociosanitario	
Assistant en soins pour animaux	Tierheilpraktiker/in	Assistente di cura per animali	

- 📄 1. Gesundheitswesen

Bezeichnung			Bemerkung
Alternativmedizin und Komplementärtherapien*	Médecine alternative et thérapies complémentaires*	Medicina alternativa e terapie complementari*	Betrifft insbesondere: Akupunktur, Atemtherapie, Ayurveda, Bioresonanz, Kunsttherapie, Heileurythmie, Homöopathie, Reflexologie, Shiatsu, Traditionelle Chinesische Medizin (TCM), Traditionelle Europäische Naturheilkunde (TEN). Siehe die Liste: www.emr.ch .
Apotheker/in	Pharmacien	Farmacista	

Figure 2.22: The example compares the first elements of a list in three languages, taken from the official website of the Swiss confederation.³⁵ The list is in alphabetical order, and it changes according to the language.

Since some of these lists were relatively short, their content was merged altogether in a single segment. However, this solution was not ideal for longer ones, as it would have resulted in extremely long segments. Therefore, if it was the case, elements were aligned line per line.

Here follows an example of the file structure for the IT-FR pair, taken from the `fr-it_aligned.xml` file: every line corresponds to the alignment of two or more sentences; first, “link type = ‘1-1’” indicates that the alignment is 1:1, so one sentence in Italian is aligned to one sentence in French; the IDs of the two sentences follow, i.e. ‘agr000_it_1’ and

³⁵ <https://www.fedlex.admin.ch/eli/cc/2013/510/it> Please note that examples are taken from the old version dated 01.10.2019, as it is the version used in the creation of the corpus.

‘agr000_fr_1’. Eventually, “status = ‘man’” indicated whether the alignment was made manually or automatically (in this case ‘auto’ was inserted instead of ‘man’).

```
<?xml version='1.0' encoding='utf-8'?>

<linkGrp toDoc='cheu-lex_it.xml' fromDoc='cheu-lex_fr.xml'>

<link type='1-1' xtargets='agr000_it_1;agr000_fr_1' status='man'/>

<link type='1-1' xtargets='agr000_it_2;agr000_fr_2' status='man'/>

<link type='1-1' xtargets='agr000_it_3;agr000_fr_3' status='man'/>

<link type='1-1' xtargets='agr000_it_4;agr000_fr_4' status='man'/>

<link type='1-1' xtargets='agr000_it_5;agr000_fr_5' status='man'/>

<link type='1-1' xtargets='agr000_it_6;agr000_fr_6' status='man'/>

<link type='1-1' xtargets='agr000_it_7;agr000_fr_7' status='man'/>

<link type='1-1' xtargets='agr000_it_8;agr000_fr_8' status='man'/>

<link type='1-1' xtargets='agr000_it_9;agr000_fr_9' status='man'/>
```

Figure 2.23: The structure of the aligned IT-FR file.

2.4.2.5. POS tagging and dependency parsing

In this last phase, the Part-Of-Speech tagging (POS tagging) and lemmatization were carried out for each sub-corpus. POS tagging consists in marking up every single word in a corpus as corresponding to a particular part of speech, or grammatical category, usually depending on its definition and/or its context. This is done by means of algorithms that associate terms with a set of descriptive tags. Lemmatization is the algorithmic process that associates the correct *lemma* to every inflected form of a word. *Lemma* is, by definition, the basic form of a word which is typically found in dictionaries. So, for instance, the lemma of the word “kids” is “kid”. For the Italian and German sub-corpora, these two operations were carried out on *SketchEngine*, using the taggers available for each language, which are the Italian Treetagger (Marco Baroni tagset)³⁶ and the German RFTagger³⁷, respectively. A different

³⁶ <https://www.sketchengine.eu/italian-treetagger-part-of-speech-tagset/>

³⁷ <https://www.sketchengine.eu/german-rftagger-part-of-speech-tagset/>

operation was carried out for the French corpus, for which the *French TreeTagger*³⁸ was used.

Tag sets were also modified to adapt to some specific features of the analyzed genre, and this was done through Regular Expressions. In particular, the “*LI*” tag (list tag) was added, to tag alphanumeric list markers such as “1.”, “a.”, “1a.” etc., which often occur at the beginning of titles, paragraphs and lists both in Laws and Agreements. Also, the tag ABR (abbreviation) was added both in the Italian and German corpus, whereas the FW tag was added in the French corpus, to indicate the presence of a foreign word. The inclusion of the Latin form [OMISSIS] during the text-cleaning phase made it necessary to add a specific tag for it, named OMISSIS.

Further problems were created by some unmodifiable settings, such as the annotation of the corpus with <s> to mark sentence boundaries, and the addition of <g/> tag, called *glue tag*; the first tags were successfully removed using Regular Expressions, whereas the latter, since they were only present in the Italian and German sub-corpora, were eventually added in the French corpus by means of Regular Expressions.

The final format of the POS-tagged and lemmatized corpus is a word-per-line (WPL) .vert file, which is the format required by *SketchEngine* and *NoSketchEngine*. Each token is on a single line and is separated from its structural metadata tag by a tab; this structure divides information into columns; the first one contains tokens and structures, the second one contains the part-of-speech tag and the third one contains the *lempos*, which consists of the lemma followed by a hyphen and a one-letter abbreviation referring to the part of speech. The following example clarifies the description above:

³⁸ <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

```

<title>
<s id="agr000_de_1">
Vereinbarung      N.Reg.Nom.Sg.Fem      Vereinbarung-n
Über              APPR                   über-i
die               ART.Def.Acc.Sg.Fem    die-x
rechtliche        ADJA.Pos.Acc.Sg.Fem   rechtlich-j
Stellung          N.Reg.Acc.Sg.Fem      Stellung-n
der               ART.Def.Gen.Sg.Fem    die-x
Europäischen      ADJA.Pos.Gen.Sg.Fem   europäisch-j
Investitionsbank  N.Reg.Gen.Sg.Fem      Investitionsbank-n
in                APPR                   in-i
der               ART.Def.Dat.Sg.Fem    die-x
Schweiz           N.Name.Dat.Sg.Fem     Schweiz-n
</s>
</title>

```

Figure 2.24: An example of the WPL.vert file.

Some errors were found in all sub-corpora, and therefore each .vert file was checked and corrected, as far as possible, using Regular Expressions. Some problems were specifically related to each language version, while others were shared by the three corpora almost to the same extent. In order to track down the most obvious errors, POS tag-based research was carried out on *SketchEngine*, and the results were then sorted by frequency.

First, alphanumeric list markers were tagged using the “LI” tag, trying to be as consistent as possible throughout all the three corpus language versions. Every possible exception (i.e., capital letters followed by a full stop which are not list markers) had to be avoided. For instance, “N.” could either be a list marker or the abbreviation for “number”. On the other hand, wrongly annotated acronyms and abbreviations were corrected as well, assigning the tag for “proper noun” (which is different for every tagset) for cases such as “EU”, “EIONET” etc. The tag “ABR” was annotated, and in this specific case, abbreviations were merged with their following full stop when wrongly separated by the tagger. More in general, a lot of corrections were carried out, both by means of Regular Expressions or by hand, concerning wrongly annotated words, symbols, uppercase stand-alone letters (widely present in tables), roman numbers and punctuation marks. For what concerns foreign words, these were correctly tagged only by the German Tree Tagger with “FW”, while the other taggers did not recognize them and usually inserted the “<UNKNOWN>” element. These were corrected only in the French version, as they were easier to detect, whereas in the Italian corpus they were left unchanged due to their large number and variety.

Here is a sample taken from the Italian .vert file:³⁹

```
<section name="Title">
<s id="agr000_it_1">
Accordo      NOUN      accordo-n
sullo        ARTPRE    sul-x
statuto      NOUN      statuto-n
giuridico    ADJ       giuridico-j
in           PRE       in-i
Svizzera     NPR       Svizzera-n
della        ARTPRE    della-x
Banca        NPR       Banca-n
Europea      NPR       Europea-n
</g/>
```

Figure 2.25: An example of the .vert file.

And here is an example of the use of the LI tag:

```
</g/>
approvazione NOUN      approvazione-n
della        ARTPRE    della-x
Banca        NPR       Banca-n
nazionale    ADJ       nazionale-j
svizzera     NOUN      svizzera-n
1)           LI        [list]-x
prima        ADV       prima-r
di           PRE       di-i
emettere     VER:infi  emettere-v
un           ART       un-x
prestito     NOUN      prestito-n
sul          ARTPRE    sul-x
mercato      NOUN      mercato-n
svizzero     ADJ       svizzero-j
</g/>
```

Figure 2.26: An example of the use of the LI tag.

The last operation carried out was dependency parsing. It is the process of analyzing the syntactic structure of a sentence based on the words which constitute that sentence. This operation was carried out by means of the *Spacy*⁴⁰ library. The file structure resembles that of the .vert file, therefore having each token on a single line and each annotation level divided by a single *tab* character. There are 8 columns in total, and each column has a different element: *TEXT*, which is the original word text; *LEMMA*, the base form of the token; *POS*, which is the simple part-of-speech tag; *TAG*, i.e. the detailed POS tag; *DEP*, holding the syntactic dependency, therefore the relation between tokens; *HEAD TEXT*,

³⁹ Here the *tab* format has been adapted to clearly show the structure of the file.

⁴⁰ <https://spacy.io/>

which is the original text of the token head; *HEAD POS*, referring to the POS tag of the token head and *CHILDREN*, which holds the immediate syntactic dependents of the tokens. This structure is retrieved from the Spacy documentation.⁴¹ In this phase, no manual correction was performed. Thus, it is highly likely to find a lot of classification mistakes, especially because the models used were not trained on legal texts. Here follows an example from the Italian corpus:

⁴¹ <https://spacy.io/usage/linguistic-features>

<u>TEXT</u>	<u>LEMMA</u>	<u>POS</u>	<u>TAG</u>	<u>DEP</u>	<u>HEAD TEXT</u>	<u>HEAD POS</u>	<u>CHILDREN</u>
Il	Il	DET	RD__Definite=Def Gender=Masc Number=Sing PronType=Art	det	testo	NOUN	-
testo	testare	NOUN	S__Gender=Masc Number=Sing	nsubj:pass	redatto	VERB	Il
dell'	dell'	DET	DD__Number=Sing PronType=Dem	det	accordo	NOUN	-
accordo	accordare	NOUN	S__Gender=Masc Number=Sing	nsubj:pass	redatto	VERB	dell'
è	essere	AUX	VA__Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	aux:pass	redatto	VERB	-
redatto	redatto	VERB	V__Gender=Masc Number=Sing Tense=Past VerbForm=Part	ROOT	redatto	VERB	testo, accordo, è, greco, fa, .
in	in	ADP	E__	case	greco	NOUN	-
greco	greco	NOUN	S__Gender=Masc Number=Sing	obl	redatto	VERB	in
e	e	CONJ	CC__	cc	fa	VERB	-
fa	fare	VERB	V__Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	conj	redatto	VERB	e, fede, testi
fede	fede	NOUN	S__Gender=Fem Number=Sing	obj	fa	VERB	pari
al	al	ADP	E__	case	pari	ADJ	-
pari	pario	ADJ	A__	amod	fede	NOUN	al
dei	dio	DET	DI__Gender=Masc Number=Plur PronType=Ind	det	testi	NOUN	-
testi	testo	NOUN	S__Gender=Masc Number=Plur	obj	fa	VERB	dei, originali
originali	originale	ADJ	A__Number=Plur	amod	testi	NOUN	-
.	.	PUNCT	FS__	punct	redatto	VERB	-

Table 2.2: An example of the structure of the .vert file obtained during this phase. It is possible to observe 8 columns referring to the 8 different elements: *text*, *lemma*, *POS*, *tag*, *dep*, *head text*, *head pos* and *children*. For every *token*, i.e. the smallest unit that a corpus consists of, all the categories above are described.

2.4. Methodology

This Section introduces the methodology adopted in the case study which makes the object of this thesis. Since the methodology strictly follows that adopted by Mori (2018c), for every Section, a quick overview of the methodology and a short explanation of the result obtained in the frame of the *Eurolect Observatory Project* is set forth.

During the *Eurolect Observatory Project*, a common reference template (section 2.4.1.) was conceived by Mori to answer the research question whether two different legislative varieties (the Eurolect and the domestic variety) do exist for at least some the official languages of the European Union. Keeping in mind that “the term *Eurolect* is used to refer to those cases in which language variation is found to affect several linguistic levels” (Mori and Szmrecsanyi, 2021), the common reference template aimed at carrying out an intra-lingual comparison at various levels, i.e., lexical, morphological, morphosyntactic, syntactic and textual. Three heuristic macro-areas of research were chosen: *EU Rooted-phenomena*, *Contact-induced features* and *Intra-linguistic variability* (section 1.5.3.).

This thesis aims at analyzing whether the impact of the European legal drafting process can be detected on the Swiss legal Italian (section 2.1.). To tackle this research, linguistic data extracted from the CHEU-Lex corpus (representing the Swiss variety of legal Italian used to implement the bilateral agreements) were compared to those obtained from Corpus B (representing the domestic Italian of implementation laws) (Mori, 2018c). Eventually, Corpus A (Mori, 2018c) was used as a *tertium comparationis*, as it could provide further evidence to support the previous results: indeed, to be sure that results obtained from the CHEU-Lex Corpus can be interpreted as the results of an eventual influence of the European legal drafting process on the Swiss Italian, it is necessary to compare them to an example of European language. Corpus A collects Italian legal texts (written within the EU environment (EU directives), thus being therefore an optimal candidate for this role.

Given that data obtained from the CHEU-Lex corpus had to be directly comparable to those of Corpus B, it was necessary to set up a research method which closely followed that of Mori (2018c). This means that the procedure followed in the framework of the *Eurolect Observatory Project*, was accurately replicated on the CHEU-Lex corpus, to make the data extracted as comparable as possible. In this sense, the research template conceived by Mori (section 2.4.1.) represented a precious guide to follow.

Nonetheless, some inconsistencies necessarily persisted between the two methods, and they are worth mentioning. First, two different programs were used: Wordsmith⁴² was used for the *Eurolect Observatory Project*, whereas SketchEngine⁴³ was used for the CHEU-Lex corpus. The whole CHEU-Lex corpus is, indeed, available for free via the *NoSketchEngine* interface⁴⁴, which is the open-source, limited version of *SketchEngine*. However, a set of tools and operations were not available on this interface. Therefore, it was necessary to upload the original file of the Italian CHEU-Lex corpus on the personal SketchEngine account of the author of this thesis, to take advantage of all the tools available.

It is worth remembering that, as already explained in Section 2.1., the research had to be conducted exclusively on the *laws* sub-section. This did not represent a limit, since SketchEngine allows to restrict the research to a specific sub-section (i.e. selecting *laws* texts only). However, Mori states that “In this study the focus will be primarily on results extracted from enacting parts.” (2018c: 205), which means that the research had to be further narrowed down to the *enacting parts* of the *Laws* sub-corpus. This new subdivision posed a problem: to correctly identify the *enacting parts* sub-section.⁴⁵

To properly identify the *enacting parts* sub-section, it is essential to underline an important difference between the structure of the CHEU-Lex corpus and that of Corpus B. Both corpora were divided into different functional sections through the use of XML-like markup (see Section 2.3.2.1.), but since names are chosen in an arbitrary manner, these do not always match. Sections of Corpus B are named *Title*, *Preamble*, *Disposition* and *Annex* (Tomatis, 2018: 30), whereas the CHEU-Lex internal structure includes *Title*, *Title Info*, *Preamble*, *Body*, *Article title*, *Article text*, *Annex*, *Annex Title* and *Annex Text* (Section 2.3.2.1.).

The *Standard Structure* published in the *Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of the European Union legislation* (European Union, 2015: 24) states that “All acts of general application shall be

⁴² <https://www.lexically.net/wordsmith/>

⁴³ <https://www.sketchengine.eu/>

⁴⁴ <http://corpora.fti.unige.ch/crystal/#open>

⁴⁵ Another, minor problem was that SketchEngine did not allow to select more than one sub-section at time (i.e. *Laws* and the section corresponding to the *enacting parts*). To solve this problem, a smaller corpus, made up of Italian laws only, was created. The new corpus was then uploaded on SketchEngine. Since the new corpus contained *Laws* only, it was not necessary to select “Laws” as a sub-section, and “Body” was the only sub-section to be selected. Thus, the research was performed on the *enacting parts* of the *Laws* of the Italian CHEU-Lex corpus.

drafted according to a standard structure (Title - Preamble - Enacting Terms – Annexes, where necessary)”. It also states that:

“The ‘enacting terms’ are the legislative part of the act. They are composed of articles, which may be grouped into parts, titles, chapters and Sections, and may be accompanied by annexes.” (European Union, 2015: 24)

In view of the above, the *enacting parts* correspond to the *Dispositions* sections in Corpus B, and to the *Body* sections in the CHEU-Lex Corpus.

As already made explicit in Section 2.1., in this thesis the name “CHEU-Lex corpus” refers to the corpus containing Italian *laws* only, and not to the whole CHEU-Lex corpus, unless otherwise specified.

Another problem is that Wordsmith and SketchEngine share a set of tools which are similar in nature, but not completely identical; in particular, the standardized Type/Token ratio (section 2.4.2.) is available on Wordsmith only. Therefore, to extract this value it was necessary to upload the corpus on Wordsmith. For what concerns the other tools such as keywords and concordances, research conducted on Wordsmith was replicated on SketchEngine as accurately as possible. In particular, the Concordance, Keyword list, Wordlist and N-grams tools were used.

Concordance tool was the most used tool. Its *advanced* option allows a set of different types of searches. In particular, *simple* and *CQL* are the most used query types.

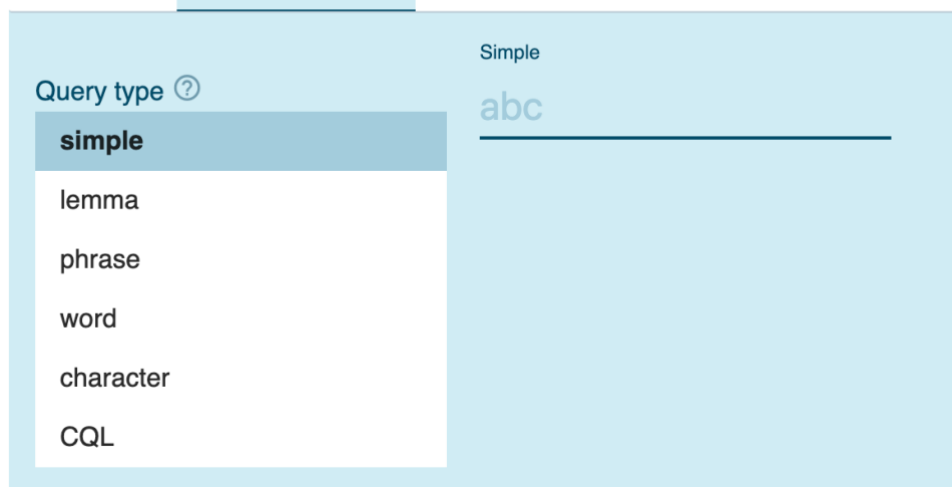


Figure 2.27: The concordance setting panel on SketchEngine, where it is possible to select different types of query type, as “simple” or “CQL”.

Simple query is case insensitive, which means that no distinction is made between uppercase and lowercase. Also, it looks for every inflected form of the word. So, if the research is, for instance, “cittadino”, results might include “Cittadino”, “cittadino”, “Cittadina”, “cittadina”, “Cittadini”, “cittadini”, “Cittadine” and “cittadine”.

The CQL query type is a special query language which allows to search for complex grammatical or lexical patterns:

“The Corpus Query Language (CQL) is a special code or query language used in Sketch Engine to search for complex grammatical or lexical patterns or to use search criteria which cannot be set using the standard user interface.”⁴⁶

Thus, if the research was meant to find “Cittadino/a/i/e dell’Unione”, the CQL used is: `[word=”Cittadin[o|a|i|e]”][word=”dell”][word=”Unione”]`.

The CQL query type is sometimes used as a sort of litmus test to check the accuracy of the simple query type. Indeed, the results must be the same of those of the simple search. If they do not match, it means that an error was made, and it is necessary to go back and solve it.

Keywords and terms extraction tool is used to retrieve words, or a set of words, which are typical of the corpus analyzed:

⁴⁶ <https://www.sketchengine.eu/documentation/corpus-querying/>

“[...] word and phrases typical for your corpus because they appear in your corpus more frequently than they would in general language. They can be used to define or understand the main topic of the corpus. Sketch Engine combines statistics with linguistic criteria to extract keywords and terms.” (SketchEngine⁴⁷, n.a.)

Keyword extraction tool searches for single words (tokens) which appear in the focus corpus more often than they do in general language (or in a reference corpus). The same applies for the terms extraction tool, with the difference that terms extraction refers to multi-words units (phrases).

Wordlist tool generates a frequency word list which can be of various kind, such as part of speech (nouns, verbs and adjectives), words containing certain characters, or attributes (word forms, tags, lemmas etc.). It can also be a combination of the three options above.

The N-Grams tool produces frequency lists of sequences of tokens. This tool is extremely helpful when searching for lexical bundles, or more in general, for sequences of words which appear in the observed corpus with a higher frequency than they do in ordinary language. Here, N is a variable that depends on the length of the string. This means that a 3-Gram, for instance, indicates a sequence of 3 tokens. Also,

“[...] N-grams are called multi-word expressions (or MWEs) or lexical bundles. The user has a choice of filtering options including regular expressions to specify in detail which n-grams should have their frequency generated. N-grams can be generated on any attribute with word and lemma being the most frequently used ones.” (SketchEngine⁴⁸, n.a.)

The tools above mentioned were used to replicate as closely as possible the research performed by Mori in her research (2018c). If, however, some operations were not available on SketchEngine, such as the sum of the normalized frequency of n-Grams (section 2.4.5.), an Excel⁴⁹ spreadsheet was used to perform the calculations.

⁴⁷ <https://www.sketchengine.eu/quick-start-guide/keywords-and-terms-lesson/>

⁴⁸ <https://www.sketchengine.eu/guide/n-grams-multiword-expressions/>

⁴⁹ <https://www.microsoft.com/it-it/microsoft-365/excel>

Results obtained in the CHEU-Lex Corpus were compared to those of Corpus A and Corpus B both through the normalized frequencies and by means of the Log Likelihood ratio (LL ratio), following the same method used by Mori (2018c).

Usually, the LL ratio calculates how statistically significant the difference between two measurement is. The higher the LL value, the more statistically different the use of the item and the lower the possibility that this case is due to randomness (Mori, 2018c). The tool used to calculate the LL ratio is the online LL wizard developed at Lancaster University.⁵⁰ It compares the relative frequencies of a specific element in two given corpora. The LL ratio itself is always a positive number. However, the script insert “+” or “-“ to indicate an over-underline of the element in corpus 1 compared to corpus 2.

	Corpus 1	Corpus 2
Frequency of word	<input type="text"/>	<input type="text"/>
Corpus size	<input type="text"/>	<input type="text"/>

Figure 2.28: The online LL wizard. To calculate the LL ratio, it is necessary to insert the raw frequencies of the element analyzed as well as the corpus size for both corpora.⁵¹ The tool automatically calculates the LL ratio and add a “+” or a “-“ to indicate the over/underline of the element in corpus 1 compared to corpus 2. (Adapted from <https://ucrel.lancs.ac.uk/llwizard.html>)

In the case of Mori (2018c), Corpus 1 refers to Corpus A, and Corpus 2 to Corpus B. Take, for instance, this example:

⁵⁰ <https://ucrel.lancs.ac.uk/llwizard.html>

⁵¹ Number of tokens of Corpus A: 1,439,069. Number of tokens of Corpus B: 1,978,795. Number of CHEU-Lex tokens were extracted from SketchEngine and are a total of 780,763. Data are retrieved from Mori (2018c: 205). The total of tokens refers to the enacting parts (body) sub-corpora only, since the analysis is mainly performed on the *enacting part* sections. Concerning the *preambles* sections, the total of tokens is as follows: Corpus A = 619,181; Corpus B = 72,134; CHEU-Lex corpus = 6,934.

EU-noun phrases	Corpus A		LL ratio	Corpus B	
	Raw frequency (RF)	Normalised frequency per million (NF/pm)		Raw frequency (RF)	Normalised frequency per million (NF/pm)
<i>Cittadino/a/i/e dell'Unione</i> ³⁵	82	57	+2.59	88	44
<i>Commissione (europea)</i>	5,516	3,833	+5,427.09	864	437
<i>Comunità (europea/e)</i>	1,673	1,163	+1,570.04	289	146
<i>Consiglio (dell'Unione europea)</i>	2,425	1,685	+1,119.18	1,024	517
<i>Paesi Terzi</i>	1,174	816	+731.53	372	188
<i>Parlamento europeo</i>	1,603	1,114	+1,205.40	401	203
<i>Stato/i Membro/i</i>	16,275	11,309	+17,528.52	2,058	1,040
<i>Unione europea</i>	654	454	+5.21	797	403

Table 2.3: A random sample showing the use of LL ratio. In particular, this table compare the use of EU-noun phrases both in Corpus A and Corpus B (Adapted from Mori, 2018 c: 210).

Here, the LL ratio is in the middle between the two corpora results. Take for instance *Commissione (europea)*: the LL ratio is + 5,427.09. This means that there is a high overuse of this EU-noun phrase in Corpus A compared to Corpus B.

This tool is useful when comparing the normalized frequencies of the elements, as it tells you how statistically different the frequency is. Also, “if the LL value is higher than 6.63, there is a 99% probability that the frequency difference is statistically significant” (Mori, 2018c: 208). However, having a positive LL ratio does not necessarily mean that the item is more frequent in Corpus A if compared to Corpus B; this type of information can be deducted only by looking at their normalized frequencies.

Eventually, at the end of all Sections, results were represented in graphs, in order to have a visual representation of the distribution of items in the three corpora. Graphs were realized by means of an Excel spreadsheet⁵². The normalized frequencies of the items selected in each Section were either represented alone or summed up together for every corpus. These, together with the LL ratio and the comparison of the normalized frequencies, provided a further comparison of the data.

⁵² <https://www.microsoft.com/it-it/microsoft-365/excel>

2.4.1. The Research Template

The following table represents *The Eurolect Observatory Research Template* conceived by Mori (2018b) as a reference to be followed when carrying out the comparison between the two corpora (Corpus A and Corpus B) for each of the 11 languages chosen for the project:

Levels	Heuristic categories		
	EU-rooted phenomena	Contact-induced features	Intra-linguistic variability
Lexis	Europeisms: a. EU newly-coined words b. EU-noun phrases (3-grams, 4-grams) c. EU-based metaphors d. Semantic Europeisms e. European acronyms	Contact-induced variants: Loanwords, structural (derivation/composition) Calques and semantic Calques Eurolect-related lexical items	Differences in lexical distribution: – Higher/lower register variants – Latinisms – Mixed register – Old-fashioned words – Lexical variation (Type/Token ratio)
Lexical Morphology		Morpheme borrowing; Eurolect-driven morphological choices	
		International prefixes and neo-classical compounds	
		Contact-induced degree of morphological productivity in word-formation	
Verb Morphology		Range of tenses and moods according to legislators' aim, text function and legal system; Presence of distinctive features of legal language: – Deontic modality – Impersonal structures – Modals – Passive voice – Past and present participles	
Morphosyntax	Eurolect-related bundles (N-grams)		Nominal vs. verbal style; Different use of legal-administrative collocations; Complex prepositions for legal mapping
Syntax	Eurolect-related Word order (SVO and Non-Pro Drop preference)		Differences in sentence complexity; Use of typical inter-sentential connectives
		Syntactic calques	
Textuality/ Discourse	Macro-structural text organisation	Inference markers and logical connectors	
	Eurolect-related cohesion devices		
			Patterns for legal framing

Table 2.4: A visual representation of the scheme conceived by Mori. (Adapted from Mori, 2018: b, 19)

Linguistic levels, i.e. *Lexis*, *Lexical Morphology*, *Verb Morphology*, *Morphosyntax*, *Syntax and Textuality/Discourse* are represented in rows, and must be analyzed in concordance with the three criteria described in columns: *EU-rooted phenomena*, *Contact-induced phenomena*, and *Intra-linguistic variability* (section 1.5.3.).

A specific procedure was followed in this research: for each level, results were analyzed according to the three heuristic categories. This means that for each level, a small Section is devoted to the description of the analysis, and, if present, the three criteria are taken into consideration to interpret the results obtained.

Contact-induced phenomena were not considered. Indeed, as the name suggests, this type of phenomenon originates from the contact between two languages, in this case from the contact between the source language (SL) and the target language (TL). However, the two linguistic varieties under analysis (Swiss and domestic legal Italian of implementation laws) do not share the same implementation procedure and, most important, do not share the same source language (see Section 2.2.). This implies that the comparison of *Contact-induced phenomena* would be difficult to carry out without the third part of the comparison, i.e. a common source language, that would legitimate this type of research. This category was therefore avoided unless a particular motivation was given.⁵³

In the light of the above, for every linguistic level, corpus-based and/or corpus-driven analysis was carried out, in order to analyze the frequency and the distribution in the CHEU-Lex Corpus of specific items selected by Mori (2018b).

2.4.2. Lexical analysis

The first level analyzed by Mori (2018c) is the *Lexical level*, which was proved extremely profitable in certain legal areas such as finance law (Caterina and Rossi, 2008), and it is therefore expected to be the most relevant one.

First, lexical variety was analyzed. In Mori's chapter (2018c), for each corpus (Corpus A of European directives and Corpus B of their Italian laws of implementation) the standardized

⁵³Something different happens for Latinisms, which, despite being considered as contact-induced phenomena, are worth analyzing, since the "source language" with which both the target languages got in touch with is Latin. For further explanation see Section 2.4.2.3.

Type/Token ratio (sTTR) of two sub-sections, i.e. the *enacting terms* and the *preambles*, was compared:

	Corpus A (EU directives)	Corpus B (National Implementation Laws)
Enacting terms		
Types	18,660	22,881
Standardised Type/Token Ratio	31.96	34.46
Preambles		
Types	14,717	2,480
Standardised Type/Token Ratio	37.32	23.22

Table 2.5: The results obtained by Mori during her research. (Adapted from Mori, 2018: c 208).

The sTTR of the *enacting terms* of Corpus A is 31.96, whereas the sTTR of the *enacting terms* of Corpus B is 43.46. These results show that the *enacting terms* of Corpus A are less varied than those of Corpus B. However, lexical variety reaches higher values in *preambles* of Corpus A (37.32) if compared to those of Corpus B (23.22). In general, these results underline the fact that there tend to be

“[...] a preference for synonymy in legal Italian compared to Italian Eurolect, which is rather affected by literal translation from EU versions in which Eurolect lexical types tend to be more repeated (less lexically rich), thus affecting cohesion as well. Needless to say that in Italian Eurolect lexical choices mirror lexical variants adopted in master texts (basically English and French) and, consequently, they may also be interpreted as contact-induced phenomena”
(*Observing Eurolect*, 2018: 208)

To compare the lexical variety of the CHEU-Lex corpus to that of Corpus B, it was necessary to extract the sTTR from the same sub-sections, i.e. *enacting terms* (see Section 2.4.) and *preambles*. However, the sTTR tool is available on Wordsmith only (section 2.4.). Therefore, the *enacting terms* and the *preambles* of the CHEU-Lex Corpus were uploaded on Wordsmith and the sTTR was extracted for both sub-sections. Results were then compared to those of Corpus B and Corpus A (Mori, 2018c).

2.4.2.1. EU-rooted lexical features

Another interesting category that was analyzed is that of *Europeisms*. These are defined as

“EU newly-coined words [...] lexical items referring to concepts, institutions, policies, principles and practices and are highly contextualized since they are clearly rooted in the EU context. Among possible lexical strategies, newly-coined words that denote EU matters are limited, if compared to other categories based on contact-induced phenomena, such as calques – either structural or semantic [...]” (Mori, 2018c: 209).

In the framework of the *Eurolect Observatory Project*, a list of items was selected from previous qualitative analyses by Mori, and statistics of the cross-corpora distribution of Europeisms and EU-noun phrases were compared as shown in the following tables (Table 2.6 and 2.7):

Table 3. Selected EU-noun phrases

EU-noun phrases	Corpus A		LL ratio	Corpus B	
	Raw frequency (RF)	Normalised frequency per million (NF/pm)		Raw frequency (RF)	Normalised frequency per million (NF/pm)
<i>Cittadino/a/i/e dell'Unione</i> ³⁵	82	57	+2.59	88	44
<i>Commissione (europea)</i>	5,516	3,833	+5,427.09	864	437
<i>Comunità (europea/e)</i>	1,673	1,163	+1,570.04	289	146
<i>Consiglio (dell'Unione europea)</i>	2,425	1,685	+1,119.18	1,024	517
<i>Paesi Terzi</i>	1,174	816	+731.53	372	188
<i>Parlamento europeo</i>	1,603	1,114	+1,205.40	401	203
<i>Stato/i Membro/i</i>	16,275	11,309	+17,528.52	2,058	1,040
<i>Unione europea</i>	654	454	+5.21	797	403

Table 2.6.: The raw and normalized frequency of the list of the EU-noun phrases selected by Mori both in Corpus A and Corpus B. (Adapted from Mori, 2018c: 210).

Results comparing the use of semantic Europeisms are shown in the following table (Table 2.7.). These are defined as “[...] words with a generic meaning that belong to several word classes and have acquired a contextualized meaning thanks to predominant use in EU discourse.” (Mori, 2018c: 211).

Table 4. Semantic Europeisms

Semantic word classes	Semantic Europeisms	Corpus A		LL ratio	Corpus B	
		RF	NF/pmw		RF	NF/pmw
Noun	<i>adesione</i> (inclusion of a new Member State)	111	77	+142.13	8	4
	<i>armonizzazione</i>	86	60	+25.46	49	25
	<i>attuazione</i>	732	509	+241.15	392	198
	<i>direttiva/e</i> ³⁶	16,387	11,387	+16,000.80	2,609	1,318
	<i>ravvicinamento</i> ³⁷	66	46	+37.61	23	12
	<i>recepimento</i>	10	7	-198.70	243	123
	<i>regolamento/i</i> (CE/CEE)	710	493	+86.60	581	294
	<i>trasposizione</i>	7	5	+3.17	3	2
Adjective	<i>armonizzata/e/o/i</i>	168	117	+37.68	110	56
	<i>attuativa/e/o/i</i>	4	3	-115.11	132	67
	<i>comunitaria/e/o/i</i>	1,105	768	-12.07	1728	873
Verbs (all forms)	<i>armonizzare</i>	200	139	+47.94	127	64
	<i>attuare</i>	164	114	+7.16	168	85
	<i>recepire</i>	30	21	-2.37	58	29

Table 2.7. The raw and normalized frequency of the list of the semantic europeisms selected by Mori both in Corpus A and Corpus B. (Adapted from Mori, 2018: c, 211).

Thanks to the normalized frequency (NF/pmw), it was possible to establish that both the EU-noun phrases and the semantic Europeisms are extremely productive in the EU law-drafting process, and to a less extent in their Italian implementation laws.

By means of the *advanced Concordance Corpus Tool* available on *SketchEngine*, the same corpus-based research was carried out on the CHEU-Lex corpus. This means that the same list of items was adopted, and every single item, be it an EU-noun phrase or a semantic Europeism, was retrieved on the CHEU-Lex corpus by means of the *advanced Concordance Corpus Tool*. For this research, the query type “simple” was selected in the first place. Then, to check accuracy, results were cross-checked with the results obtained using the CQL query type (see Section 2.4.).

Eventually, the frequencies were normalized (NF/pmw) and compared to those of Corpus B obtained by Mori (see figures 20 and 21. Also check Mori, 2018c). To better interpret the results, the LL ratio was also retrieved by means of the online LL wizard developed at Lancaster University⁵⁴ (see also Section 2.4.1.). Further proves were provided by the comparison with results obtained in Corpus A (Mori, 2018c).

⁵⁴ <https://ucrel.lancs.ac.uk/llwizard.html>

2.4.2.2. Intra-linguistic lexical variability

Previous studies led by Mori on the Italian of EU directives underlined the use of a less formal register, aiming at making EU Law more approachable by European citizens. This means that ordinary lexical variants are expected to be widely present in the EU directives (Mori, 2018c).

In this sense, some distinctive features of legal and administrative Italian were analyzed, such as technical verbs or some marked orthographic choices. These items act as register markers and could be omitted without any significant semantic loss. The variation of their conventional use within the two legislative varieties of Italian analyzed by Mori (Italian of EU directives and the Italian of domestic laws of implementation) makes them a reference point when analyzing the intra-linguistic lexical variability.

Table 7. Distribution of legal verbs

Legal verbs (inflected forms)	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>adire</i>	24	17	-0.30	38	19
<i>adottare</i>	4,685	3,256	+1,511.14	2,541	1,284
<i>agire</i>	143	99	-1.95	228	115
<i>applicare</i>	3,180	2,210	+94.58	3,440	1,738
<i>avvalere</i>	157	109	-8.88	289	146
<i>depositare</i>	104	72	+1.27	123	62
<i>emanare</i>	29	20	-612.03	740	374
<i>ottemperare</i>	95	66	+0.15	124	63

Table 2.8. The distribution of some legal verbs both in Corpus A and Corpus B (Adapted from Mori, 2018c: 216).

The table 2.8. shows the results obtained when comparing the distribution of legal verbs in Corpus A of EU directives and Corpus B of their domestic laws of implementation. In this case, the LL ratio helps comparing how statistically relevant the difference is. The same research was performed on register markers and intra-genre variants, as shown in table 2.9:

Table 6. Register markers and intra-genre variants

Item	Corpus A		LL ratio	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>altresi</i> ⁴³	195	136	-164.16	700	354
<i>apposito</i>	14	10	-192.93	255	129
<i>attraverso</i>	179	124	-94.51	543	274
<i>a titolo di</i>	23	16	-27.22	95	48
<i>concernente/i</i>	447	311	+6.98	518	262
<i>inerente/i</i> ⁴⁴	73	51	-21.12	185	93
<i>di concerto</i>	22	15	-639.31	732	370
<i>in materia di</i>	823	572	-44.33	1,505	761
<i>in quanto</i>	130	90	-9.42	248	125
<i>in sede di</i>	43	30	-126.22	279	141
<i>in tema di</i>	12	8	-32.93	75	38
<i>mediante</i>	585	407	-49.11	1,142	577
<i>minore/i</i>	95	66	-13.17	203	103
<i>ovvero</i> ⁴⁵	358	249	-753.08	1,942	981
<i>per mezzo di/del/della/dei/delle</i>	24	17	-0.05	35	18
<i>presso</i>	321	223	-224.91	1,075	543
<i>riguardante/i</i>	394	274	+46.76	325	164
<i>se del caso</i> ⁴⁶	595	413	+339.65	207	105
<i>tramite</i>	241	167	-29.00	501	253

Table 2.9. The table shows the results of the list of register markers and intra-genre variants (Adapted from Mori, 2018c: 215)

Overall, the research conducted in the *Eurolect Observatory Project* framework underlined the conservative character of national legislative Italian of implementation laws, which reveals itself in the use of higher register variants (Mori, 2018c).

To the purpose of this research, the same elements belonging both to the legal verbs and the register markers and intra-genre variants were extracted from the CHEU-Lex corpus by means of the *advanced Concordance Corpus Tool*. Legal verbs were firstly searched through the “simple” query type, and then results were checked thanks to the CQL query type. The same applies to the register markers and intra-genre variants. However, a research such as “per mezzo di/del/della/dei/delle”, required the use of the CQL query type only, due to its complex structure (the simple query search “per mezzo di” would not find “per mezzo del” or “per mezzo della”, as they are not inflected forms, nor they differ in upper-lower case letters.)

Then, normalized frequencies were compared to those obtained by Mori in Corpus B both by means of the LL ratio and the normalized frequencies, which was calculated on the online

LL wizard⁵⁵. The same comparison was also carried out between the CHEU-Lex Corpus and Corpus A, as previously explained.

2.4.2.3. Latinisms

Although Latinisms are considered a contact-induced lexical phenomenon and should therefore not be considered part of this study (see Section 2.4.1.), they are worth analyzing, since the contact does not take place straightforwardly from a single language, but it is instead a deeper and older contact, dating back to different periods in history and concerning different languages (Mori, 2018c). Indeed, Latinisms are used in almost all legal languages in western culture.

Their function in the European environment is related to the choice of the register, and this is true in many of the European languages, as far as the legal field is concerned.

Table 5. Eurolect renderings

	Corpus A		LL ratio	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>ad hoc</i>	16	11	+14.39	3	2
<i>acquis</i>	55	38	+95.15	0	0
<i>autorità competente/i</i>	3,521	2,447	+1,362.81	1,698	858
<i>conformemente</i>	1,604	1,115	+876.09	583	295
<i>mutatis mutandis</i>	42	29	+72.66	0	0
<i>post mortem</i>	11	8	+2.62	7	4
<i>regolamentare/i</i>	812	564	+600.33	208	105
<i>regolamentazione</i> ³⁹	602	418	+540.93	113	57
<i>status</i>	141	98	+52.30	70	35

Table 2.10 Results of the use of Eurolect renderings both in Corpus A and Corpus B (Adapted from Mori, 2018c: 213).

In the frame of this research, the same items showed in Table 2.10 were extracted in the CHEU-Lex corpus by means of the *advanced Concordance Corpus Tool*. In this case, as well as in the previous ones, the “simple” query type was selected, and results were then checked by performing the same research using the CQL query type.

Eventually, results were compared using the normalized frequencies and the LL ratio.⁵⁶ If the latter happens to be mostly positive, then an overuse of Latinism can be observed in the CHEU-Lex corpus compared to Corpus B. This could imply the influence of the European legal drafting process on the Swiss legal Italian. Eventually, a cross-comparison was

⁵⁵ <https://ucrel.lancs.ac.uk/llwizard.html>

⁵⁶ <https://ucrel.lancs.ac.uk/llwizard.html>

performed by comparing then the results with those of Corpus A, in order to detect any possible likelihood between the two corpora, and therefore between the Italian of EU directives and the Swiss Italian of implementation laws, which would either prove or disprove (in case of evident discrepancies between the corpora) the possible European influence.

2.4.3. Lexical morphology

For what concerns word formation, there is a particular phenomenon that is worth analyzing: the overuse of international prefixes. These are prefixes which are widely used in different languages with the same meaning. An example to this can be “extra”, which is profitable in Italian as well as in other European Languages such as French, English, German etc. These are strictly connected with semantic cross-language transparency, and they might underline the effect of Europeanisation in some legal languages (Biel, 2014). Nonetheless, it is also subjected to cross-languages differences. In particular, “As far as Italian is concerned, its neo-Latin origin results in an extensive use of classic prefixes for word formation.” (Mori, 2018c: 217).

Table 8. International prefixes

Item	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>anti</i> + N/Adj.	333	231	+112.08	176	89
<i>bio</i>	97	67	-5.59	179	90
<i>eco</i>	72	50	+0.16	93	47
<i>extra</i> + Adj.	31	22	-46.87	143	72
<i>intra</i> + Adj.	96	67	+8.33	86	43
<i>multi</i> + Adj.	121	84	+9.10	112	57
<i>post</i> + N/Adj.	17	12	+3.19	12	6
<i>semi</i> + N/Adj.	15	10	-1.19	29	15
<i>sub</i> + N/Adj.	15	10	-94.81	156	79

Table 2.11. The use of the international prefixes both in Corpus A and Corpus B. (Adapted from Mori, 2018c: 217)

Based on the list in table 2.11, the same set of prefixes both from Latin and Greek was selected and, by means of the *Concordance Corpus Tool*, their frequency was compared to that of Corpus B. This time it was necessary to use the CQL tool only, as was necessary to perform a more complex research, such as “*prefix + Noun*” or “*prefix + Adjective*”, to look for the specific items.

For instance, the research “multi + Adj” aims at obtaining a list of adjectives which start with “multi”. This means that results such as “multiculturalità”, which is a noun (actually found in CHEU-Lex corpus, raw frequency:1), had to be excluded. This was possible thanks to the use of the CQL query search, which, in this case, was [word="multi.*" & tag="ADJ"]. Once the results were obtained, a further selection was carried out by hand to eliminate all those words which happened to share the same first letters with the international prefixes but actually had nothing to do with them (for instance, when searching for words with “anti” as prefix, the word *antico* and all its derived words were to be omitted, since they do not actually begin with the prefix “anti”).

The LL ratio was retrieved online, to check whether any statistically relevant difference in the use of some items could be observed. Then, a cross-comparison was carried out between the CHEU-Lex corpus and the results of Corpus A. If some likelihood is found at this level, this might imply that there is some influence of the European legal drafting process on the Swiss Italian of implementation laws.

As explained by Mori, both varieties of legal Italian (Italian of EU directives and the domestic variety of the laws of implementation) are characterized by Latin prefixes for two main reasons: first, for what concerns European legislation, this complies with the need for equality among the different languages, which can be achieved also through borrowing, like in this case. Second, the Italian corpus of Implementation laws simply mirrors a word-formation method that is extremely productive both in technical and scientific varieties of Italian (Mori, 2018c).

2.4.4. Verb morphology

Another interesting phenomenon worth analyzing is verb morphology. Verb tenses are among the most relevant features to be observed in legal language. Their peculiarity is due to the fact that, according to Mattila (2013), law aims at regulating hypothetical future events on the basis of experience drawn from the real world. Therefore, their timespan is characterized by a certain universality, which is reflected, among other features, using verbs in legislative language. Within the Eurolect project, different verb tenses were analyzed, and results underlined a reduced range of tenses and moods (which were mainly present and indicative, respectively), “[...] according to what is expected in legal Italian” (Mori, 2018c: 220).

A corpus-driven analysis was carried out by Mori (2018c), who compared verbs in the present indicative. According to Mori, this specific tense was the most represented from the keyword analysis of Corpus A (Mori, 2018c: 220). In particular, the third person (both singular and plural) was considered. Wordlists of the enacting parts of both corpora were retrieved and compared, as showed in the following chart (table 2.12).

Table 9. The cross-corpora distribution of most frequent inflected verbs (first 1,000 words)

Corpus A	Wordlist ranking	Corpus B	Wordlist ranking
<i>possono</i>	n. 57	<i>può</i>	n. 67
<i>sia</i>	n. 66	<i>sia</i>	n. 79
<i>siano</i>	n. 69	<i>possono</i>	n. 80
<i>può</i>	n. 73	<i>deve</i>	n. 120
<i>adottano</i>	n. 122	<i>siano</i>	n. 132
<i>ha</i>	n. 126	<i>devono</i>	n. 137
<i>applicano</i>	n. 163	<i>ha</i>	n. 164
<i>applica</i>	n. 197	<i>applica</i>	n. 209
<i>provvedono</i>	n. 211	<i>provvede</i>	n. 285
<i>hanno</i>	n. 201	<i>hanno</i>	n. 289
<i>comunicano</i>	n. 214	<i>modifica</i>	n. 467
<i>entra</i>	n. 284	<i>intende</i>	n. 533
<i>contengono</i>	n. 290	<i>abbiano</i>	n. 625
<i>informano</i>	n. 304	<i>sarà</i>	n. 703
<i>prescrivono</i>	n. 653	<i>viene</i>	n. 706
<i>trattino</i>	n. 664	<i>provvedono</i>	n. 712
<i>stabilisce</i>	n. 779	<i>riguarda</i>	n. 746
<i>intendono</i>	n. 785	<i>comunica</i>	n. 760
<i>soddisfano</i>	n. 804	<i>adotta</i>	n. 814
<i>garantiscono</i>	n. 808	<i>costituisca</i>	n. 854
<i>vengano</i>	n. 849	<i>emana</i>	n. 881
<i>contiene</i>	n. 877	<i>trasmette</i>	n. 907
<i>comunica</i>	n. 932	<i>dispone</i>	n. 927
<i>indica</i>	n. 978	<i>intendono</i>	n. 938
		<i>costituisce</i>	n. 944

Table 2.12. The list of the most frequent inflected verbs among the first 1000 words from wordlists extracted from the two corpora. (Adapted from Mori, 2018c: 221).

To carry out the same research, a wordlist (see Section 2.4.) was extracted from the CHEU-Lex corpus. Ranking of the verbs in Corpus B were then compared to the ranking of the same verbs in the CHEU-Lex wordlist. The same comparison was carried out with the ranking of verbs in Corpus A. Only those within the first 1000 results were considered.

Mandatory information is another important feature that verbs in legal language need to express. To do so, impersonal structures such as *è necessario* and *occorre* were used. Mori observed a normalized frequency (NF) of *è necessario* higher in Corpus A (561 tokens) than in Corpus B (42 tokens). *Occorre*, instead, happens to be very limited in both corpora (NF: 33 in Corpus A and NF: 20 in Corpus B).

The use of the modal verbs *dovere*⁵⁷ and *potere*, were also analyzed (Table 2.13):

Table 10. Modals *dovere* and *potere* (3rd person singular and plural)

Obligation and permission modals	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>deve</i>	1,188	826	-30.53	1,997	1,009
<i>devono</i>	1,139	791	-21.56	1,863	941
<i>può</i>	2,545	1,769	-0.00	3,500	1,769
<i>possono</i>	3,664	2,546	+616.81	2,696	1,362

Table 2.13. The use of the 3rd person plural and singular of the modal verbs *dovere* and *potere* in both corpora. (Adapted from Mori, 2018c: 223)

In this specific research, to carry out the same analysis, the *advanced Concordance Corpus tool* was used. “Simple” query type was selected, and the research was performed for every inflected form shown in Table 2.13. Results were then compared to those of Corpus B and those of Corpus A, either by means of the LL ratio and through the normalized frequencies.

The use of passive voice is a feature that is also worth considering. In this particular verb form, the object of the action is brought into the foreground, whereas the subject is put in the background. It is worth noting that, according to Mori (2018c), the agent of the passive form is mainly made explicit in the Italian of EU directives through the prepositional phrase *da essi/esse* (NF: 116 for Corpus A; NF: 43 for Corpus B). Log Likelihood ratio in the following table (table 2.14) shows a statistically high preference for passive forms in the Italian of EU directives, whereas passive forms codifying a deontic value (i.e. past participle preceded by either the modal verb *dovere* or *andare*) are more frequent in domestic variety (their LL ratios are -8.68 and -13.31, respectively).

⁵⁷ Interestingly enough, according to guide styles obligations expressed through the use of *dovere* should be avoided. Nonetheless, it is found in both legislative varieties (Mori, 2018).

Table 11. Passive forms

	Corpus A		LL ratio	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>è stato/a</i> + PP	341	237	+16.07	345	174
<i>sono state/i</i> + PP	323	224	+25.41	296	150
<i>viene</i> + PP	315	219	+33.03	269	136
<i>vengono</i> + PP	223	155	+27.95	181	91
<i>va</i> + PP	19	13	-8.68	55	28
<i>vanno</i> + PP	41	28	+14.50	21	11
<i>deve essere</i> + PP	279	194	-48.50	625	316
<i>devono essere</i> + PP	405	281	-13.31	698	353

Table 2.14. The use of the passive forms both in Eurolect Italian (Corpus A) and in the legal Italian of laws of implementation (Corpus B). (Adapted from Mori, 2018c: 224).

To perform the same-sort analysis, the same items observed in Table 2.14 were extracted from the CHEU-Lex corpus by means of the *advanced Concordance Corpus tool*. This time, “CQL” query type was selected (see Section 2.4.). Indeed, to perform this type of research, it was required to search for specific patterns. For instance, to obtain the normalized frequency of “è stato/a + Past Participle”, it was necessary to set a research query that matched this specific pattern. In particular, “Past Participle”, as well as other verb forms, can be searched only through the “CQL” query type, as on SketchEngine these elements are tagged. Therefore, to search the example above mentioned, the CQL entered was the following: [word=”è”][word=”stato|stata”][tag=”VER:ppast”], where “VER:ppast” is the tag used by SketchEngine to mark the Past Participle. Same-sort queries were created for the other items.

Eventually, the normalized frequencies were compared to those of Corpus B and Corpus A obtained by Mori (2018c). Log Likelihood ratio was retrieved from the online Log-likelihood wizard⁵⁸, to check whether some important differences do emerge from the comparison with the two corpora.

2.4.5. Morphosyntax

Phrasemes, fixed formulas, and ready-made sentences fall into the category of morphosyntax. Keeping in mind that morphosyntax studies the relation among morphemes and words when creating a sentence, this specific analysis was based on the concept of

⁵⁸ <https://ucrel.lancs.ac.uk/llwizard.html>

collocations, i.e. the list of the most frequent words occurring in a specific position near the selected search word. When facing fixed formulas such as “*fatt* salv**” (Table 2.15) for instance, it is highly likely that, in such specific legal corpora, “*salv**” will occur very often near “*fatt**”, due to the fact that this fixed formula is widely used in this genre. A corpus-based analysis was carried out on selected legal collocations, with the aim of spotting any remarkable cross-corpora difference (Mori, 2018c).

Table 12. Legal-administrative collocations

	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>fatt* salv*</i>	875	608	+32.37	919	464
<i>ferm* restando</i>	59	41	-144.01	346	175
<i>quanto + PP</i>	138	96	-1,064.63	1644	831
<i>in deroga a*</i>	319	222	+31.39	277	140
<i>in base a*</i>	692	481	+50.75	644	325
<i>sulla base d*</i>	352	245	-174.55	1046	529
<i>in ottemperanza a*</i>	7	5	+0.35	7	4
<i>nel rispetto d*</i>	136	95	-478.42	979	495
<i>tenuto conto di*</i>	158	110	+166.87	21	11
<i>a seguito d*</i>	136	95	-0.00	188	95
<i>a carico d*</i>	84	58	-414.09	743	375
<i>in seguit a*</i>	79	55	+30.69	38	19
<i>secondo quanto + PP</i>	58	40	-64.35	233	118
<i>senza pregiudizio d*</i>	44	31	+29.12	13	7
Total	3,137	2,180	-570.61	7,098	3,587

Table 2.15. The distribution of a set of legal-administrative collocations both in Corpus A and Corpus B. (Adapted from Mori, 2018c: 225).

The same-sort information was extracted from the CHEU-Lex corpus, to check to what extent the use of such items is similar to Corpus B or, otherwise, to Corpus A. The *advanced Concordance Corpus Tool* was used. A combination of simple and CQL query types was used for every item searched, to check the accuracy of results. For instance, the query for *in deroga a** must find all combinations of *in deroga* and *a|ai|alla|alle|allo|agli|all'*, and it is therefore necessary to include these options in the CQL query, which is as follow: [word="in"][word="deroga"][word="a|ai|alla|alle|allo|agli|all'"]. However, other queries such as “*fatt* salv**” were easier to retrieve, and for these the simple query type was sufficient (although a CQL query type was performed anyway, to check the accuracy of results). Eventually, results were compared to those obtained by Mori in Corpus B and in Corpus A (2018c), both by means of the LL ratio and by looking at the normalized frequencies.

“Lexical bundles” were also retrieved by Mori in both corpora. These are described as “[...] multi-word sequences made of strings of variable size, regardless of their meaning and of any syntactic tie” (Mori, 2018c: 226). To do so, the N-Grams function by *WordSmith*⁵⁹ was used, setting “N” from 3 to 6 and the minimum threshold set at 5. Then, results for every set of N-Grams (therefore of 3-Grams, 4-Grams etc.) were analyzed according to their frequency.

Table 13. Distribution of N-Grams

Lexical bundles (N-Grams)	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
3-Grams	34,076	23,679	-10.75	47,957	24,235
4-Grams	23,095	16,049	+12.83	30,781	15,555
5-Grams	45,970	31,944	+20,442.53	20,000	10,107
6-Grams	11,227	7,802	+89.16	13,685	6,916

Table 2.16. The distribution of the various N-Grams in both Corpus A and Corpus B. (Adapted from Mori, 2018c: 226)

A same-sort analysis was carried out on the CHEU-Lex Corpus as well, since *SketchEngine* offers the same N-Gram tool as *WordSmith* (see Section 2.4). Since N-Grams length ranges from 3-Grams up to 6-Grams, four different types of research were performed, each for every different length. Frequency minimum was also set to 5. However, unlike *WordSmith*, *SketchEngine* did not allow to get the sum of the normalized frequency of all the lexical bundles. Instead, only the total frequency, i.e. the sum of the raw frequencies of every item, was displayed.

⁵⁹ It is worth remembering that the leading research was carried out by means of *WordSmith*, whereas this research was carried out using *SketchEngine*.

	Word	Frequency ?	Frequency per million ?
1	di cui all'	977	1,251.34 ...
2	cui all' articolo	889	1,138.63 ...
3	Il Consiglio federale	733	938.83 ...
4	secondo l' articolo	663	849.17 ...
5	di cui al	477	610.94 ...

Table 2.17. The screenshot is taken from SketchEngine, showing the N-Grams interface. It is possible to see that only the total frequency of the 3-Grams is available, which, in this case, happens to be 181,481.

However, to overcome this issue, the .xlsx file was downloaded for every research, and the sum of the normalized frequencies (frequency per million) was calculated by means of the Excel⁶⁰ spreadsheet.⁶¹

Once the total normalized frequency was calculated, results were compared to those obtained by Mori in Corpus B and Corpus A. Eventually, the LL ratio was also extracted to better compare the distribution of the various N-Grams in the CHEU-Lex and the two corpora (Corpus A and Corpus B).

2.4.6. Syntax

Previous studies showed a relevant difference concerning the level of sentence complexity between the two legal varieties, i.e. the Italian of Eu directives and the Italian used in the domestic laws of implementation. In particular, the Italian domestic variety of implementation laws was observed to be much more complex and denser than the Italian of EU directives (Mori, 2018c). Differences were analyzed considering the EU-rooted phenomena and Intra-linguistic variability.

⁶⁰ <https://www.microsoft.com/it-it/microsoft-365/excel>

⁶¹ To calculate this, the following mathematical proportion could have been used, too: $f:fpm=tf:x$, where “f” stands for the frequency of a random item, “fpm” stands for its normalized frequency and “tf” stands for the total frequency. X is the value to be calculated and refers to the total normalized frequency. However, results were approximated, and the “SUM” formula on Excel was observed to be more accurate.

2.4.6.1. EU-rooted syntactical phenomena

For what concerns sentence structure, previous results underlined EU-derived features which distinguish Italian of EU directives from its domestic variety of implementation laws (Mori, 2018c). For instance, a rather fixed EU syntactic order was observed in the EU variety: SVO, namely Subject, Verb and Object (Mori, 2018c). Also, it has been noted that subjects occurring at the beginning of sentences are generally not omitted in the Italian of EU directives, and this has to do with the “syntactic skeleton of EU master texts written-negotiated-revised in English and French” (Mori, 2018c: 227).⁶²

On the one hand, cohesion in the Italian of EU directives is assured by recalling subjects throughout the text either by means of partial repetition or the use of anaphoric subject pronouns. On the other hand, in national legal Italian subject ellipsis is more widespread. This can be observed in table 2.18: indeed, there is a high overuse of subject pronouns in Corpus A when compared to Corpus B, underlying the fact that the Italian of EU directives tends to avoid the subject ellipsis. Instead, legal Italian of laws of implementation is more likely to omit the subject pronouns shown in table 2.18.

Table 14. Subject pronouns

	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>esso</i>	335	233	+99.08	191	97
<i>essa</i>	417	290	+350.42	175	88
<i>essi</i>	1,397	971	+1,320.16	238	120
<i>esse</i>	263	183	+107.49	122	62

Table 2.18. The overuse of the subject pronouns in Corpus A compared to Corpus B (Adapted from Mori, 2018c: 226).

A same-sort analysis was carried out on the CHEU-Lex Corpus, looking for concordances of “*esso*”, “*essa*”, “*essi*” and “*esse*” by means of the *Concordance Corpus tool*; the simple query type was selected. Normalized frequencies were compared to those of Corpus B and, later, to Corpus A, both by means of the normalized frequencies and the LL ratio. It is worth remembering that if the LL value is higher than the absolute value of 6.63, then it is highly likely that the frequency difference is statistically significant (see Section 2.4.1.).

⁶² English and French are described as “Non-Pro Drop languages” (Mori, 2018: 227), which means that, when drafting laws in these languages, subject ellipsis does not take place.

2.4.6.2. Intra-linguistic syntactic variability

Like the EU-rooted phenomena, previous qualitative studies by Mori (Mori, 2018c) underlined patterns characterizing the Italian of EU directives such as shorter sentences, lower structural complexity, and avoidance of embedded clauses. Again, this could be interpreted as the result of the need for simple and accessible legislative texts which characterize the European legal drafting process. These features were investigated by looking for inter-sentential connectives (table 2.19), which are reliable indicators of sentence complexity (Mori, 2018c). Interestingly enough, implicit and explicit connectives such as “*affinché + subjunctive*” (implicit) and “*per + infinitive*” (explicit), which introduce purpose subordinates, are more used in Corpus A than in Corpus B. This is due to the fact that EU directives “[...] have to define goals to be achieved by member states in their domestic laws” (Mori, 2018c: 231).

Table 17. Inter-sentential connectives

Grammatical information		Corpus A		LL	Corpus B	
		RF	NF/pmw		RF	NF/pmw
Purpose (explicit)	<i>affinché</i> ⁶³ + subj.	1,319	917	+1,191.94	245	124
Purpose (implicit)	<i>per + infinitive</i>	2,683	1864	+356.00	2,140	1,081
	<i>al fine di + infinitive</i>	435	302	-76.16	976	493
Cause	<i>dal momento che</i>	24	17	+41.52	0	0
	<i>dato che</i>	69	48	+88.24	5	3
	<i>perché</i>	41	28	+2.07	41	21
Relative	<i>cui</i>	13,169	9,151	-395.09	22,480	11,360
	<i>il/la quale; i/le quali</i>	773	537	-164.02	1,812	916
Restrictive and hypothetical	<i>se</i>	4,145	2,880	+542.10	3,321	1,678
Hypothetical	<i>qualora + subj.</i>	1,562	1,085	+65.02	1,612	815
Restrictive conditional	<i>a condizione che + subj.</i>	332	231	+26.60	303	153
	<i>laddove</i>	147	102	+33.93	95	48
	<i>nel caso in cui</i>	139	97	-141.77	539	272
	<i>nella misura in cui</i>	155	108	+111.93	41	21
	<i>ove</i>	466	324	-114.06	1,131	572
	<i>purché + subj.</i>	482	335	+84.08	350	177
	<i>sempreché</i> ⁶⁴ + subj.	55	38	+14.79	33	17

Table 2.19. The use of specific syntactical structures both in Corpus A and Corpus B. (The Adapted from Mori, 2018c: 230).

By means of the *Concordance Corpus Tool*, the same research was carried out on the CHEU-Lex corpus. In this case, the CQL query type was selected, as it was necessary to set up a

query search corresponding to complex patterns. Indeed, as it happened for verbs (see Section 2.3.4.), it was sometimes necessary to search for patterns containing specific verb forms, such as the subjunctive (*subj.*). So, for queries such as *affinché* + *subj* it was necessary to use the CQL language. In this case, the query would be `[word="[A|a]ffinché"][]?[tag="VER:fin"]`, where “VER:fin” indicates the presence of the finite form of a verb.⁶³ The same happened for elements which required the infinitive form of the verb. In this case, the tag changed from “VER:fin” to “VER:infi”. Eventually, results were cross-checked with those obtained from Corpus B and with those of Corpus A (Figure 41). The comparison took place thanks to the LL ratio, as well as the comparison of the normalized frequencies.

2.4.7. Textual level

The last level which was analyzed is the textual one. Due to their highly specific typology as well as their technical content, the structure of legal texts is expected to be fixed. Legal texts are, indeed, organized according to a hierarchy, which divides texts into units and sub-units (Mori, 2018c., and European Union, 2015.).

2.4.7.1. EU-rooted textual phenomena

Both the EU and the domestic varieties are characterized by a well-defined macro-structure. Nonetheless, these macro-structures differ in their internal organization: EU directives are divided into “*capitoli*”, “*paragrafi*”, “*articoli*” and “*lettere*”, whereas national laws present “*capì*”, “*titoli*”, “*articoli*” and “*comma*”. This difference is well-shown in the following table (table 2.20), reporting the frequency results of the previously mentioned sections in both corpora:

⁶³ After “*affinché*”, as well as after the other elements mentioned in figure 41, it is mandatory to have the subjunctive form of a verb. Therefore, the absence of a specific tag for the subjunctive does not affect the research.

Table 18. Nomenclature of text sections

Item	Corpus A		LL	Corpus B	
	RF	NF/pmw		RF	NF/pmw
<i>articolo/i</i>	21,979	15,273	+2,065.79	19,309	9,758
<i>capitol/i</i> ⁶⁶ (as internal text section)	191	133	+212.95	22	11
<i>capo/i</i>	935	650	-0.19	1,310	662
<i>comma/i</i>	1,076	748	-16,034.62	20,818	10,521
<i>lettera/e</i>	2,544	1,768	-406.96	5,598	2,829
<i>paragrafo/i</i>	10,753	7,472	+14,582.03	596	301
<i>titolo/i</i>	591	411	-337.02	1,840	930

Table 2.20. The overuse of certain items in Corpus A if compared to corpus B. (Adapted from Mori, 2018c: 232).

As explained by Mori (2018c), this difference can be considered the result of the influence of the EU legal drafting process, since this specific structure is adopted within the EU and shared among the European languages.

The same research was performed on the CHEU-Lex corpus. Through the use of the simple query type integrated in the *Concordance Corpus tool*, the same elements present in Table 2.20 were extracted from the CHEU-Lex corpus and compared by means of the normalized frequency and the LL ratio.

2.5. General remarks, limitations to this research and future developments

Overall, the research method presented above mirrors the method followed in *Observing Eurolects* (Mori, 2018c), with the aim of comparing results obtained analyzing the CHEU-Lex Corpus to those obtained by Mori, with a few exceptions. Contact-induced phenomena are not analyzed in this research since the two types of Italian do not share the same background. On the one hand, texts of Corpus B (Italian laws of implementation) derive from their European Italian versions which, in turn, are translated from the English or French EU version; on the other hand, the CHEU-Lex Corpus collects laws which are translated from the Swiss laws of implementation, which are in German (see Section 2.2.). As previously mentioned, Latinisms are an exception to this (see Section 2.4.2.3.). Nonetheless, analyzing the possible differences due to the different contact scenarios could be the object of future research.

Also, limitations to the research of this thesis are due to the fact that, unlike the CHEU-Lex Corpus, Corpus B of Italian laws of implementation analyzed by Mori is not annotated, which means that some qualitative research could not be carried out, such as the representation of the impersonal subject “*si*” as well as the passive “*si*”, or the use of “*che*” (Mori, 2018: 223-231).

CHAPTER 3

3.1. Introduction

This chapter analyzes and discusses the results of the research carried out following the methodology described in Chapter 2. The structure of this chapter mirrors that of Section 2.4. (Methodology). In particular, every section is devoted to the description of a linguistic phenomenon and corresponds to sections from 2.4.2. to 2.4.7.1. Note that references to previous research led by Mori (2018c) were inserted at the beginning of every section in order to keep track of it and thus be able to carry out as clear a comparison as possible.

It is worth remembering that, in the following Sections, data are extracted from the CHEU-Lex Corpus, i.e. the corpus of Swiss laws of implementation, and are compared both to Corpus B, which is the Italian corpus of implementation laws, and Corpus A, i.e. the corpus of Italian EU directives.

3.2. Lexical analysis

The first level analyzed during this study is the lexical level. In particular, the sTTR of the CHEU-Lex corpus was obtained by means of *WordSmith Tools*⁶⁴, as this operation is not performed by SketchEngine (see Section 2.4.2.).

The following table (Table 3.1) shows the results extracted from the two sections of the CHEU-Lex corpus, i.e. *enacting terms* and *preambles*, compared to the results for Corpus B, i.e. the corpus of Italian laws of implementation.

⁶⁴ <https://www.lexically.net/wordsmith/>

	CHEU-Lex corpus	Corpus B
Enacting Terms		
Types	16,921	22,881
Standardized	34.5	34.46
Type/Token ratio		
Preambles		
Types	453	2,480
Standardized	17.48	23.22
Type/Token ratio		

Table 3.1: Standardized Type/Toke ratio of *preambles* and *enacting parts* of the CHEU-Lex Corpus compared to Corpus B.

The sTTR of the *enacting terms* of the CHEU-Lex corpus is almost the same as that of Corpus B (34.5 and 34.46, respectively). Previous results from the *Eurolect Observatory Project* highlighted a greater lexical variety in the *enacting parts* of corpus B when compared to those of corpus A (section 2.4.2.). According to Mori (2018c: 208), this difference is due to the fact that Italian of implementation laws is more likely to use synonyms, resulting in greater lexical variety, whereas Italian of EU directives is affected by literal translation from the European versions in which “[...] Eurolect lexical types tend to be more repeated”, giving rise to a less lexically varied sub-corpus.

The fact that the *enacting terms* of the CHEU-Lex corpus has a standardized Type/Token ratio almost identical to that of Corpus B would suggest that the lexicon of CHEU-Lex’s *enacting terms* is as varied as that of Corpus B. Therefore, this could imply that the use of synonymy in the Italian of Swiss laws of implementation (CHEU-Lex Corpus) is as frequent as in the Italian of implementation laws (Corpus B).

On the contrary, lexical variety of the *preambles* of the CHEU-Lex Corpus is lower than that of Corpus B (17.48 compared to 23.22), which might suggest that the lexicon in CHEU-Lex’s *preambles* is more standardized than that of *preambles* in Corpus B.

The sTTR of both *preambles* and *enacting parts* was then compared to those of Corpus A, i.e. the corpus of European directives, with the aim of observing any relevant discrepancy or similarity.

	CHEU-Lex corpus	Corpus A
Enacting Terms		
Types	16,921	18,660
Standardized	34.5	31.96
Type/Token ratio		
Preambles		
Types	453	14,717
Standardized	17.48	37.32
Type/Token ratio		

Table 3.2: Standardized Type/Toke ratio of *preambles* and *enacting parts* of the CHEU-Lex Corpus compared to Corpus A.

CHEU-Lex’s *enacting parts* have a more varied lexicon if compared to those of Corpus A (34.5 compared to 31.96), but the same does not apply for *preambles*, where the lexical variety is greater in Corpus A rather than in the CHEU-Lex Corpus (37.32 compared to 17.48, respectively).

With regard to preambles, an explanation could be that the structure of *preambles* in the CHEU-Lex corpus is rather fixed: it usually starts either with “*Il consiglio federale svizzero*” or “*L’assemblea federale della confederazione svizzera*”, followed by “*visti gli articoli*” and ends either with “*ordina*” or “*decreta*”. By means of the *Keywords and terms extraction tool* (see Section 2.4.), a frequency list⁶⁵ of single words and n-Grams of length from 3 to 6 of the sub-section *preambles* was extracted which further confirms this: in the n-Gram list, “*visti gli articoli*” occupies the 1st position, with a normalized frequency of 78.74, followed by “*Il consiglio federale svizzero*” (n° 3, NF: 58.52) and “*assemblea federale della confederazione svizzera*” (n° 12; NF: 38.3). For the single-word list, the attribute “lemma” was first selected, which resulted in “*decretare*” in position n°14 (NF: 40.43), followed by “*ordinare*” (n°23). Then, the attribute “word” was selected, and “*decreta*” appears in position n°4, followed by “*ordina*” in position n°5.

⁶⁵ The reference corpus used is Italian Web 2016 (itTenTen2016).

Here follows an example showing how a preamble is usually structured:

“Il Consiglio federale svizzero,
vista la legge federale del 16 dicembre 2005 sugli stranieri e la loro integrazione (LStrI);
visto l’articolo 119 della legge federale del 26 giugno 1998 sull’asilo (LAsi),
ordina:”

Overall, the CHEU-Lex Corpus seems to be as lexically varied as the Corpus of Italian laws of implementation (Corpus B), except for preambles, which seem to have a more fixed form in the CHEU-Lex Corpus, resulting in a rather homogeneous lexicon.

3.2.1. EU-rooted lexical features

Another category analyzed was that of Europeisms, i.e. newly-coined words that originated in the European context. It is worth remembering that, in the framework of the *Eurolect Observatory Project*, a set of selected EU-noun phrases and semantic Europeisms were chosen, and their normalized frequency was retrieved from both Corpus A and Corpus B, and then compared by means of the LL ratio and the normalized frequency. Results showed that these elements were widely used in the corpus of EU directives (Corpus A) and, to a lesser extent, in their domestic laws of implementation (Corpus B).

For the purposes of this research, the same process was adopted, and the normalized frequencies of the same set of elements were retrieved from the CHEU-Lex Corpus. The results were eventually compared to those obtained from Corpus B (see Section 2.4.2.1.).

The following table (Table 3.3) shows the results obtained from the CHEU-Lex Corpus (see Section 2.4.2.1. for the methodology adopted) and those of Corpus B:

EU-noun phrases	CHEU-Lex corpus		LL ratio	Corpus B	
	Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
<i>Cittadino/a/i/e dell'Unione</i>	0	0	-58.53	88	44
<i>Commissione (europea)</i>	8	8.51	-503.92	864	437
<i>Comunità (europea)</i>	90	95.76	-4	289	146
<i>Consiglio (dell'unione Europea)</i>	2	2.12	-657.22	1024	517
<i>Paesi Terzi</i>	56	59.58	-56.73	372	188
<i>Parlamento Europeo</i>	0	0	-266.73	401	203
<i>Stato/i Membro/i</i>	212	225.57	-495.4	2058	1040
<i>Unione Europea</i>	104	110.66	-148.15	797	403

Table 3.3: The normalized frequency of the list of EU-noun phrases in the CHEU-Lex corpus compared to Corpus B (Adapted from Mori, 2018:c).

Results in Table 3 show that there tend to be a general underuse of the EU-noun phrases in the CHEU-Lex Corpus, i.e. the corpus of Swiss laws of implementation, when compared to the domestic legal Italian used in the laws of implementation (Corpus B). Interestingly enough, “*Cittadino/a/i/e dell'Unione*” and “*Parlamento Europeo*” are not even present in the enacting parts of the CHEU-Lex Corpus. Concerning the other elements, each of them shows a normalized frequency which is significantly lower in the CHEU-Lex Corpus when compared to Corpus B.

If we consider that in her research Mori (2018c) observed an underuse of EU-noun phrases in the domestic legal Italian (Corpus B) if compared to the Italian of EU directives (Corpus A), results obtained in figure 45 show an even lower usage of these elements in the *enacting parts* of the CHEU-Lex Corpus. This is also evident in the comparison of the CHEU-Lex results with the results for Corpus A.

EU-noun phrases	CHEU-Lex corpus		LL ratio	Corpus A	
	Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
<i>Cittadino/a/i/e dell'Unione</i>	0	0	-71.08	82	57
<i>Commissione (europea)</i>	8	8.51	-4677.79	5516	3833
<i>Comunità (europea)</i>	90	95.76	-927.54	1673	1163
<i>Consiglio (dell'unione Europea)</i>	2	2.12	-2073.94	2425	1685
<i>Paesi Terzi</i>	56	59.58	-679.31	1174	816
<i>Parlamento Europeo</i>	0	0	-1389.6	1603	1114
<i>Stato/i Membro/i</i>	212	225.57	-12284.1	16275	11309
<i>Unione Europea</i>	104	110.66	-178.1	654	454

Table 3.4: The normalized frequency of the EU-noun phrases of the CHEU-Lex corpus compared to that of Corpus A (Mori, 2018c).

In table 3.4, the normalized frequencies of the items in Corpus A are higher than those in the CHEU-Lex Corpus. Also, LL ratios are all negative and considerably higher than those in Table 3.4, confirming the idea that there is a statistically significant difference in the use of the EU-noun phrases. These elements seem to be widely used in Italian of EU directives, but they do not seem to have influenced the Swiss Italian of implementation laws. However, it is worth remembering that these differences might be due to the different content of the two corpora, or even the different environments in which the two corpora were born. Indeed, on the one hand, the CHEU-Lex corpus collects the transposition laws of the bilateral agreements between the European Union and the Swiss Confederation, which aim at regulating the relations between Swiss and the EU. On the other hand, the EU directives “[...] indicate the policies that different EU Member States should put into force” (Tomatis, 2018: 28). This could explain why these elements are less present in the CHEU-Lex Corpus.

The following graph shows the distribution of the selected EU-noun phrases in all the corpora analyzed:

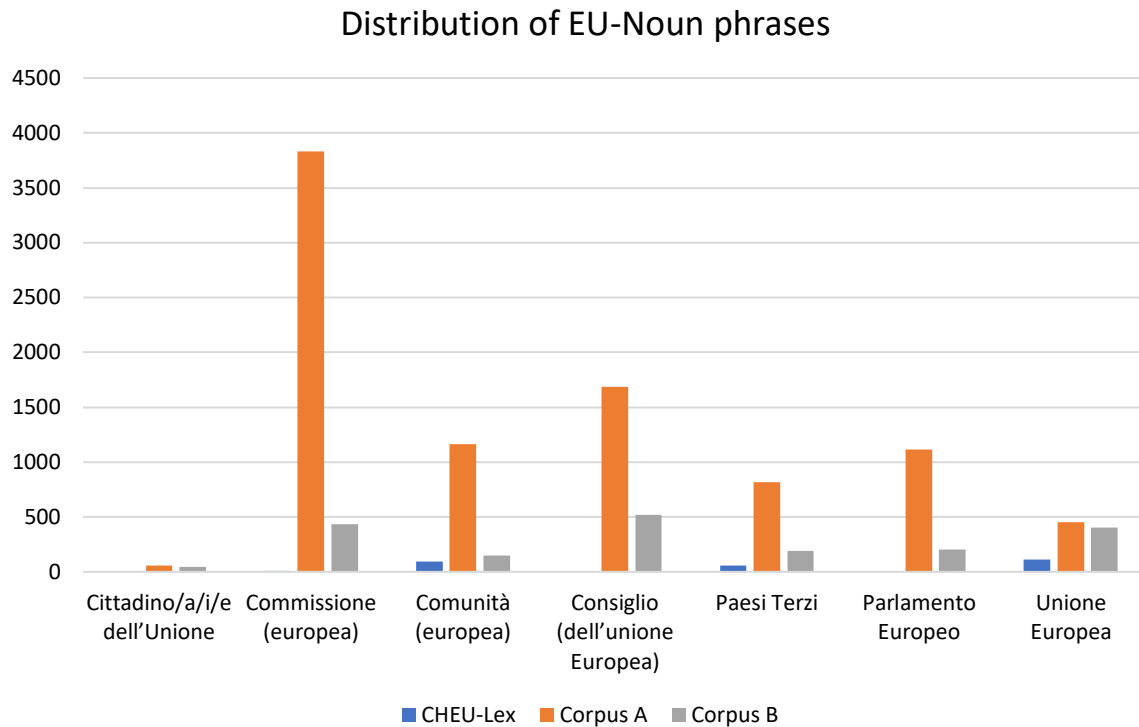


Figure 3.1: Distribution of the different EU-Noun phrases in the three corpora (normalized frequencies). Note that, due to the high normalized frequency in Corpus A of *Stato/i Membro/i* (NF: 11309), it was necessary to delete it from the graph, in order to be able to better observe the other data.

Thanks to Figure 3.1, it is possible to observe that there generally is an overuse of EU-noun phrases in the Italian of EU directives (Corpus A), whereas these elements tend to be used less in the Italian of Swiss laws of implementation (CHEU-Lex Corpus) as well as in the Italian of implementation laws (Corpus B).

Semantic Europeisms were another category analyzed, which also proved to be widely used in Corpus A (EU directives) and, to a lesser extent, in Corpus B (domestic laws of implementation). The following table shows the results obtained from the CHEU-Lex Corpus and compares them to those of Corpus B.

Semantic word classes	Semantic Europeisms	CHEU-Lex corpus		LL ratio	Corpus B	
		Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
Noun	<i>adesione</i>	11	11.7	+7.23	8	4
	<i>armonizzazio ne</i>	13	13.83	-1.74	49	25
	<i>attuazione</i>	74	78.74	-39.69	392	198
	<i>direttiva/e</i>	150	159.6	-948.9	2,609	1,318
	<i>ravvicinamen to</i>	0	0	-15.3	23	12
	<i>recepimento</i>	2	2.13	-143.5	243	123
	<i>Regolamento/ i (CE/CEE)</i>	149	158.54	-23.87	581	294
	<i>trasposizione</i>	3	3.19	+1.25	3	2
Adjective	<i>armonizzata/ e/o/i</i>	98	104.27	+32.97	110	56
	<i>attuativa/e/o/ i</i>	0	0	-87.8	132	67
	<i>comunitaria/ e/o/i</i>	2	2.13	-1123.41	1728	873
Verbs (all forms)	<i>armonizzare</i>	101	107.46	+26.41	127	64
	<i>attuare</i>	29	30.86	-20.35	168	85
	<i>recepire</i>	2	2.13	-26.09	58	29

Table 3.5: The normalized frequency of the list of semantic europeisms both in CHEU-Lex corpus and Corpus B.

Unlike EU-noun phrases, the use of semantic Europeisms varies according to the item analyzed. Almost all the elements show a significant underuse in the CHEU-Lex Corpus, with lower normalized frequencies when compared to Corpus B. Only three elements have their normalized frequencies higher in the CHEU-Lex Corpus: *adesione* (NF for the CHEU-Lex Corpus: 11.7; NF for Corpus B: 4), *armonizzata/e/o/i* (NF for the CHEU-Lex Corpus: 104.27; NF for Corpus B: 56) and *armonizzare* (NF for the CHEU-Lex Corpus: 107.46; NF for Corpus B: 64). LL ratios further confirm the overuse of these elements in the CHEU-Lex Corpus.

It is interesting to notice that the highest normalized frequencies are those of the verb *armonizzare* and its inflected forms. This overuse can be linked to the content of the CHEU-Lex corpus: on the one hand, bilateral agreements that form the CHEU-Lex Corpus aim at ensuring the intergovernmental cooperation between the Swiss Confederation and the European Union. On the other hand, *harmonization of law* is juridically defined as:

“[...] The process by which two or more states, sometimes under the auspices of an interstate or international organization, change their legislation relevant to some area of common concern to conform their statutes and to facilitate compliance and enforcement across borders.” (General Multilingual Environment Thesaurus (GEMET), 2022).

In this sense, it is possible to argue that the overuse of the verb *armonizzare* could be due to the necessity of conforming the Swiss legal system to the European one and eliminate major differences. Therefore, this hypothesis suggests that this difference might be due to the different function of the texts, and it is not strictly related to linguistic aspects.

In the following table, results from the CHEU-Lex Corpus are compared to those obtained from Corpus A.

Semantic word classes	Semantic Europeisms	CHEU-Lex corpus		LL ratio	Corpus A	
		Raw Frequency	Normalized Frequency		Raw Frequency	Normlized Frequency
Noun	<i>adesione</i>	11	11.7	-45.3	111	77
	<i>armonizzazion e</i>	13	13.83	-24.72	86	60
	<i>attuazione</i>	74	78.74	-294.8	732	509
	<i>direttiva/e</i>	150	159.6	-12809.43	16,387	11,387
	<i>ravvicinament o</i>	0	0	-57.21	66	46
	<i>recepimento</i>	2	2.13	-2.03	10	7
	<i>Regolamento/i (CE/CEE)</i>	149	158.54	-134.3	710	493
	<i>trasposizione</i>	3	3.19	-0.12	7	5
Adjective	<i>armonizzata/e/o/i</i>	98	104.27	+0.32	168	117
	<i>attuativa/e/o/i</i>	0	0	-3.47	4	3
	<i>comunitaria/e/o/i</i>	2	2.13	-932.81	1,105	768
Verbs (all forms)	<i>armonizzare</i>	101	107.46	-0.53	200	139
	<i>attuare</i>	29	30.86	-39.43	164	114
	<i>recepire</i>	2	2.13	-15.22	30	21

Table 3.6: The normalized frequencies of the semantic Europeisms in the CHEU-Lex Corpus compared to those of Corpus A (EU directives).

Overall, there is a general underuse of the semantic Europeisms in the CHEU-Lex Corpus when compared to Corpus A, implying that the influence of the EU law making process does not seem to impact the Swiss legal Italian to a great extent. “*armonizzato/a/o/i*” and “*armonizzare*” have a normalized frequency in the CHEU-Lex Corpus which is close to that

of Corpus B (Italian laws of implementation), although lower (104.27 compared to 117 for *armonizzato/a/o/i*; 107.46 compared to 139 for *armonizzare*).

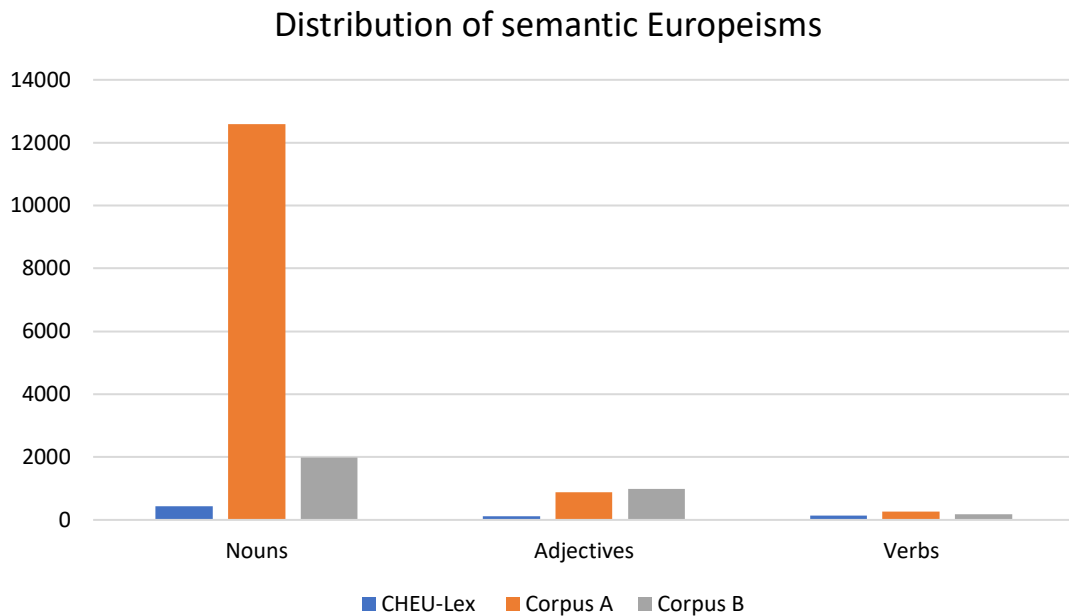


Figure 3.2: The sums of the normalized frequencies of each category (nouns, adjectives and verbs) of semantic Europeisms in CHEU-Lex Corpus, Corpus A and Corpus B.

From the graph (Figure 3.2) it is possible to observe a high overuse of “European” nouns in Corpus A, followed by Corpus B and CHEU-Lex Corpus. For what concerns “European” adjectives, there is a slight overuse in Corpus B if compared to Corpus A, whereas the use of the three verbs selected seems to be limited in all three corpora. However, what stands out is that the use of semantic Europeisms is extremely low in the CHEU-Lex Corpus.

The following chart compares the distribution of EU-noun phrases and semantic Europeisms in the three corpora:

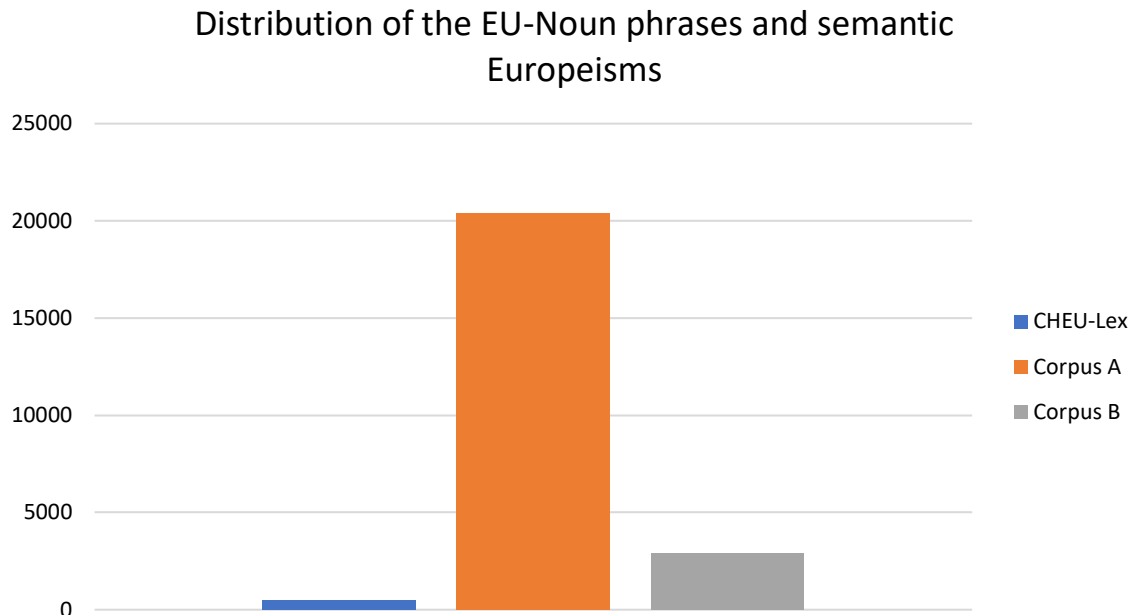


Figure 3.3: Sum of the normalized frequencies of the semantic Europeisms and the EU-Noun phrases in CHEU-Lex Corpus, Corpus A and Corpus B.

Once again, the overuse of these elements in Corpus A and the relative underuse in the CHEU-Lex Corpus is extremely evident.

3.2.2. Intra-linguistic lexical variability

This section analyzes the level of complexity of the register. As already mentioned in Section 2.4.2.2., the Italian of EU directives tends to use a less formal register, in compliance with the need for more approachable legislative texts for the European citizens. Instead, national legislative Italian tends to use a more complex lexicon. A corpus-based analysis was carried out by Mori (2018c) to observe the cross-corpora distribution of some legal verbs.

The same study was performed on the CHEU-Lex Corpus (see Section 2.4.2.2.), and results were compared to those obtained for Corpus B.

Legal verbs (inflected forms)	CHEU-Lex Corpus		LL ratio	Corpus B	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>adire</i>	24	25.54	+3.12	38	19
<i>adottare</i>	178	189.39	-825	2,541	1,284
<i>agire</i>	55	58.52	-11.8	228	115
<i>applicare</i>	846	900.15	-166.2	3,440	1,738
<i>avvalere</i>	8	8.51	-138.82	289	146
<i>depositare</i>	83	88.31	+13.64	123	62
<i>emanare</i>	435	462.84	+41.83	740	374
<i>ottemperare</i>	20	21.28	-16.94	124	63

Table 3.7: The normalized frequencies of the legal verbs for Corpus B (Mori, 2018c) compared to those for the CHEU-Lex corpus.

From the results shown in Table 3.7, it is possible to observe that there tends to be an underuse of legal verbs in the CHEU-Lex corpus compared to Corpus B. Among the eight items, only three of them have a normalized frequency higher in the CHEU-Lex Corpus when compared to Corpus B: *adire* (NFs: 25.54 and 19, respectively), *depositare* (NFs: 88.31 and 62, respectively) and *emanare* (NFs: 462.84 and 374, respectively). LL ratios further prove the overuse of these items in the CHEU-Lex Corpus if compared to Corpus B.

Legal verbs (inflected forms)	CHEU-Lex Corpus		LL ratio	Corpus A	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>adire</i>	24	25.54	+4.42	24	17
<i>adottare</i>	178	189.39	-2,906.36	4,685	3,256
<i>agire</i>	55	58.52	-4.93	143	99
<i>applicare</i>	846	900.15	-384.83	3,180	2,210
<i>avvalere</i>	8	8.51	-88.79	157	109
<i>depositare</i>	83	88.31	+6.74	104	72
<i>emanare</i>	435	462.84	+717.26	29	20
<i>ottemperare</i>	20	21.28	-17.88	95	66

Table 3.8: The normalized frequencies of legal verbs in Corpus A and CHEU-Lex corpus.

Results in Table 3.8 show quite a balanced use of legal verbs between the CHEU-Lex Corpus and Corpus A: three elements show a normalized frequency higher in the CHEU-Lex Corpus when compared to Corpus A, i.e. *adire* (NFs: 25.54 and 17, respectively), *depositare* (NFs: 88.31 and 72, respectively) and *emanare* (NFs: 462.84 and 20, respectively); the same happened when comparing the CHEU-Lex Corpus to Corpus B. The other verbs have a normalized frequency lower in the CHEU-Lex Corpus when compared to Corpus B.

Adottare and *applicare*, which are considered ordinary verbs (Mori, 2018c: 216), are observed to be overused in Corpus A compared to Corpus B. The same happens when comparing Corpus A to the CHEU-Lex Corpus: their normalized frequencies in the CHEU-Lex Corpus (189.39 and 900.15, respectively) are lower when compared to those in Corpus A (3,256 and 2,210, respectively), thus confirming the underuse of these elements in CHEU-Lex corpus. This might indicate the tendency to use complex structures and sentences: this feature could suggest that the European legal drafting process, which requires the use of simple texts, did not significantly affect the Italian of Swiss laws of implementation.

The following chart (Figure 3.4) shows the distribution of legal verbs in all the corpora under observation, and gives a visual representation of what previously said:

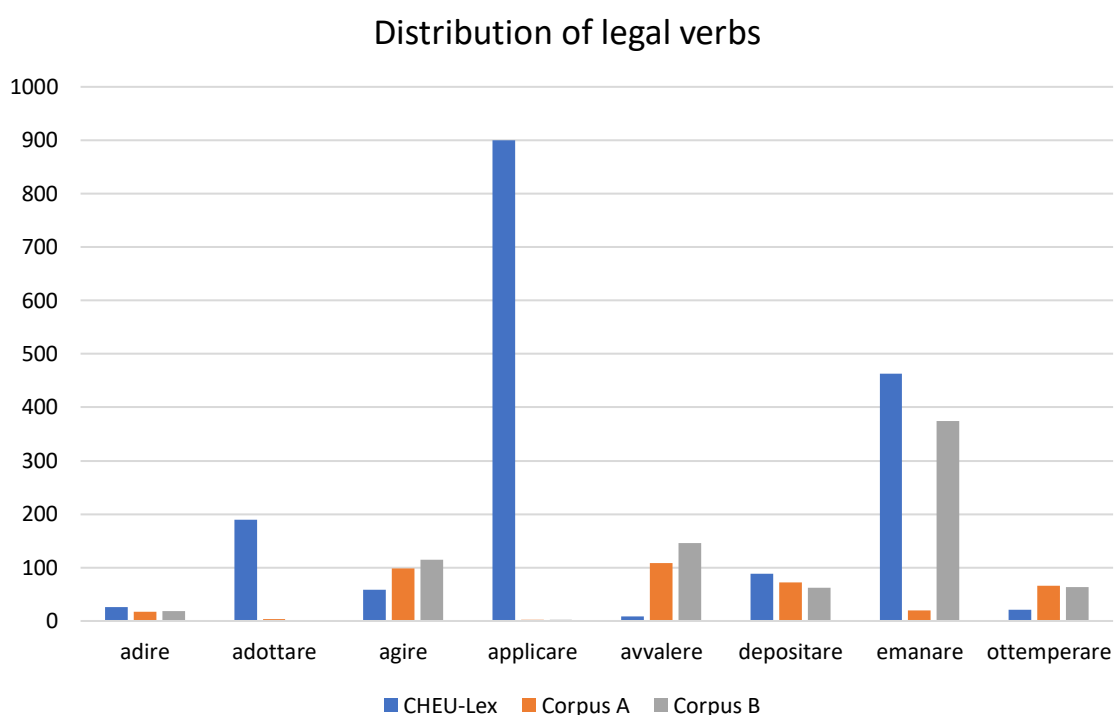


Figure 3.4: The single normalized frequencies of legal verbs in CHEU-Lex Corpus, Corpus A and Corpus B.

The next table (table 3.9) compares the use of some register markers as well as some intra-genre variants. During the *Eurolect Observatory Project*, results showed that the “[...] conservative character of national legislative Italian reveals itself mainly in register-related variability [...]” (Mori, 2018c: 215). In other words, higher register variants were observed to be widely more used in Corpus B (the corpus of Italian laws of implementation) rather than Corpus A (the corpus of EU directives).

The same research was conducted on the CHEU-Lex corpus, with the aim of observing whether any significant difference between the CHEU-Lex Corpus and Corpus A and Corpus B in the use of the selected items could be observed.

Results are shown in the table below (table 3.9).

Item	CHEU-Lex corpus		LL Ratio	Corpus B	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>altresì</i>	22	23.41	-324.24	700	354
<i>apposito</i>	17	18.09	-85.36	255	129
<i>attraverso</i>	64	68.1	-113.84	543	274
<i>a titolo di</i>	34	36.18	-0.24	95	48
<i>concernente/i</i>	223	237.27	+1.17	518	262
<i>inerente/i</i>	120	127.68	+17.21	185	93
<i>di concerto</i>	2	2.13	-464.33	732	370
<i>in materia di</i>	480	510.72	-17.07	1,505	761
<i>in quanto</i>	112	119.17	+1.38	248	125
<i>in sede di</i>	4	4.26	-153.67	279	141
<i>in tema di</i>	0	0	-49.89	75	38
<i>mediante</i>	448	476.678	-0.01	1,142	577
<i>minore/i</i> ⁶⁶	/	/	/	203	103
<i>ovvero</i>	16	17.02	-1146.46	1,942	981
<i>per mezzo di/del/della/dei/delle/</i>	38	40.43	+18.16	35	18
<i>presso</i>	333	354.31	-15.51	1,075	543
<i>riguardante/i</i>	90	95.76	-9.42	325	164
<i>se del caso</i>	57	60.65	-6.17	207	105
<i>tramite</i>	80	85.12	-69.58	501	253

Table 3.9: Register markers and intra-genre variants in Corpus B compared to those in the CHEU-Lex corpus.

Like the results obtained by Mori, where the overuse of certain elements in Corpus B was extremely evident, results in Table 3.9 show a noticeable underuse in the distribution of the selected items in the CHEU-Lex Corpus if compared to Corpus B. Among the 19 items, only two of them show a normalized frequency higher in the CHEU-Lex Corpus when compared to Corpus B: *inerente/i* (NFs: 127.68 and 93, respectively) and *per mezzo di/del/della/dei/delle/* (NFs: 40.43 and 18, respectively). The other items are significantly

⁶⁶ The item “minore/i” could not be analyzed since during the POS tagging phase (see Section 2.3.2.5) an error occurred and no distinction was made between the noun and the adjective. Therefore, all the occurrences of “minore” and “minori” were tagged as adjectives.

underused in the Italian of Swiss laws of implementation (CHEU-Lex Corpus) when compared to the Italian of implementation laws (Corpus B).

The same comparison was carried out between the CHEU-Lex Corpus and Corpus A:

Item	CHEU-Lex corpus		LL Ratio	Corpus A	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>altresì</i>	22	23.41	-72.62	195	136
<i>apposito</i>	17	18.09	+4.98	14	10
<i>attraverso</i>	64	68.1	-8.71	179	124
<i>a titolo di</i>	34	36.18	+14.11	23	16
<i>concernente/i</i>	223	237.27	-1.06	447	311
<i>inerente/i</i>	120	127.68	+58.07	73	51
<i>di concerto</i>	2	2.13	-9.48	22	15
<i>in materia di</i>	480	510.72	+1.57	823	572
<i>in quanto</i>	112	119.17	+12.61	130	90
<i>in sede di</i>	4	4.26	-18.27	43	30
<i>in tema di</i>	0	0	-10.4	12	8
<i>mediante</i>	448	476.678	+29.54	585	407
<i>minore/i</i>	/	/	/	95	66
<i>ovvero</i>	16	17.02	-211.62	358	249
<i>per mezzo di/del/della/dei/delle/</i>	38	40.43	+17.46	24	17
<i>presso</i>	333	354.31	+67.76	321	223
<i>riguardante/i</i>	90	95.76	-64.7	394	274
<i>se del caso</i>	57	60.65	-248.23	595	413
<i>tramite</i>	80	85.12	-15.63	241	167

Table 3.10: The normalized frequencies and the LL ratios of the intra-genre variants and the register markers in Corpus A and in the CHEU-Lex Corpus.

Unlike the comparison with Corpus B, when comparing the CHEU-Lex Corpus with Corpus A, there is a sort of balance in the use of the register markers chosen, with seven items showing a normalized frequency higher in the CHEU-Lex Corpus when compared to Corpus A: *apposito* (NFs: 18.09 and 10, respectively), *a titolo di* (NFs: 36.18 and 16, respectively), *inerente/i* (NFs: 127.68 and 51, respectively), *in quanto* (NFs: 119.17 and 90, respectively), *mediante* (NFs: 476.678 and 407, respectively), *per mezzo di/del/della/dei/delle/* (NFs: 40.43 and 17, respectively) and *presso* (NFs: 354.31 and 223, respectively).

Overall, considering that according to Mori (2018c) higher register variants were observed to be widely more used in Corpus B (the corpus of Italian laws of implementation) rather

than Corpus A (the corpus of EU directives), the use of register markers in the CHEU-Lex Corpus is closer to Corpus A, as shown in the following chart (figure 3.5):

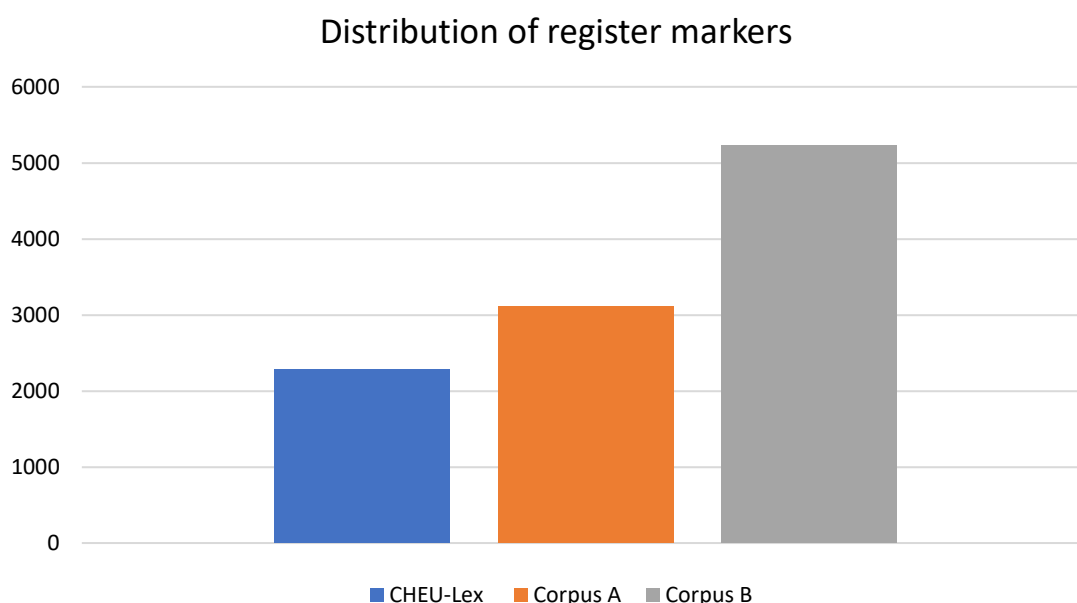


Figure 3.5: Sum of the normalized frequencies of the register markers selected in the CHEU-Lex Corpus, Corpus A and Corpus B.

The graph shows that Corpus B has a higher normalized frequency of the item selected when compared to Corpus A and the CHEU-Lex Corpus. As already shown by Mori (2018c), the Italian variety used in the European directives (Corpus A) tends to remain as simple as possible, maintaining a simple structure as well as using more common lexical variants. Considering this, it seems as though the CHEU-Lex Corpus tends to use an even simpler structure and register.

3.2.3. Latinisms and Eurolect renderings

Latinisms and Eurolect renderings, i.e. contact-induced lexical items, were the last lexical items analyzed. As already explained, Latinisms belong to the “contact-induced phenomena” group. It is worth remembering that, due to the different languages each variety of Italian got in contact with, contact-induced phenomena were not considered in this research (Italian of Swiss laws of implementation is translated from the Swiss legal German, whereas the Italian of implementation laws used during the implementation procedure is “translated” from the Italian of EU directives, i.e. the variety of Italian born within the European Context. For this reason, it would have been difficult to compare the two varieties of Italian).

Nonetheless, Latinisms constitute an exception to this since they are shared among the European languages due to their historical background (see Section 2.4.1.) and were therefore worth analyzing. Eurolect renderings were considered as well in this section. Results are shown in the following table.

Item	CHEU-Lex Corpus		LL ratio	Corpus B	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>ad hoc</i>	3	3.19	+1.25	3	2
<i>acquis</i>	4	4.26	+10.1	0	0
<i>autorità competente/i</i>	394	419.22	-100.12	1,698	858
<i>conformemente</i>	789	839.5	+509.15	583	295
<i>mutatis mutandis</i>	0	0	0	0	0
<i>post mortem</i>	0	0	-138.35	208	105
<i>status</i>	0	0	-46.56	70	35
<i>regolamentare/i</i>	73	77.67	-0.75	208	205
<i>regolamentazione</i>	37	39.37	-1	113	57
<i>regolazione</i>	0	0	-31.93	48	Not given⁶⁷

Table 3.11: Use of some Latinisms and Eurolect renderings both in the CHEU-Lex Corpus and in Corpus B.

It is interesting to notice that four elements out of ten do not appear in the *enacting parts* of the CHEU-Lex Corpus: *mutatis mutandis*, *post-mortem*, *status* and *regolazione*. Among the others, *conformemente* stands out, with a normalized frequency in the CHEU-Lex Corpus of 839.5 (compared to 295 in Corpus B), followed by *autorità competente/i* (NF:419.22). These two items were observed to be widely used in Corpus A (Mori, 2018c: 213), these being Eurolect renderings. The fact that they appear to be widely used in CHEU-Lex Corpus could be considered as a clue of the influence of the European legal drafting process on the Italian of Swiss laws of implementation. However, there are other factors that must be taken into account. Indeed, the fact that *conformemente* is the only item having a normalized frequency higher in the CHEU-Lex Corpus than in Corpus B could reveal that the use of *conformemente* does not actually come from the Eurolect influence (since, if this were the case, the influence would probably be evident with respect to other items as well). Therefore, it could be used widely in Swiss legal Italian for reasons other than European contact, such as, for instance, the fact that the Swiss Confederation has three official languages, among which there is French. Considering that *conformemente* is probably a calque from the French

⁶⁷ Only attested in corpus B with a raw frequency of 48 (Mori, 2018c: 213).

conformément (Mori, 2018c), this could be interpreted as a contact-induced phenomenon between Swiss legal French and Swiss legal Italian. The fact that the reception of the European directives in Switzerland is done starting from the European German version (and not from the French version), further supports this thesis. This would disprove the idea that *conformemente* is a calque from the French Eurolect. To check this hypothesis, the first attempt to be done is to check the German and the French renderings of *conformément* in the CHEU-Lex Corpus: by means of the *Parallel Concordance Tool* available on *NoSketchEngine*⁶⁸, it is possible to search the parallel concordances of *conformemente* both in the French and in the German CHEU-Lex sub-corpora. Results show that, for what concerns the German version, *gemäß* is the most used rendering in parallel with *conformemente* (NF:103.84), followed by *in Übereinstimmung mit* (NF: 8.51). If, instead, the French rendering is observed, *conformément* occurs in parallel with *conformemente* with a normalized frequency of 240.6, higher than that of *gemäß*. This could be a further clue of the possible origin as a French calque of the word *conformemente*. However, these results do not show whether it is a calque from the French Eurolect or, instead, from the Swiss legal French. This point could be further discussed when comparing the Italian Swiss corpus of implementation laws to a corpus of Italian laws unrelated to the European context.

The next table compares the results of the CHEU-Lex Corpus with those of Corpus A.

Item	CHEU-Lex Corpus		LL ratio	Corpus A	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>ad hoc</i>	3	3.19	-3.57	16	11
<i>acquis</i>	4	4.26	-26.78	55	38
<i>autorità</i>	394	419.22	-1319.28	3521	2447
<i>competente/i</i>					
<i>conformemente</i>	789	839.5	+5.14	1604	1115
<i>mutatis mutandis</i>	0	0	-36.41	42	29
<i>post mortem</i>	0	0	-9.54	11	8
<i>status</i>	0	0	-703.9	812	564
<i>regolamentare/i</i>	73	77.67	-211.87	602	418
<i>regolamentazione</i>	37	39.37	-17.6	141	98

Table 3.12: The usage of some Latinisms and Eurolect renderings both in the CHEU-Lex Corpus and Corpus A.

⁶⁸ <http://corpora.fti.unige.ch/crystal/#open>

Unlike the comparison with Corpus B, the normalized frequencies in the CHEU-Lex Corpus are all lower than those of Corpus A. Also, *conformemente* has a lower normalized frequency in the CHEU-Lex Corpus when compared to that of Corpus A (839.5 compared to 1115, respectively). However, its normalized frequency is the highest among those of the CHEU-Lex Corpus. To understand if the word *conformemente* comes from the contact with the Swiss legal French or if, instead, it results from the influence of the European legal drafting process, it is necessary to compare the CHEU-Lex Corpus to a corpus representing the Swiss legal Italian not related to the EU environment, i.e. laws which are born in the Swiss Confederation.

The following chart (Figure 3.6) shows the distribution of the Latinisms and the Eurolect renderings chosen in the three corpora.

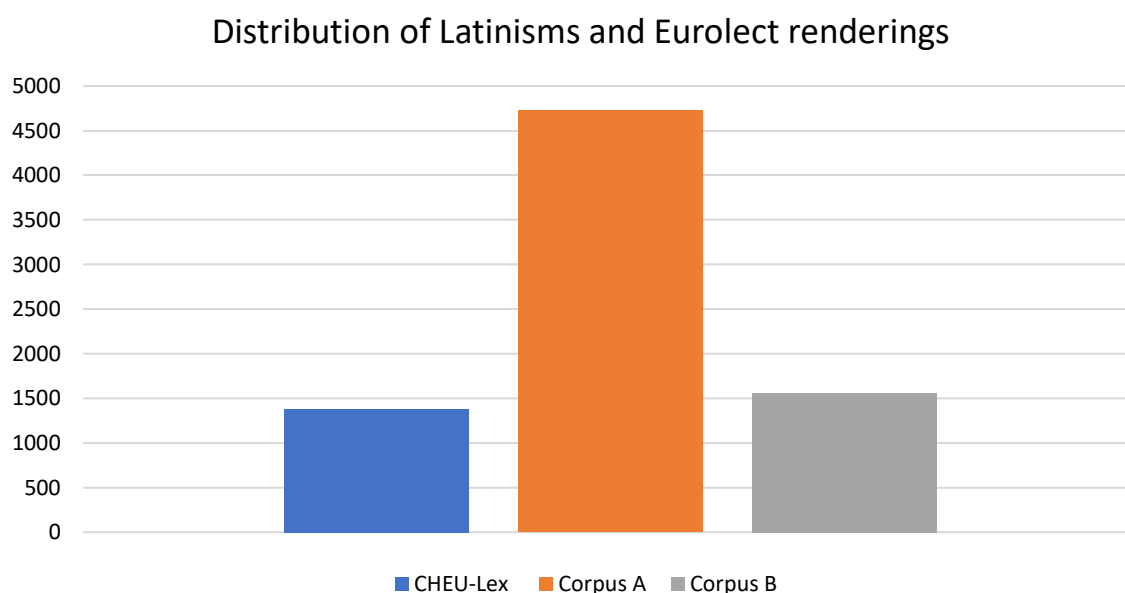


Figure 3.6: The sums of the normalized frequencies of Latinisms and Eurolect renderings in all three corpora.

From the graph above (Figure 3.6), it is possible to see how, overall, Latinisms and European renderings are more present in Corpus A, and to a lesser extent in the two other corpora which are almost at the same level.

3.3. Lexical morphology

Lexical morphology analyzes the use of certain international prefixes in the word-formation process (see Section 2.4.3.). As stated by Mori, “the overuse of international prefixes is

connected with their semantic cross-language transparency; therefore, they were described as “[...] cues of the Europeanisation process in some legal languages at the EU level” (2018c: 217).

Item	CHEU-Lex Corpus		LL ratio	Corpus B	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>anti + N/Adj</i>	160	170.24	+56.05	176	89
<i>bio</i>	187	198.97	+84.05	179	90
<i>eco</i>	501	533.07	+811.43	93	47
<i>extra + Adj.</i>	15	15.96	-33.83	143	72
<i>intra + Adj.</i>	13	13.83	-13.06	86	43
<i>multi + Adj.</i>	3	3.19	-54.27	112	57
<i>post + N/Adj.</i>	23	24.47	+21.06	12	6
<i>semi + N/Adj.</i>	7	7.45	-1.5	29	15
<i>sub + N/Adj.</i>	38	40.43	-7.8	156	79

Table 3.13: Use of some international prefixes in the CHEU-Lex Corpus and in Corpus B.

Results in Table 3.13 show a general overuse of these elements in the CHEU-Lex Corpus if compared to Corpus B. In particular, *eco* (NFs: 533.07 and 47, respectively), *bio* (NFs: 198.97 and 90, respectively), *anti* (NFs: 170.24 and 89, respectively) and *post* (NFs: 24.47 and 6, respectively) stand out.

Item	CHEU-Lex Corpus		LL ratio	Corpus A	
	Raw Frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>anti + N/Adj</i>	160	170.24	-1.62	333	231
<i>bio</i>	187	198.97	+110.2	97	67
<i>eco</i>	501	533.07	+676.18	72	50
<i>extra + Adj.</i>	15	15.96	-0.13	31	22
<i>intra + Adj.</i>	13	13.83	-30.72	96	67
<i>multi + Adj.</i>	3	3.19	-82.9	121	84
<i>post + N/Adj.</i>	23	24.47	+8.25	17	12
<i>semi + N/Adj.</i>	7	7.45	-0.11	15	10
<i>sub + N/Adj.</i>	38	40.43	+29.26	15	10

Table 3.14: The normalized frequencies of the international prefixes in the CHEU-Lex Corpus and in Corpus A.

The chart in Table 3.14 shows that the CHEU-Lex Corpus uses some Latin and Greek prefixes to a greater extent than Corpus A. In particular, *bio* (NFs: 198.97 and 67, respectively), *eco* (NFs: 533.07 and 50, respectively), *post + N/Adj* (NFs: 24.47 and 12,

respectively) and *sub + N/Adj* (NFs: 40.43 and 10, respectively) are the items which are more widely used in the CHEU-Lex Corpus when compared to Corpus A.

Overall, the results show that in general all three legislative varieties seem to be characterized both by Latin and Greek prefixes. In particular, it seems as though Greek prefixes (*bio*, *eco*, and to a certain extent also *anti*⁶⁹) characterize better the CHEU-Lex Corpus if compared to Corpus A and Corpus B.

The following chart compares the distribution of each element in the three corpora. It can be observed how *eco* and *bio* are more widely used in the CHEU-Lex Corpus if compared to the other prefixes.

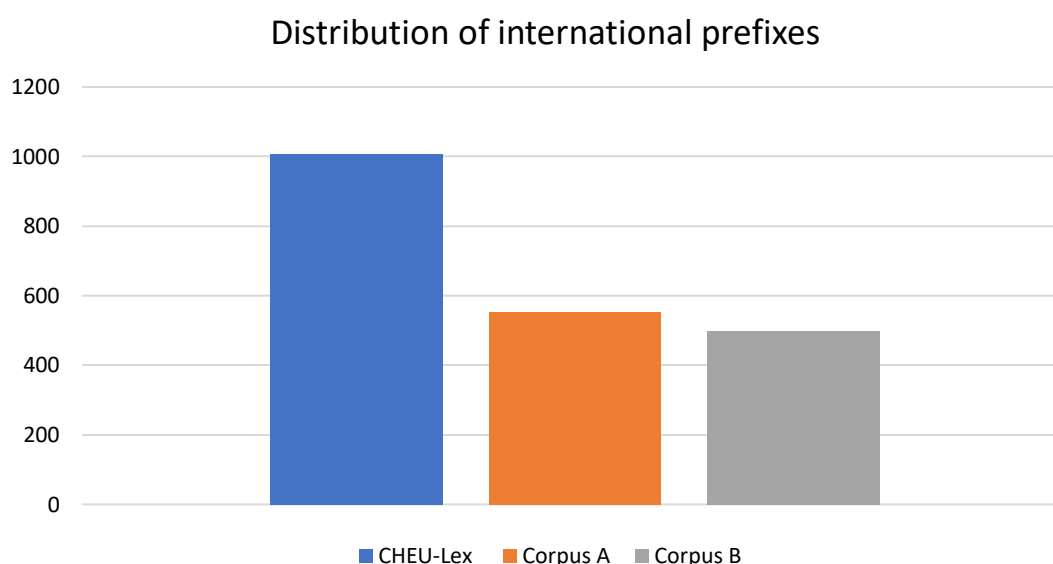


Figure 3.7: The distribution of the international prefixes in each corpus.

However, since no historical evidence⁷⁰ is given to explain this overuse, it could be interpreted as the result of the influence of the European legal drafting process on the Italian or Swiss laws of implementation. Indeed, as happens for Corpus B when compared to Corpus A, this could “[...] mirror a word-formation option extremely productive in technical and scientific varieties of Italian” (Mori, 2018c: 217). To confirm this hypothesis, it would be

⁶⁹ *Anti* can either be Greek or Latin, according to the meaning.

⁷⁰ Think, for example, to the Romansh, one of the descendant languages of the spoken Latin language and one of the official national languages of the Swiss Confederation: this could explain the widespread presence of Latin suffixes, but it is hard to link Greek prefixes to this. Besides, considering that *bio-* and *eco-* are extremely productive in different languages, it is more likely that their use is the outcome of contact situations.

necessary to carry out a further analysis, comparing the CHEU-Lex Corpus to a corpus of Swiss national Italian legislation: if Greek prefixes are observed to be widely present in the latter, then results might be interpreted as the evidence of a typical national trait in the Swiss legislation. If, instead, no evidence is found of a remarkable use of these elements in the Swiss legal Italian variety, then results might be interpreted as clues of a possible European influence.

3.4. Verb morphology

Verb morphology was another interesting feature analyzed in the frame of the *Eurolect Observatory Project*. A cross-corpora comparison of word lists of the enacting parts of both corpora was carried out on verbs in the present indicative (Mori, 2018c).

As already explained in Section 2.4.4., the same analysis was carried out, and a word list of the enacting parts of the CHEU-Lex Corpus was compared to that of Corpus A and Corpus B.

CHEU-Lex Corpus	Wordlist ranking	Corpus B	Wordlist ranking	Corpus A	Wordlist ranking
<i>è</i>	n° 17	<i>è</i>	n° 24	<i>è</i>	n° 29
<i>sono</i>	n° 27	<i>sono</i>	n° 31	<i>sono</i>	n° 34
<i>può</i>	n° 31	<i>può</i>	n° 67	<i>possono</i>	n° 57
<i>deve</i>	n° 53	<i>sia</i>	n° 79	<i>sia</i>	n° 66
<i>possono</i>	n° 54	<i>possono</i>	n° 80	<i>siano</i>	n° 69
<i>ha</i>	n° 67	<i>deve</i>	n° 120	<i>può</i>	n° 73
<i>devono</i>	n° 75	<i>siano</i>	n° 132	<i>adottano</i>	n° 122
<i>sia</i>	n° 134	<i>devono</i>	n° 137	<i>ha</i>	n° 126
<i>hanno</i>	n° 161	<i>ha</i>	n° 164	<i>applicano</i>	n° 163
<i>modifica</i>	n° 171	<i>applica</i>	n° 209	<i>applica</i>	n° 197
<i>siano</i>	n° 221	<i>provvede</i>	n° 285	<i>provvedono</i>	n° 211
<i>applica</i>	n° 249	<i>hanno</i>	n° 289	<i>hanno</i>	n° 201
<i>viene</i>	n° 519	<i>modifica</i>	n° 467	<i>comunicano</i>	n° 214
<i>intende</i>	n° 576	<i>intende</i>	n° 533	<i>entra</i>	n° 284
<i>provvede</i>	n° 988	<i>abbiano</i>	n° 625	<i>contengono</i>	n° 290
<i>sarà</i>	Not available in the first 1000 results	<i>sarà</i>	n° 703	<i>informano</i>	n° 304
<i>abbiano</i>	Not available in the first 1000 results	<i>viene</i>	n° 706	<i>prescrivono</i>	n° 653

<i>provvedano</i>	Not available in the first 1000 results	<i>provvedono</i>	n°712	<i>trattino</i>	n° 664
-------------------	---	-------------------	-------	-----------------	--------

Table 3.15: The table compares the ranking of the verbs taken from the word lists of Corpus A and Corpus B to those of the CHEU-Lex Corpus.

Interestingly enough, as happens for both Corpus A and Corpus B, the first verbs to appear in the word list are *è* and *sono*. However, the high ranking of these verbs might be due to various factors, such as the wide use of passive forms. To retrieve more information on these elements, it would be necessary to compare the CHEU-Lex Corpus to an annotated corpus; indeed, it is worth remembering that, unlike CHEU-Lex Corpus, Corpus A and Corpus B are not annotated. Beside these items, the vast majority of verbs in the word list of the CHEU-Lex Corpus happen to be 3rd person singular and plural forms in present indicative, followed by the subjunctive mood, as shown in Table 3.15. The only difference that can be observed is that, beside the subjunctive of the verb “to be”, i.e. *sia* (n°134) and *siano* (n°221), there are no other subjunctive forms available among the first 1000 results of the CHEU-Lex Corpus. Corpus A and Corpus B, instead, both contain the subjunctive of the verb “to be” in a higher position and include other subjunctives, such as *abbiano* and *trattino* (n° 625 in corpus A and n° 664 in corpus B, respectively). This could be read as, once again, the need for a language as simple as possible, avoiding complex verb forms.

È necessario and *occorre* were analyzed to observe the use of impersonal structures. During the *Eurolect Observatory Project*, it was observed that *è necessario* is used in Corpus A almost ten times more when compared to Corpus B, whereas the use of the verb *occorre* is very limited in both corpora (Mori, 2018c).

Occurrences of these two items were extracted from the CHEU-Lex Corpus (see Section 2.4.4.), and their normalized frequency was compared to that of the same elements extracted from Corpus A and Corpus B.

Results are shown in Tables 3.16 and 3.17.

CHEU-Lex Corpus	Normalized Frequency	Corpus B	Normalized frequency
<i>è necessario</i>	126.62	<i>è necessario</i>	42
<i>occorre</i>	125.55	<i>occorre</i>	20

Table 3.16: The use of *è necessario* and *occorre* between the CHEU-Lex Corpus and Corpus B. LL ratios are not retrieved here since the raw frequencies of the elements in Corpus B were not available.

CHEU-Lex Corpus	Normalized Frequency	Corpus A	Normalized frequency
<i>è necessario</i>	126.62	<i>è necessario</i>	561
<i>occorre</i>	125.55	<i>occorre</i>	33

Table 3.17: The use of *è necessario* and *occorre* between the CHEU-Lex Corpus and Corpus A. LL ratios are not retrieved here since the raw frequencies of the elements in Corpus A were not available.

On the one hand, *è necessario* is used almost five times less in the CHEU-Lex Corpus (NF: 126.62) if compared to Corpus A (NF: 561). On the other, *occorre* is widely more used in the CHEU-Lex Corpus (NF: 125.55) if compared to Corpus A (NF: 33) and Corpus B (NF: 20).

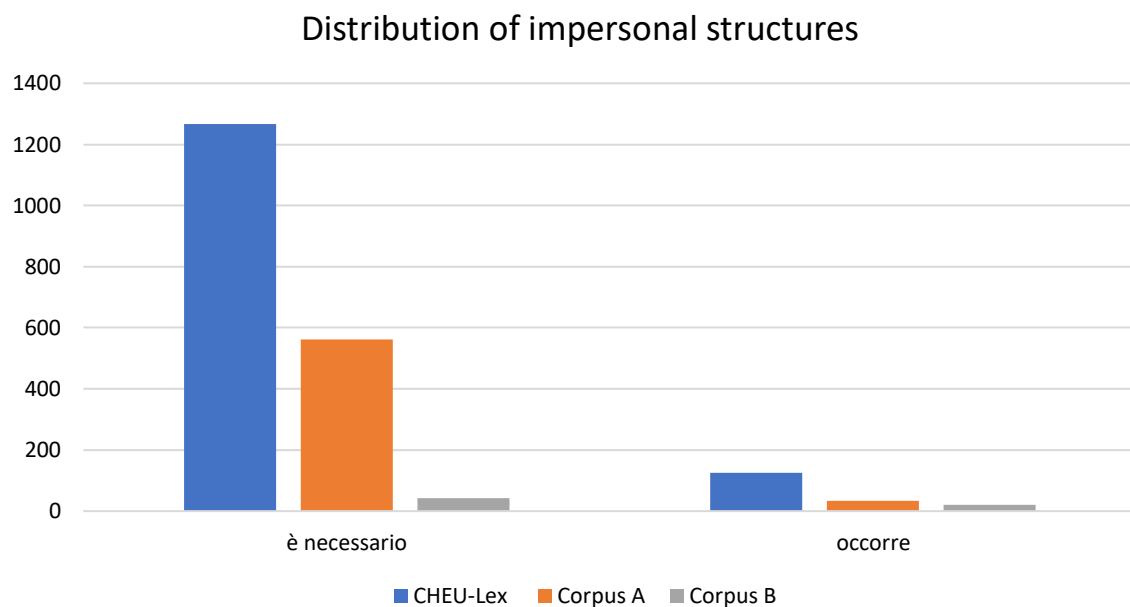


Figure 3.8: The distribution of *è necessario* and *occorre* in all three corpora.

Other modal verbs that were observed are the inflected forms of the verbs *dovere* and *potere*, i.e. *deve*, *devono*, *può* and *possono*. Mori explains that, according to the styles guide, the

obligation expressed with the verb *dovere* should be avoided both at the EU and at national level. Nonetheless, the latter is widely used in both corpora (Mori, 2018c: 223).

The same items were extracted from the CHEU-Lex Corpus and compared both to Corpus B and Corpus A. Results are shown below.

ITEM	CHEU-Lex Corpus		LL RATIO	Corpus B	
	Raw frequency	Normalized Frequency		Raw frequency	Normalized frequency
<i>deve</i>	1510	1,606.64	+347.36	1,997	1,009
<i>devono</i>	1056	1,123.59	+85.16	1,863	941
<i>può</i>	3093	3,290.96	+2023.54	3,500	1,769
<i>possono</i>	1,464	1,557.7	+93.45	2,696	1,362

Table 3.18: The normalized frequency of the items in the CHEU-Lex Corpus and in Corpus B.

ITEM	CHEU-Lex Corpus		LL RATIO	Corpus A	
	Raw frequency	Normalized Frequency		Raw frequency	Normalized frequency
<i>deve</i>	1510	1,606.64	+483.79	1188	826
<i>devono</i>	1056	1,123.59	+154.45	1139	791
<i>può</i>	3093	3,290.96	+907.45	2545	1,769
<i>possono</i>	1,464	1,557.7	-101.97	3664	2,546

Table 3.19: The normalized frequency of the items in the CHEU-Lex Corpus and in Corpus A.

Surprisingly enough, these elements are widely more present in the CHEU-Lex Corpus compared both to Corpus A and Corpus B. In particular, in Corpus A these items seem to be way less used than in the other corpora, with the exception of *possono* which has a normalized frequency of 2,546 in Corpus A, compared to 1,362 in Corpus B and 1,557.7 in the CHEU-Lex Corpus. The overuse of these elements is also confirmed by the wordlists (table 3.15): *può* (n° 31), *deve* (n° 53), *possono* (n° 54) and *devono* (n° 75) in CHEU-Lex Corpus and are attested in Corpus B as well (*può* in position n°67, *possono* in position n°80, *deve* in position n°120 and *devono* in position n°137). In Corpus A, *possono* appears in position n°57 and *può* in position n°63, whereas *deve* and *devono* are not attested in the first 1000 results. However, it would be necessary to check whether the same style guides are applied in the Swiss confederation as well (i.e. to avoid the verb *dovere*), although results show the opposite.

The following chart shows the distribution of modal verbs *dovere* and *potere* in all the corpora analyzed.

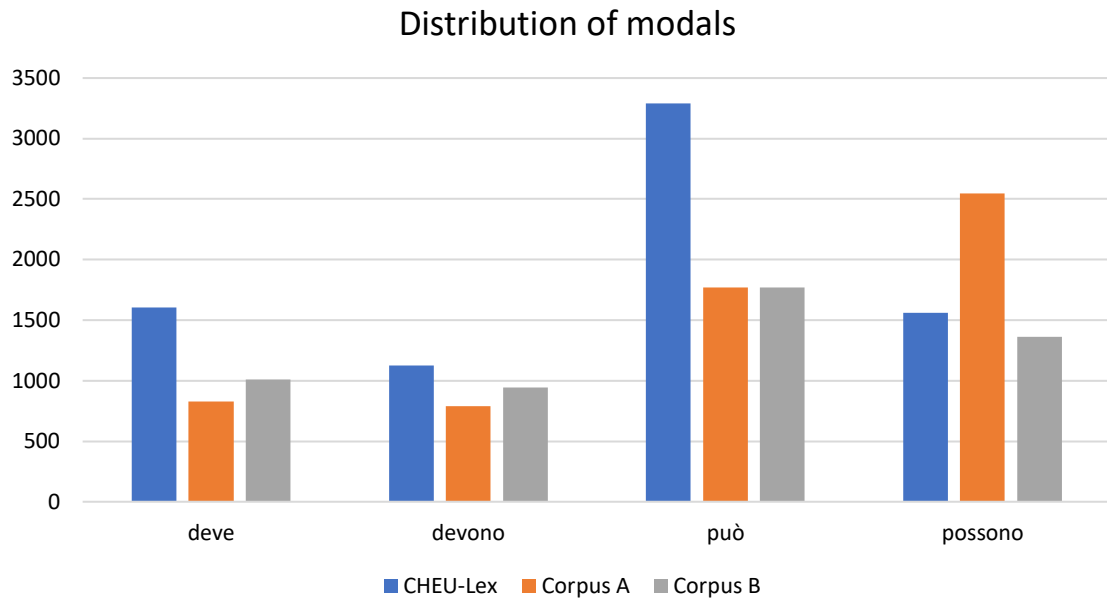


Figure 3.9: The normalized frequencies of the items in the three corpora.

From the chart above (Figure 3.9), it is possible to observe that the use of modals is more widespread in the CHEU-Lex Corpus, with the exception of *possono*, which has a higher normalized frequency in Corpus A.

The last analysis concerning the verb morphology observed the use of passive forms. Previous results obtained from Corpus A (Italian of EU directives) and Corpus B (domestic legal Italian of implementation laws), showed “[...] a preference for passive in Italian of EU directives” (Mori, 2018c: 224) if compared to the domestic variety (see Section 2.4.4).

The same research was performed on the CHEU-Lex Corpus and results are reported in the following chart (Table 3.20):

ITEM	CHEU-lex Corpus		LL RATIO	Corpus B	
	Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
<i>è stato/a</i> + <i>PP</i>	406	431.99	+218.53	345	174
<i>sono</i> <i>state/i</i> + <i>PP</i>	203	215.99	+35.16	296	150
<i>viene</i> + <i>PP</i>	151	160.66	+11.58	269	136
<i>vengono</i> + <i>PP</i>	108	114.91	+11.11	181	91
<i>va</i> + <i>PP</i>	60	63.84	+28.88	55	28
<i>vanno</i> + <i>PP</i>	50	53.2	+54	21	11
<i>deve</i> <i>essere</i> + <i>PP</i>	289	307.5	+4.87	625	316
<i>devono</i> <i>essere</i> + <i>PP</i>	504	536.26	+102.06	698	353

Table 3.20: The normalized frequencies of the passive forms in the CHEU-Lex Corpus and those obtained for Corpus B.

Statistics in Table 3.20 show that there tends to be a general overuse of the passive forms in the CHEU-Lex Corpus if compared to Corpus B. Indeed, the normalized frequencies are overall higher for the CHEU-Lex Corpus, underlying the fact that the passive form is generally more widespread in the CHEU-Lex Corpus than in Corpus B.

The same happens when comparing the CHEU-Lex Corpus to Corpus A, although to a lesser extent:

ITEM	CHEU-lex Corpus		LL RATIO	Corpus A	
	Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
<i>è stato/a + PP</i>	406	431.99	+62.51	431	237
<i>sono state/i + PP</i>	203	215.99	+24.25	232	224
<i>viene + PP</i>	151	160.66	-1.59	315	219
<i>vengono + PP</i>	108	114.91	-0.95	223	155
<i>va + PP</i>	60	63.84	+54.7	19	13
<i>vanno + PP</i>	50	53.2	+14.77	41	28
<i>deve essere + PP</i>	289	307.5	+58.58	279	194
<i>devono essere + PP</i>	504	536.26	+155.02	405	281

Table 3.21: The normalized frequencies of the passive forms in the CHEU-Lex Corpus and in Corpus A.

Unlike the comparison with Corpus B, some items have a lower normalized frequency in the CHEU-Lex Corpus if compared to those of Corpus A: *sono state/i + PP* (NF: 215.99 in the CHEU-Lex Corpus and 224 in Corpus A), *viene + PP* (NF: 160.66 in the CHEU-Lex Corpus and 219 in Corpus A) and *vengono + PP* (NF: 114.91 in CHEU-Lex Corpus and 155 in Corpus A).

Overall, there tend to be an overuse of passive form in the Swiss Italian of implementation laws if compared both to Italian of EU directives and the Italian of laws of implementation, as shown in the following chart.

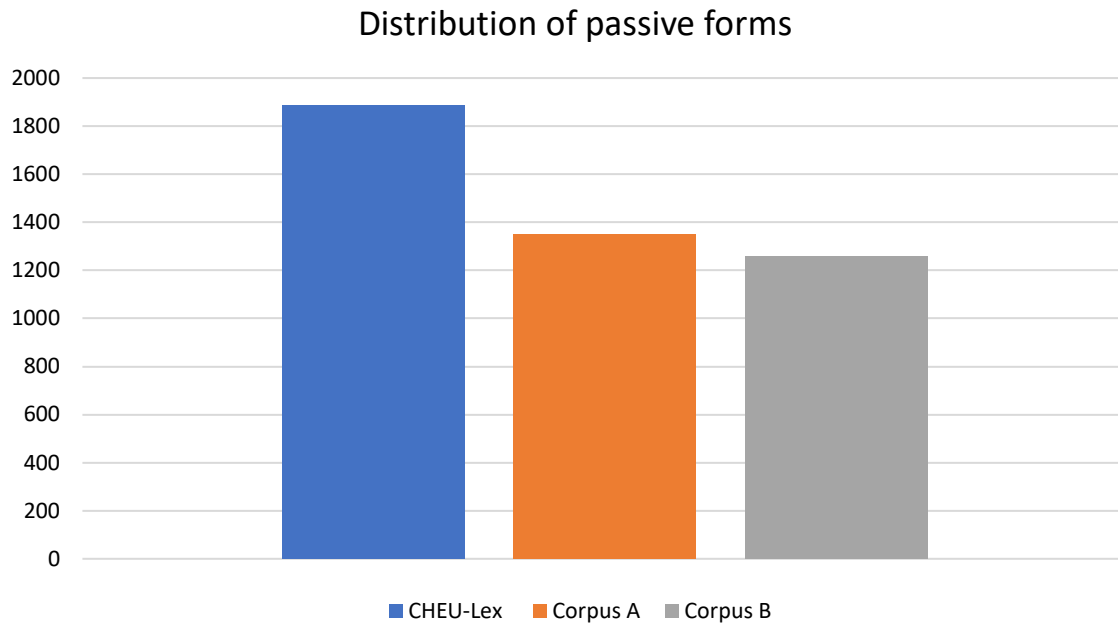


Figure 3.10: The sums of the normalized frequencies of the passive forms in the three corpora analyzed.

3.5. Morphosyntax

In this section, a set of legal-administrative collocations was selected and the normalized frequency of each of them was extracted from Corpus A and Corpus B, with the aim of observing any possible relevant discrepancy between the corpora. Results show that these items are mainly exploited in national legislative Italian of implementation laws (Corpus B) (Mori, 2018c: 225).

The same analysis was carried out on the CHEU-Lex Corpus. Results were then compared with those obtained from Corpus B and Corpus A, as shown in Tables 3.22 and 3.23.

Item	CHEU-Lex Corpus		LL RATIO	Corpus B	
	Raw frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>fatt* salv*</i>	116	123.42	-177.97	919	464
<i>ferm* restando</i>	1	1.06	-218.98	346	175
<i>quanto + PP</i>	10	10.64	-996.68	1644	831
<i>in deroga a*</i>	38	40.43	-48.25	277	140
<i>in base a*</i>	177	188.33	-19.38	644	325
<i>sulla base d*</i>	130	38.32	-206.35	1046	529
<i>in ottemperanza a*</i>	0	0	-4.66	7	4
<i>nel rispetto d*</i>	15	15.96	-533.49	979	495
<i>tenuto conto di*</i>	29	30.86	+19.17	21	11
<i>a seguito d*</i>	15	15.96	-55.91	188	95
<i>a carico d*</i>	42	44.69	-272.61	743	375
<i>in seguito a*</i>	54	57.46	+36.89	38	19
<i>secondo quanto + PP</i>	1	1.06	-144.6	233	118
<i>senza pregiudizio d*</i>	4	4.26	-0.2	13	7

Table 3.22: The normalized frequencies of the legal-administrative collocations in the CHEU-Lex Corpus and in Corpus B.

According to the results showed in Table 3.22, most of the item selected seem to be more widely used in Corpus B than the CHEU-Lex Corpus. The collocation *in ottemperanza a** appears to have 0 occurrences in the CHEU-Lex Corpus, whereas the only two elements showing a normalized frequency higher in the CHEU-Lex Corpus are *tenuto conto di** (NF: 30.86 in the CHEU-Lex Corpus and 11 in Corpus B) and *in seguito a** (NF: 57.46 in the CHEU-Lex Corpus and 19 in Corpus B).

Almost the same happens when comparing the results with Corpus A:

	CHEU-Lex Corpus		LL RATIO	Corpus A	
	Raw frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>fatt* salv*</i>	116	123.42	-285.4	875	608
<i>ferm* restando</i>	1	1.06	-43.06	59	41
<i>quanto + PP</i>	10	10.64	-67.33	138	96
<i>in deroga a*</i>	38	40.43	-113.89	319	222
<i>in base a*</i>	177	188.33	-91.28	692	481
<i>sulla base d*</i>	130	38.32	-14.83	352	245
<i>in ottemperanza a*</i>	0	0	-6.07	7	5
<i>nel rispetto d*</i>	15	15.96	-51.51	136	95
<i>tenuto conto di*</i>	29	30.86	-36.22	158	110
<i>a seguito d*</i>	15	15.96	-51.51	136	95
<i>a carico d*</i>	42	44.69	-0.19	84	58
<i>in seguito a*</i>	54	57.46	+1.68	79	55
<i>secondo quanto + PP</i>	1	1.06	-42.23	58	40
<i>senza pregiudizio d*</i>	4	4.26	-18.97	44	31

Table 3.23: The normalized frequencies of the legal-administrative collocations in the CHEU-Lex Corpus and in Corpus A.

Here, only the normalized frequency of *in seguito a** is slightly higher in the CHEU-Lex Corpus than in Corpus A (57.46 in CHEU-Lex Corpus and 55 in Corpus A), whereas all the other items show a significant underuse in the CHEU-Lex Corpus if compared to Corpus A. This is also attested by the LL ratios. Thus, legal-administrative collocations appear to be more widely used in Corpus A than in the CHEU-Lex Corpus.

Overall, it seems as though the distribution of the legal-administrative variants selected are widely more present in Corpus B and, to a lesser extent in Corpus A and CHEU-Lex Corpus, as shown in the following graph.

Distribution of legal-administrative collocations

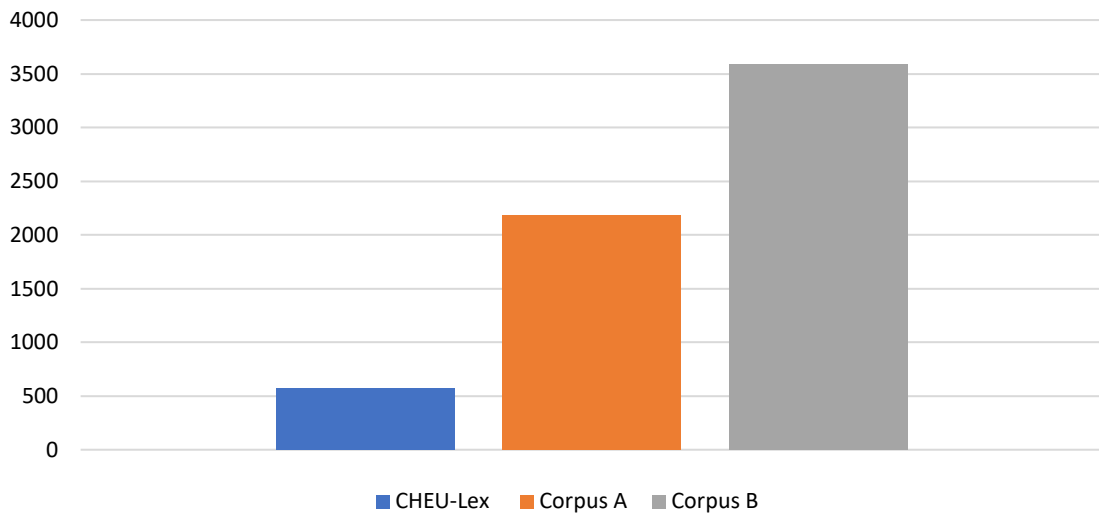


Figure 3.11: The sums of the normalized frequencies of the legal-administrative collocations in all three corpora.

Besides the distribution of legal-administrative collocations, “lexical bundles” were retrieved as well in both corpora (Corpus A of EU directives and Corpus B of Italian laws of implementation). The same analysis was carried out on *SketchEngine* by means of the *N-Grams* tool. However, as already stated in Section 2.4.5., *SketchEngine* does not provide the sum of the normalized frequencies of all the lexical bundles (i.e., for instance, the sum of the normalized frequencies of all the 3-Grams); indeed, only the normalized frequency of every single N-Gram retrieved was calculated automatically. Therefore, an Excel spreadsheet was used to calculate the sum of the normalized frequencies of every set of N-Grams. Results are shown in the table below (Tables 3.24 and 3.25), and were then compared to those obtained in Corpus B and Corpus A.

Lexical bundles (N-grams)	CHEU-Lex LAWS		LL RATIO	Corpus B	
	Raw (total) frequency	Normalized (total) Frequency		Raw frequency	Normalized frequency
3Grams	181,481	232,440	+254913.59	47,957	24,235
4Grams	75,603	96,832.2	+83387.81	30,781	15,555
5Grams	34,760	44,520.55	+29190.62	20,000	10,107
6Grams	18,390	23,553.88	+11766.79	13,685	6,916

Table 3.24: The sums of the normalized frequencies of the N-Grams of length from 3 to 6 in the CHEU-Lex Corpus and in Corpus B.

Lexical bundles (N-grams)	CHEU-Lex LAWS		LL RATIO	Corpus A	
	Raw (total) frequency	Normalized (total) Frequency		Raw frequency	Normalized frequency
3Grams	181,481	232,440	+220632.19	34,076	23,679
4Grams	75,603	96,832.2	+70622.42	23,095	16,049
5Grams	34,760	44,520.55	+2138.62	45,970	31,944
6Grams	18,390	23,553.88	+8856.15	11,227	7,802

Table 3.25: The sums of the normalized frequencies of the N-Grams of length from 3 to 6 in the CHEU-Lex Corpus and in Corpus A.

It is quite interesting that the normalized frequencies are higher in the CHEU-Lex Corpus if compared both to Corpus A and Corpus B. An explanation for this could be the use of CAT tools and translation memories within the Swiss institutions⁷¹. Consistency among legal texts could be another explanation, but since this feature is common to the other corpora as well (Corpus A and Corpus B), the difference of the normalized frequencies (along with the LL ratios) should not be so important.

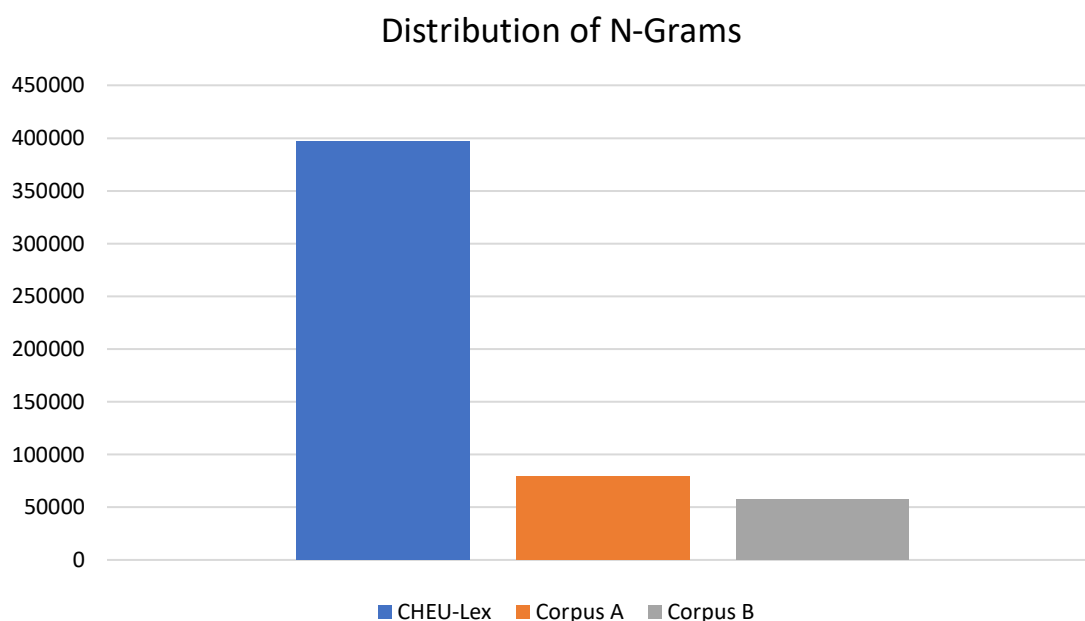


Figure 3.12: The sum of the normalized frequencies of every item (3-Gram, 4-Gram, 5-Gram and 6-Gram) in the three corpora.

According to the graph in Figure 3.12, it is evident that the normalized frequency of 3-Grams in the CHEU-Lex Corpus reaches a peak (NF: 232,440). The normalized frequency of the

⁷¹ The same explanation was given by Mori when analyzing the huge number of longer N-grams in Corpus A. In particular, she states that the values retrieved from Corpus A coincide with clauses which are “[...] identically repeated for internal and cross-text consistency [...]” (Mori, 2018c: 227).

items in the CHEU-Lex corpus then decreases until it reaches with 6-Grams almost the same value (although still bigger) of the normalized frequencies of Corpus A and Corpus B. Also, the two corpora (Corpus A and Corpus B) seem to share more or less the same amount of N-grams.

Overall, this could be interpreted as a sign of high standardization of Italian or Swiss laws of implementation, if compared to the two other legal varieties. This could, however, also be the result of the environment in which the legal texts are drafted. Indeed, it is important to remember that the Swiss Confederation is a multilingual nation with four national languages.⁷² It could be argued that the need for a standardized and consistent phraseology as well as the use of CAT tools was already present in Switzerland long before the agreements with the European Union, and this could have led to the high standardization of the legal texts. However, this hypothesis could be confirmed only with further studies, e.g. comparing the CHEU-Lex Corpus with a corpus of Swiss national Italian legislation unrelated to the European environment.

3.6. Syntax

Syntax is another level which is worth analyzing. Indeed, language complexity can be considered as a yardstick when comparing two legislative varieties of a given language.

“Differing degrees of sentence complexity between the two legislative varieties were already reported in previous qualitative studies by Mori, where a less complex intra- and inter-sentential texture in EU law was noted.” (Mori, 2018c: 227)

In the following sections, EU-rooted phenomena and intra-linguistic variability were analyzed, in order to observe some statistically significant differences or similarities between the CHEU-Lex Corpus and Corpus B. Then, a comparison with Corpus A was also carried out.

⁷² I.e. German, French, Italian and Romansh.

3.6.1. EU-rooted syntactical phenomena

As already stated in Section 2.4.6.1, for what concerns sentence structure, some EU-derived features were detected during previous studies which distinguish the Italian of EU directives from the domestic variety of Italian. In particular, the use of subject pronouns was observed to be widespread in Italian of EU directives when compared to Italian of implementation laws. This is because Italian of EU directives tends to repeat the subjects, whereas the Italian of implementation laws is more likely to use the subject ellipsis (“zero anaphora”; Mori, 2018c: 228). To prove this, a corpus-driven analysis was carried out by Mori, observing the distribution of the subject pronouns *esso*, *essa*, *essi* and *esse*.

The same analysis was performed on the CHEU-Lex Corpus, with the aim of observing any interesting syntactical difference between the Swiss Italian of implementation laws and the domestic Italian of implementation laws. A comparison with Corpus A (Eurolect Italian) was also carried out.

Item	CHEU-Lex Corpus		LL ratio	Corpus B	
	Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
<i>esso</i>	437	369.21	+524.58	191	97
<i>essa</i>	332	353.25	+301.31	175	88
<i>essi</i>	206	219.18	+65.28	238	120
<i>esse</i>	124	131.94	+53.25	122	62

Table 3.26: The distribution of the subject pronouns in the CHEU-Lex corpus and Corpus B.

Item	CHEU-Lex Corpus		LL ratio	Corpus A	
	Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
<i>esso</i>	437	369.21	+146.95	335	233
<i>essa</i>	332	353.25	+26.64	417	290
<i>essi</i>	206	219.18	-411.89	1,397	971
<i>esse</i>	124	131.94	-1.69	263	183

Table 3.27: The distribution of the subject pronouns in the CHEU-Lex corpus and Corpus B.

Statistics reported in Tables 3.26 and 3.27 show that there is a substantial overuse of subject pronouns in the CHEU-Lex Corpus when compared to Corpus B. Almost the same happens for the comparison with Corpus A, where only *essi* and *esse* are used less in the CHEU-Lex Corpus than in Corpus A. Indeed, *esse* has a normalized frequency higher in Corpus A (183)

if compared to that in CHEU-Lex Corpus (131.94), as well as *essi* (NF: 217.18 in the CHEU-Lex Corpus and 971 in Corpus A).

Overall, there tends to be an overuse of subject pronouns in the Swiss Italian laws of implementation (CHEU-Lex Corpus), especially if compared to the domestic Italian of implementation laws (Corpus B). The same applies, to a lower extent, to the comparison with Corpus A.

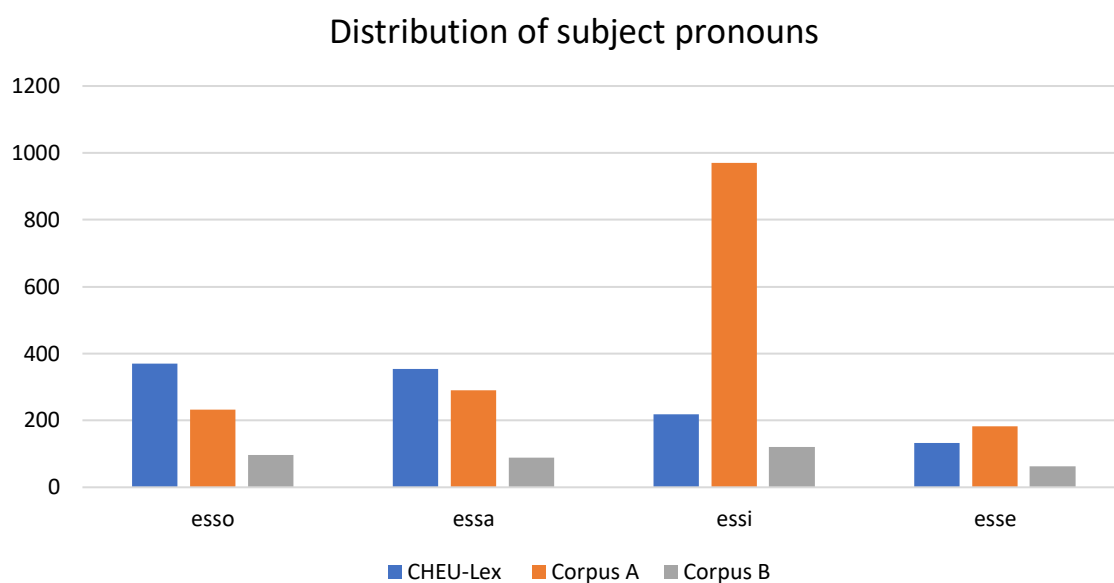


Figure 3.13: The normalized frequencies of the subject pronouns in the three corpora.

Considering the statistics and the graph shown in Figure 3.13, it is possible to argue that Swiss Italian of implementation laws behaves in a similar way to Italian of EU directives, in other words it does not seem to use subject ellipsis as much as Italian of implementation laws. This could be due to various factors: among them, the European legal drafting process, as well as the multilingual environment (Swiss Confederation) in which laws are drafted. Indeed, it is worth remembering that the repetition of the subject in the Eurolect Italian is attributed to the syntactic structure of the languages in which texts are drafted in the first place within the EU scenario: “In Italian of EU directives [...] the subject is generally not omitted, because of the syntactic skeleton of EU master texts written-negotiated-revised in English and French” (Mori, 2018c: 227. See also Section 2.4.6.1.). Considering the multilingual landscape of Switzerland, the same scenario could happen within the Swiss borders. Also, it is also important to remember that EU directives are adopted in the Swiss

Confederation starting from the German version⁷³ and are eventually translated into French and Italian (see Section 2.2.). This process could have influenced the use of subject pronouns as well.

3.6.2. Intra-linguistic syntactic variability

Another feature typical of the Italian of EU directives is the simplicity of the sentences used. In particular short sentences, a lower structural complexity, as well as the avoidance of the embedded clauses are among the most common features of Italian of EU directives. This can be linked to the need for simple legislative texts, which must be accessible to every European citizen, but also to the fact that, as previously stated, EU texts are written and revised in English and French. To investigate this feature, a set of inter-sentential connectives were selected, and their use was observed both in Corpus A and Corpus B.

The same analysis was carried out on the CHEU-Lex Corpus as well. Results were then compared to those obtained in Corpus A and Corpus B; they are listed in the following tables.

⁷³ It is important to remember that the European German version which inspires the Swiss German version is translated from either the English or the French European versions.

Grammatical Information	Item	CHEU-Lex Corpus		LL ratio	Corpus B	
		Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
Purpose (explicit)	<i>affinché + subj</i>	11	11.7	-99.98	245	124
Purpose (implicit)	<i>per + inf.</i>	1,082	1,151.25	+42.92	2,140	1,081
	<i>al fine di + inf.</i>	79	84.06	-287.24	976	493
Cause	<i>dal momento che</i>	0	0	0	0	0
	<i>dato che</i>	0	0	-3.33	5	3
	<i>perché</i>	31	32.98	+7.13	41	21
	<i>cui</i>	3,049	3,244.15	-3975.28	22,480	11,360
Relative	<i>il/la quale; i/le quali</i>	335	356.44	-191.76	1,812	916
	<i>se</i>	4,029	4,286.87	+2261.72	3,321	1,678
Restrictive and hypothetical	<i>qualora + subj.</i>	93	98.95	-585.23	1,612	815
Restrictive conditional	<i>a condizione e che + subj.</i>	18	19.15	-108.31	303	153
	<i>laddove</i>	15	15.96	-13.44	95	48
	<i>nel caso in cui</i>	40	42.56	-168.56	539	272
	<i>nella misura in cui</i>	34	36.18	+9.81	41	21
	<i>ove</i>	85	90.44	-350.71	1,131	572
	<i>purché + subj.</i>	33	35.11	-91.26	350	177
	<i>sempreché + subj.</i>	10	10.64	-0.56	33	17

Table 3.28: The distribution of the inter-sentential connectives in the CHEU-Lex Corpus and Corpus B.

Among the results obtained, *per + inf* and *se* emerge among the others for their high normalized frequencies in the CHEU-Lex Corpus (1,151.25 and 4,286.87, respectively), compared to those of Corpus B (1,081 and 1,678, respectively), followed by *nella misura in cui* (NFs: 36.18 in the CHEU-Lex Corpus and 21 in Corpus B) and *perché* (NFs: 32.98 in the CHEU-Lex Corpus and 21 in Corpus B). The other items show a normalized frequency

lower in the CHEU-Lex Corpus compared to that of Corpus B. *Dato che* and *dal momento che* do not appear in the CHEU-Lex Corpus. Quite interestingly, *dato che* does not appear in Corpus B as well, and *dal momento che* has only 5 occurrences.

Results obtained in the CHEU-Lex Corpus were then compared to those of Corpus A, as shown in the following table (Table 3.29):

Grammatical Information	Item	CHEU-Lex Corpus		LL ratio	Corpus A	
		Raw frequency	Normalized frequency		Raw frequency	Normalized frequency
Purpose (explicit)	<i>affinché + subj.</i>	11	11.7	-1039	1,319	917
Purpose (implicit)	<i>per + inf.</i>	1,082	1,151.25	-70.57	2,683	1,864
	<i>al fine di + inf.</i>	79	84.06	-101.1	435	302
Cause	<i>dal momento che</i>	0	0	-20.8	24	17
	<i>dato che</i>	0	0	-59.81	69	48
	<i>perché</i>	31	32.98	+1.91	41	28
Relative	<i>cui</i>	3,049	3,244.15	-2111	13,169	9,151
	<i>il/la quale; i/le quali</i>	335	356.44	-12.13	773	537
Restrictive and hypothetical	<i>se</i>	4,029	4,286.87	+683.18	4,145	2,880
Hypothetical	<i>qualora + subj.</i>	93	98.95	-832.25	1,562	1,085
Restrictive conditional	<i>a condizionale + che + subj.</i>	18	19.15	-183.53	332	231
	<i>laddove</i>	15	15.96	-58.82	147	102
	<i>nel caso in cui</i>	40	42.56	-13.9	139	97
	<i>nella misura in cui</i>	34	36.18	-27.29	155	108
	<i>ove</i>	85	90.44	-107.7	466	324
	<i>purché + subj.</i>	33	35.11	-241.61	482	335
	<i>sempreché + subj.</i>	10	10.64	-12.76	55	38

Table 3.29: The distribution of the inter-sentential connectives in the CHEU-Lex Corpus and Corpus A.

The comparison with Corpus A shows a greater difference in the use of the inter-sentential connectives selected. The vast majority of the results has, indeed, a lower normalized frequency in the CHEU-Lex Corpus compared to that of Corpus A; only *perché* and *se* are an exception, with the normalized frequencies of 32.98 and 4,286.87, respectively (compared to 28 for *perché* and 2,880 for *se* in Corpus A).

In Swiss legal Italian the aim of the law is mainly expressed through explicit purpose subordinates (*per* + infinitive), similarly to what happens in the Italian of EU directives. This could be due to the fact that laws in the CHEU-Lex Corpus have to define goals to be achieved both by Switzerland and the European Union. A similar explanation was given by Mori when trying to explain why, concerning the Italian of EU directives, explicit and implicit purpose connectives were more present if compared to Corpus B: “[...] because directives have to define goals to be achieved by member states in their domestic law” (Mori, 2018c: 231).

For what concerns causal connectives, *Dal momento che* and *dato che* were observed to be particularly profitable in Italian of EU directives (Mori, 2018c), but, as far as the Swiss legal Italian is concerned, they are not attested, similarly to what happens in Corpus B.

The conditional connective *se* was extremely used in Italian of EU directives (Mori, 2018c). The comparison with the CHEU-Lex Corpus shows that this element happens to be even more used in Swiss legal Italian than in the Italian of EU directives. This is also proved by the LL ratio.

Overall, it seems as though there is a massive overuse of the relative connectives in Corpus A (Italian of EU directives) if compared to the Italian of Swiss laws of implementation (CHEU-Lex Corpus) and Italian of implementation laws (Corpus B). Purpose connectives (explicit and implicit) are generally more used in the Italian of EU directives as well, whereas for what concerns restrictive and hypothetical connectives, these seem to be more widespread in the CHEU-Lex Corpus. In general, however, Corpus B seems to be the one which uses the selected connectives the least, as shown in the following graph (Figure 3.14).

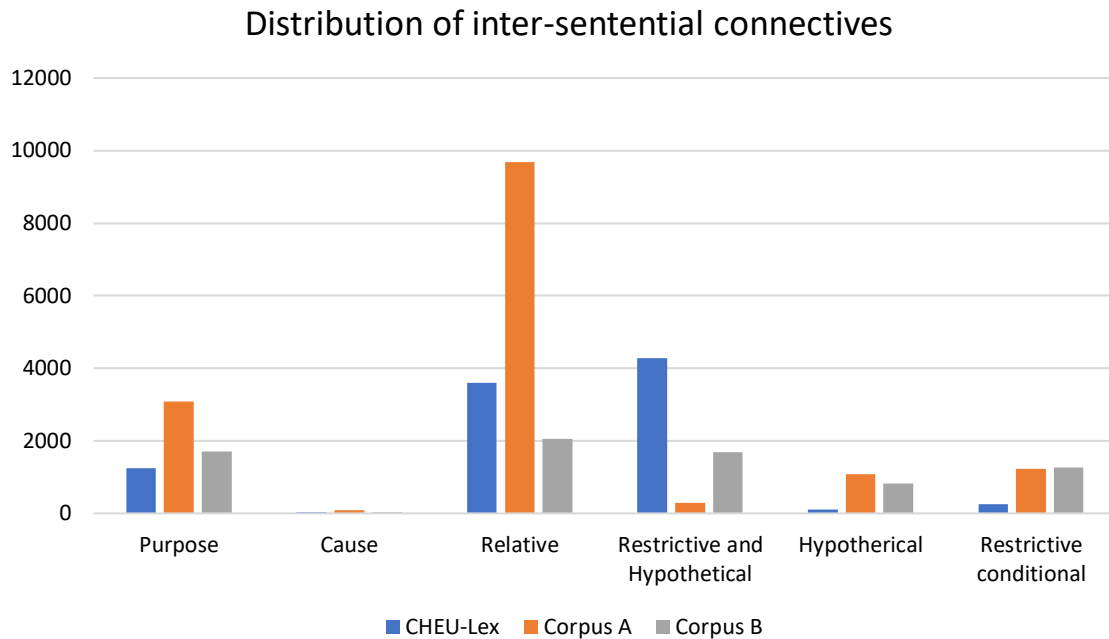


Figure 3.14: The distribution in the three corpora of the different inter-sentential connectives, according to the grammatical information they provide.

3.7. Textual level

The last dimension that was analyzed is the textual one. As explained by Mori, due to the technical content of legislative texts, a high standardization of the language usage is expected to be found. It is indeed true that legal texts are organized into a hierarchy, dividing texts into units and sub-units (Mori, 2018c and European Union, 2015.).

3.7.1. EU-rooted textual phenomena

Although, as already explained in Section 2.4.7.1., both EU and Italian legislation are characterized by a well-defined macro structure, the two structures differ: Italian national laws present “*capì*”, “*titoli*”, “*articoli*” and “*commì*”, whereas EU directives are divided into “*capitoli*”, “*paragrafi*”, “*articoli*” and “*lettere*”. The last nomenclature can be considered as an EU-related difference, as it is common to the Eurolects (Mori, 2018c).

Frequencies of the above-mentioned items were extracted from Corpus A and Corpus B during the *Eurolect Observatory Project*, and evidence was found supporting the use of a different nomenclature.

The same research was performed on the CHEU-Lex Corpus, and results (Tables 3.30 and 3.31) were then compared to those obtained from Corpus B and from Corpus A, to check whether similarities can be found between the CHEU-Lex corpus and either Corpus A or Corpus B at a macro-structural level.

Item	CHEU-Lex Corpus		LL Ratio	Corpus B	
	Raw frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>articolo/i</i>	11,416	12,146.66	+1126.97	19,309	9,758
<i>capitolo/i</i>	541	575.63	+1194.93	22	11
<i>capo/i</i>	118	125.55	-354.93	1,310	622
<i>comma/i</i>	0	0	- 13847.41	20,818	10,521
<i>lettera/e</i>	910	968.24	-754.5	5,598	2,829
<i>paragrafo/i</i>	72	76.61	-121.52	596	301
<i>titolo/i</i>	593	630.94	-18.99	1,840	930

Table 3.30: The use of the specific nomenclature in the CHEU-Lex Corpus and Corpus B.

Item	CHEU-Lex Corpus		LL Ratio	Corpus A	
	Raw frequency	Normalized Frequency		Raw Frequency	Normalized Frequency
<i>articolo/i</i>	11,416	12,146.66	-14.34	21,979	15,273
<i>capitolo/i</i>	541	575.63	+455.8	191	133
<i>capo/i</i>	118	125.55	-318.33	935	650
<i>comma/i</i>	0	0	-932.75	1,076	748
<i>lettera/e</i>	910	968.24	-123.58	2,544	1,768
<i>paragrafo/i</i>	72	76.61	-8606.54	10,753	7,472
<i>titolo/i</i> ⁷⁴	593	630.94	+110.22	591	411

Table 3.31: The use of the specific nomenclature in the CHEU-Lex Corpus and Corpus A.

Overall, there is an overuse of *articolo/i* (NF: 12,146.66) and *capitolo/i* (NF: 575.63) in the CHEU-Lex Corpus if compared both to Corpus B (NFs: 9,758 and 11, respectively). *capitolo/i* is overused in the CHEU-Lex if compared to Corpus A, too (NF: 133), followed by *titolo/i* (NFs: 630.94 in the CHEU-Lex Corpus and 411 in Corpus A). This could be due to the rather fixed structure of the enacting parts of the Swiss laws of implementation, but it is not clear whether this is an influence of the European legal drafting process or, instead, it is a rather fixed feature of Swiss legal texts. To answer this question, it would be necessary

⁷⁴ A further check would be necessary for *titolo/i*, as some occurrences such as “*a titolo di*” have been observed, and should therefore be excluded from the results obtained, as they clearly do not refer to the structure of the text.

to carry out a further analysis between the CHEU-Lex Corpus and a corpus of Swiss national laws.

3.8. General remarks, limitations of this study and future developments

The method used, which mirrored that of Mori (2018c), allowed the research to be carried out at different levels, and to collect quantitative data which were subsequently compared to those obtained by Mori both in Corpus A (EU directives) and Corpus B (Italian laws of implementation). The comparison among the three different corpora aimed at pointing out any interesting linguistic feature in the CHEU-Lex Corpus that could be linked to the European legal drafting process and, in a more indirect way, to the influence of the Eurolects on a supranational level.

Similarities between the CHEU-Lex Corpus and Corpus B (Italian of implementation laws) were found at different levels. First, the specialized Type/Token Ratio of the enacting parts of the CHEU-Lex Corpus is almost identical to that of Corpus B (34.5 and 34.46, respectively). Also, the distribution of some selected EU-Noun phrases and semantic Europeisms appear to be underused both in Corpus B and in the CHEU-Lex Corpus when compared to Corpus A (EU directives). EU-Noun phrases are another item observed to be underused in Corpus B and in the CHEU-Lex Corpus when compared to Corpus A.

However, similarities between the CHEU-Lex Corpus and Corpus A were found as well. Register markers, for instance, are extremely used in Corpus B, but to a lesser extent in Corpus A and the CHEU-Lex Corpus, thus indicating that the Italian used in the EU directives (Corpus A) and that used in the Swiss implementation laws (CHEU-Lex Corpus) tend to remain as simple as possible. Impersonal structures (*è necessario* and *occorre*) are another feature shared between Corpus A and the CHEU-Lex Corpus and are underused in Corpus B instead. Subject pronouns (*esso*, *essa*, *essi* and *esse*) are widely used both in Corpus A and in the CHEU-Lex Corpus and are, instead, underused in Corpus B (Italian laws of implementation).

Overall, only few results indicated the presence of potential traces of the European legal drafting process. However, these results do not provide a clear and strong evidence confirming the influence of the European legal drafting process on the Italian of Swiss laws of implementation. These results should therefore be further investigated through a comparison with a third corpus of Swiss legal Italian, collecting legislative texts unrelated

to the European environment: if the similarities observed within the CHEU-Lex Corpus and Corpus A are confirmed, then it is likely that these are typical features of the Swiss legal system. If, instead, these similarities are not found, or they are found to be more limited, then the hypothesis of a possible European influence could be put forward.

To conclude, this thesis provides a further point of view to look at the influence of the European legal drafting process on a national legislation: that of the Swiss Confederation. It is worth remembering that the analysis was carried out on the Swiss corpus of Italian laws implementing the bilateral agreements between the Swiss Confederation and the European Union, i.e. the CHEU-Lex Corpus. What emerged is that there is a faint trace of a possible EU influence when dealing with the Swiss federal legislation representing the reception of these agreements. Nonetheless, this evidence is not strong enough to righteously claim the effective influence of the European Legal drafting process on the Italian Swiss of implementation laws, and therefore these results need to be confirmed by carrying out further research.

CONCLUSIONS

This thesis is based on the *Eurolect Observatory Project* and on the project led by Professor Annarita Felici (University of Genève), which aimed at creating the CHEU-Lex Corpus on which the analysis of this thesis was carried out. The aim of the present thesis was to observe whether any relevant influence of the European legal drafting process could be observed in the Swiss legal Italian. To carry out this study, three different corpora were taken in consideration: Corpus A, Corpus B and the CHEU-Lex Corpus. Corpus A and Corpus B were previously analyzed by Mori (2018c) during the *Eurolect Observatory Project*, and they comprise European directives in Italian (Corpus A) and their Italian implementation laws (Corpus B). The third corpus is the Italian CHEU-Lex Corpus; it collects the Swiss federal legislation (*laws*) representing the reception of the bilateral agreements entered between the Swiss Confederation and the European Union (*agreements*).

The study was based on the idea that the CHEU-Lex Corpus (the Swiss corpus of implementation laws) might feature similar patterns to Corpus B (the corpus of Italian laws of implementation). The idea was supported by the fact that both corpora collect national legislation (Swiss and Italian, respectively) resulting from the contact with European Union directives (*Leggi di implementazione*). Corpus A, instead, collects EU directives written in Italian, and it is therefore used as a *tertium comparationis*, to further check whether influences of the European legal drafting process can be observed in Swiss legal Italian.

The method followed in this thesis closely followed that adopted during the *Eurolect Observatory Project*. Also, the same research template was adopted (Section 2.4.1). It was conceived by Mori (2018b) as a reference to be followed during the *Eurolect Observatory Project*, and it represents the various linguistic levels which must be analyzed in order to find any trace of the European drafting and translational process in any European language. These levels are *lexical*, *morphological*, *morphosyntactic*, *syntactic* and *textual*. Together with these levels, three heuristic macro-areas of research were chosen: *EU-Rooted-phenomena*, *Contact-induced features* and *Intra-linguistic variability*.

Results showed that similarities between the CHEU-Lex Corpus and Corpus B (Italian of implementation laws) were found at different levels. First, the specialized Type/Token Ratio of the enacting parts of the CHEU-Lex Corpus is almost identical to that of Corpus B (34.5 and 34.46, respectively); this means that the lexicon of CHEU-Lex's *enacting terms* is as

varied as that of Corpus B. Therefore, this could suggest that in the Italian of Swiss laws of implementation (CHEU-Lex Corpus) the use of synonyms is as frequent as in the Italian of implementation laws (Corpus B). Another feature worth considering is the distribution of some selected EU-Noun phrases and semantic Europeisms, which appear to be underused both in Corpus B and in the CHEU-Lex when compared to Corpus A (EU directives). Since these elements are considered typical features of the Eurolects, their limited presence in the CHEU-Lex Corpus could imply that the EU law making process does not seem to impact the Swiss legal Italian to a great extent.

However, similarities between the CHEU-Lex Corpus and Corpus A were found as well. Register markers, for instance, are extremely used in Corpus B, but to a lesser extent both in Corpus A and the CHEU-Lex Corpus, thus indicating that the Italian used in the EU directives (Corpus A) and that used in the Swiss implementation laws (CHEU-Lex Corpus) tend to remain as simple as possible, maintaining a simple structure. This could be due to the fact that the drafting of European laws requires a language as simple as possible, avoiding complex verb forms and structures. Therefore, this could be interpreted as a possible trace of the European legal drafting process in the Italian of Swiss laws of implementation. Impersonal structures (*è necessario* and *occorre*) are another feature shared between Corpus A and the CHEU-Lex Corpus, along with the distribution of subject pronouns (*esso*, *essa*, *essi* and *esse*). These elements are widely used both in Corpus A (EU directives) and in the CHEU-Lex Corpus and are, instead, underused in Corpus B (Italian laws of implementation). The fact that subject pronouns are overused in Corpus A underlines that Italian of EU directives tends to repeat the subjects, whereas the Italian of implementation laws (Corpus B) is more likely to use the subject ellipsis (“zero anaphora”; Mori, 2018c: 228). Since this feature is observed in the CHEU-Lex Corpus as well, this could be interpreted as a possible influence of the European legal drafting process.

Overall, similarities both with Corpus B and Corpus A were observed. These results could indeed indicate the presence of traces of the European legal drafting process. However, only few traits were noticeable, and they are not strong enough to righteously claim the effective presence of the influence of the European legal drafting process in the Swiss Italian of implementation laws. Indeed, some of the traits above mentioned might actually derive from reasons other than the European contact, such as the multilingual environment which characterizes the Swiss Confederation and/or the content of the bilateral agreement, which

might legitimate the presence of certain features. Take for instance the overuse of the verb *armonizzare*: as explained in Section 3.2.1., this could be due to the fact that the bilateral agreements entered between the two parties aim at ensuring the intergovernmental cooperation between the Swiss Confederation and the European Union.

In general, this study highlights the presence of some features in the Italian of Swiss laws of implementation (the CHEU-Lex Corpus) which might lead back to the influence of the European legal drafting process. However, these results do not provide strong evidence supporting the actual influence of the European legal drafting process on the Swiss legal Italian. These results should be checked by conducting a second study comparing the CHEU-Lex Corpus to a corpus of Swiss national legislation unrelated to the EU context, representing therefore the legal Italian used within the Swiss Confederation. In this way, it would be possible to compare the Swiss legal Italian used during the implementation procedure (which therefore might present some traits in common with the Italian Eurolect) to the legal Italian typically used in Switzerland during the drafting process.

AKNOWLEDGEMENTS

Grazie

Ai miei genitori, Stefano e Lucia, che mi hanno permesso di portare avanti i miei studi;

Alla mia famiglia tutta, che mi ha sempre sostenuta e motivata;

Ai miei amici, che mi hanno sopportata (e continuano a farlo);

A Michele, per esserci stato;

Ed infine, un grazie particolare va al mio relatore, il professor Adriano Ferraresi, che mi ha dato l'opportunità di collaborare al progetto di costruzione del corpus e, con molta pazienza, di portare avanti questa tesi.

BIBLIOGRAPHY

- Adinolfi G., Cappa E., Garzone G. & Santulli F. (2008). *Il linguaggio giuridico: prospettive interdisciplinari*. Milano: Giuffrè.
- Biber D., & Conrad S. (2009). *Register, genre, style*. Cambridge: CUP.
- Biel, L. (2014). *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt: Peter Lang.
- Bratanić M., Lončar M., (2015). *The Myth of EU Terminology Harmonization on National and EU Level*. In S. Šarcevic (Ed.), *Language and culture in EU law: Multidisciplinary perspectives* (pp. 207– 218). London: Ashgate.
- Caterina R., & Rossi P. (2008). *L'italiano giuridico*. In B. Pozzo & M. Timoteo (Eds.), *Europa e linguaggi giuridici* (pp. 185– 208). Milano. Giuffrè Editore.
- Felici, A. (2015). “Translating EU legislation from a lingua franca: Advantages and disadvantages”. In S. Šarcevic (Ed.), *Language and culture in EU law: Multidisciplinary perspectives* (pp. 123– 140). London: Ashgate.
- Fernando Prieto Ramos (2014) *International and supranational law in translation: from multilingual lawmaking to adjudication*, The Translator.
- Graziadei, M., (2015). “Law, Language and Multiculturalism in Europe: The Call for a New Legal Culture”. In S. Šarčević (Ed.), *Language and culture in EU law: Multidisciplinary perspectives*. London: Ashgate.
- Grishman, R., (1986). *Computational Linguistics. An introduction*. Cambridge University Press.
- Lubello, S., (2016), “Usi pubblici e istituzionali dell’Italiano”. In Lubello, S., *Manuale di linguistica italiana*. Berlin, Boston: De Gruyter.
- Mantovani, D., (2008). “Lingua e diritto. Prospettive di ricerca fra sociolinguistica e pragmatica”. In Adinolfi, G., Cappa E., Garzone G. & Santulli F. *Il linguaggio giuridico: prospettive interdisciplinari*. Milano: Giuffrè.

- Mattila, H. (2006). “Comparative jurilinguistics: A discipline in statu nascendi”. In B. Pozzo & V. Jacometti (Eds.). *Multilingualism and harmonisation of European Law*. The Hague: Kluwer Law International.
- Mattila, H. E. S. (2013). *Comparative Legal Linguistics: Language of Law, Latin and Modern Lingua Francas*: Vol. 2nd ed. Routledge.
- Mori, L. (2018a). *Observing Eurolects*. Amsterdam: John Benjamins Publishing Company.
- Mori, L. (2018b). “Introduction. The Eurolect Observatory Project”. In Mori, L. *Observing Eurolects*. Amsterdam: John Benjamins Publishing Company.
- Mori, L. (2018c). “Observing Eurolect. The case of Italian”. In Mori, L. *Observing Eurolects*. Amsterdam: John Benjamins Publishing Company.
- Mori, L. (2018d). “Conclusions. A cross-linguistics overview on Eurolects”. In Mori, L. *Observing Eurolects*. Amsterdam: John Benjamins Publishing Company.
- Mori L., & Szmrecsanyi B., (2020). “Mapping Eurolects: an aggregate perspective on similarities between legislative varieties”. *Languages in Contrast*. Volume 2. Issue 2: 186-216.
- Nystedt, J. (1998). *L’italiano nei documenti ufficiali della CEE. Il progetto Stoccolma: presentazione e sommario di dati stilolinguistici, statistici e quantitativi* (Cahiers de Recherche 10). Stockholm: Arcomedia.
- Pozzo, B. (2006). “Multilingualism, Legal Terminology and the problems of Harmonising European Private Law”. In B. Pozzo & V. Jacometti (Eds.), *Multilingualism and harmonisation of European Law* (pp. 21– 32). The Hague: Kluwer Law International.
- S. Šarčević (2015). *Language and culture in EU law: Multidisciplinary perspectives*. London: Ashgate.
- Swiss Federal Chancellery. (2019). *Règles d’or de la mise en œuvre du droit de l’UE dans le droit suisse*. Bern: Swiss Confederation.

Tomatis, M. S., (2018). *The Eurolect Observatory Multilingual Corpus*. In Mori, L. *Observing Eurolects*. Amsterdam: John Benjamins Publishing Company.

Turchetta, B., Mori, L., & E., Ranucci (2005). *Il mondo in italiano. Varietà e usi internazionali della lingua*. Bari: Laterza.

SITOGRAPHY

- Basile, V. (2020). “I Computer e il Linguaggio Naturale”. *Itaca*.
http://ithaca.unisalento.it/nr-16_2020/articolo_Iip_11.pdf (Retrieved: 3 January 2022).
- Biel, L., (2010). *Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential*,
https://www.academia.edu/596750/Corpus_Based_Studies_of_Legal_Language_for_Translation_Purposes (Retrieved: 3 January 2022).
- Council of the European Economic Community (1958). “REGULATION No 1 determining the languages to be used by the European Economic Community”. *Official Journal of the European Communities*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31958R0001&qid=1645376150188&from=en> (Retrieved: 25 January 2022).
- Directorate-General for Communication of the European Commission, (2021). *Eu&Me - WHAT IS THE EUROPEAN UNION?* https://op.europa.eu/webpub/com/eu-and-me/en/WHAT_IS_THE_EUROPEAN_UNION.html#:~:text=The%20European%20Union%20is%20not,much%20of%20the%20European%20continent. (Retrieved: 27 December 2021).
- European Commission, (2011). *How to write clearly*. <https://op.europa.eu/en/publication-detail/-/publication/c2dab20c-0414-408d-87b5-dd3c6e5dd9a5> (Retrieved: 29 December 2021).
- European Union. (2015). *Joint Practical Guide of the European Parliament, the Council and the Commission for persons involved in the drafting of the European Union legislation*. <https://eur-lex.europa.eu/content/techleg/KB0213228ENN.pdf> (Retrieved: 26 February 2022).
- General Multilingual Environment Thesaurus (GEMET). (2021) “*Harmonisation of law*”.
<https://www.eionet.europa.eu/gemet/en/concept/3847> (Retrieved: 11 February 2022)

- Goffin, R. (1994). “L’eurolécte : oui, jargon communautaire : non”. *Érudit*.
<https://doi.org/10.7202/002930ar> (Retrieved: 25 February 2022).
- Hanks, P., Frankenberg-Garcia, A., Krishnamurthy, R., & Rundell, M. (2021). *Create and search a text corpus / Sketch Engine*. <https://www.sketchengine.eu/> (Retrieved: 13 December 2021).
- Jenni, S. (2014), “Europeanization of Swiss Law-Making: Empirics and Rhetoric are Drifting Apart”. *Wiley Online Library*. <https://doi.org/10.1111/spsr.12098> (Retrieved: 25 February 2022).
- Lind, S., (2007), *Translation Universals (or laws, or tendencies, or probabilities, or...?)*,
https://www.academia.edu/8696942/Translation_Universals_or_laws_or_tendencies_or_probabilities_or (Retrieved: 24 January 2022).
- Robinson, W., (2008) *Drafting of EU legislation: a view from the European Commission*.
<https://www.federalismi.it/ApiOpenFilePDF.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%2Bdottrina%2B%2D%2B> (Retrieved: 26 February 2022).
- UNINT. Eurolect Observatory project. <https://www.unint.eu/en/research/research-projects/33-page/490-eurolect-observatory-project.html> (Retrieved: 26 February 2022).
- Université de Genève. (2021). *The CHEU-lex corpus – Transius*. Retrieved 24 November 2021, from <https://www.unige.ch/fti/de/a-la-une/cheu-lex-a-parallel-and-comparable-corpus-of-swiss-and-eu-law-in-three-languages/> (Retrieved: 26 February 2022).