

LEX.CH.IT: A CORPUS FOR MICRO- DIACHRONIC LINGUISTIC INVESTIGATIONS OF SWISS NORMATIVE ACTS IN ITALIAN

Paolo CANAVESE, PhD candidate

Faculty of Translation and Interpreting, University of Geneva
40, boulevard du Pont-d'Arve, 1211 Geneva 4, Switzerland
Paolo.Canavese@unige.ch

ORCID: <https://orcid.org/0000-0003-3928-8905>

Abstract: The aim of this paper is to present LEX.CH.IT, a corpus for micro-diachronic linguistic investigations of Swiss normative acts in Italian. Italian has a peculiar position as an official minority translation language within the Swiss institutional system. Until now, few studies have focused on Swiss legal Italian, but the academic interest has been growing over the last two decades. In order to further expand on research in this field, resources such as corpora are fundamental. This is why LEX.CH.IT has been compiled. This corpus was originally created in the context of a doctoral research project, which will be briefly outlined in this paper. The main goals of the project are to determine whether clarity is a feature of Swiss legislation in Italian, whether there have been relevant evolutions over the last five decades and to assess the role of translation for a clear legislation. In the future, LEX.CH.IT

could also be useful for a number of other projects aiming to shed light on the features of this language variety.

Key words: Switzerland, Swiss legal Italian, legislation, corpus, clarity, plain language

LEX.CH.IT: KORPUS DO BADAŃ JĘZYKOZNAWCZYCH I MIKRODIACHRONICZNYCH NAD SZWAJCARSKIMI AKTAMI NORMATYWNYMI W JĘZYKU WŁOSKIM

Abstrakt: Celem artykułu jest przedstawienie korpusu do badań mikrodiachronicznych i językoznawczych szwajcarskich aktów normatywnych w języku włoskim LEX.CH.IT. Język włoski w Szwajcarii posiada szczególną pozycję oficjalnego języka mniejszości. Do tej pory niewiele badań koncentrowało się na szwajcarskim prawnym i prawniczym języku włoskim, ale zainteresowanie naukowe tym tematem wzrosło w ciągu ostatnich dwóch dekad. Zasoby takie jak korpusy są fundamentalne by dalej rozwijać badania w tej dziedzinie. Dlatego też skonstruowano LEX.CH.IT. Chociaż korpus ten został pierwotnie stworzony na potrzeby doktoranckiego projektu badawczego, wyjaśniającego jasność szwajcarskiego ustawodawstwa w języku włoskim, może on być również przydatny w wielu innych projektach dotyczących wielojęzyczności prawa.

Słowa kluczowe: Szwajcaria; szwajcarski prawny i prawniczy język włoski; ustawodawstwo; korpus; jasność; prosty język

LEX.CH.IT: UN CORPUS PER RICERCHE LINGUISTICHE IN CHIAVE MICRODIACRONICA SULLA LEGISLAZIONE SVIZZERA IN ITALIANO

Abstract: L'obiettivo di questo articolo è presentare LEX.CH.IT, un corpus per ricerche linguistiche in chiave microdiacronica sulla legislazione svizzera in lingua italiana. L'italiano riveste una posizione particolare nel sistema istituzionale elvetico, in cui rappresenta una "lingua ufficiale minoritaria". Finora pochi studi si sono concentrati sull'italiano giuridico svizzero, ma negli ultimi due decenni l'accademia ha mostrato un interesse crescente in tal senso. Per sviluppare ulteriormente la ricerca in questo ambito sono fondamentali risorse quali corpora di dati. È in quest'ottica che è stato compilato LEX.CH.IT, corpus concepito ai fini di un progetto di dottorato che sarà presentato rapidamente in questo articolo. Il progetto si pone come obiettivo di determinare se la chiarezza è una caratteristica della legislazione

svizzera in italiano, se si sono verificate evoluzioni significative negli ultimi cinque decenni e di capire qual è l'impatto della traduzione sulla chiarezza legislativa. In futuro LEX.CH.IT potrà servire come base per diversi altri studi che permetteranno di far luce sulle peculiarità di questa varietà linguistica.

Parole chiave: Svizzera, italiano giuridico svizzero, legislazione, corpus, chiarezza, plain language

1. Introduction

The aim of this paper is to present LEX.CH.IT, a corpus for micro-diachronic linguistic investigations of Swiss normative acts in Italian. After a brief introduction on the presence of Italian in the Swiss legislation as an “official minority language”, this article will present the quite recent academic interest in carrying out research projects on this language variety.

While Swiss Italian has been examined – mostly from a sociolinguistic perspective – as an independent variety of Italian since the sixties, its use within the federal institutions is still underresearched. The growing interest in filling this research gap seems to go hand in hand with the expansion of legal and institutional translation studies (LTS). In this field, corpus studies have attained a central position and new resources are fundamental in this respect.

This paper offers an in-depth description of LEX.CH.IT, the rationale behind its design, as well as its compilation criteria. Moreover, the first application of this corpus is outlined: a doctoral research project that aims at investigating the level of clarity of Swiss legislation in Italian. Of course, LEX.CH.IT may serve as the basis for a number of other research projects aiming to advance research on Swiss legal Italian.

2. The presence of Italian in the Swiss legislation

Switzerland is a fascinating country for anyone interested in multilingualism and multilingual legislation. Federal authorities are obliged by law to express themselves in the three official languages of the Confederation.¹ Although institutional texts with an informative function, such as websites, sometimes neglect Italian, all federal enactments are issued simultaneously and integrally in German, French and Italian, and all language versions are equally binding.² This means that the three official publications issued by the federal authorities are *de facto* trilingual, i.e.:

- the **Official Compilation of Federal Legislation**, which publishes all enactments of the Confederation (such as the Federal Constitution, federal acts, ordinances of the Federal Assembly and of the Federal Council, federal decrees, international treaties and law decrees binding in Switzerland);
- the **Classified Compilation of Federal Legislation**, which is a collection of all Swiss legislation in force, i.e. all the texts published in the Official Compilation (plus the cantonal constitutions), organized by subject and regularly updated;
- the **Federal Gazette**, which is the journal of the Swiss Government. It includes, for instance, dispatches, reports, drafts of proposed laws by the Federal Council to the Federal Assembly, normative acts issued by the Federal Assembly that are subject to a referendum, and reports by parliamentary committees.

Equalization, however, at least in the field of legislation, is quite a recent achievement. The only two Italoophone cantons (Ticino and Graubünden) have been part of the modern Confederation since its foundation in 1803, but it was not until 1848 that Italian was recognized as a national language. While in the beginning, the presence of Italian within the institutional system had a merely prescriptive nature, the status of this minority language (Schweizer,

¹ Cf. Languages Act of 5 October 2007 (status as of 1 January 2017, CC 441.1), in particular art. 3, 5, 6, 8, 9, 10, 12.

² Cf. art. 14 of the Publications Act of 18 June 2004 (status as of 26 November 2018, CC 170.512).

Borghesi et al. 2011, 430) has constantly improved over the years.³ 1974 is a key year, as it is considered to be the beginning of a *new language regime* (Pini 2017, 79-90). Starting from 1974, all federal normative acts are voted and adopted in their Italian version (and not only in German and French, as was the case before) and the Federal Gazette is translated in its entirety into the “third language”.

From 1974 until today, the status of Italian has significantly evolved. The translation sector of the Federal Administration has been reorganized to be more efficient and the number of Italophone translators has increased considerably. Several measures have been adopted to guarantee a higher quality of institutional and legal texts: guidelines, seminars and the publication of a specialized journal dealing with the quality of legislation (LeGes), among many others. Moreover, the interactions and exchange of best practices with other national and supranational institutions has increased. Finally, the idea that the federal institutions should guarantee an equal treatment of the official languages and that they should write clear, comprehensible texts was integrated in the law. All these contextual changes have probably left some (linguistic) tracks in the Italian legal and institutional texts produced from the Seventies to present day.

Today, however, full equalization of Italian to German and French still remains a milestone to be attained.⁴ In the field of legislation, this is proven, for example, by the fact that normative acts at the federal level are almost never produced in Italian. Zwicky and Kübler (2018, 17-21) show that only around 1% of them are originally drafted in Italian. In other words, Italian is systematically a “translation language” (Egger 2015, 164; Snozzi 2005, 321) within an intra-systemic translation context (Biel 2017, 78). One should bear all these elements in mind when investigating Swiss legal Italian.

³ Pini (2017) offers a very exhaustive description of how the status of Italian has changed over the last two centuries.

⁴ More on this e.g. in Egger, Ferrari and Lala (2013b, 13) and in Egger and Grandi (2013, 213-214).

3. The institutional use of Swiss Italian: an area in need of more research?

The interest in studying Swiss Italian from a sociolinguistic perspective dates back to the sixties. Several influential works have been published since then (e.g. Lurati 1976; Berruto 1984; more recently: Moretti 2004, 2005; Pandolfi 2009). The existence of an institutional variety of Italian has been acknowledged since the beginning. Suffice it to refer to a paper by Biscossa (1968), which first introduced the concept of “federal Italian”. Berruto (1984) would later describe the variety of Italian used by the Swiss institutions using the term “Helvetic Italian”.

However, the interest in carrying out in-depth investigations into Swiss institutional and legal Italian is much more recent. This could be correlated to the fact that in the mid-2000s, as explained by Prieto Ramos (2014, 272), the field of legal translation studies entered a “period of consolidation and expansion”. It may have extended to a legal context (the Swiss one) that had scarcely been investigated up until then⁵: it is precisely in 2005 that an important volume, *Lingua e diritto. La presenza della lingua italiana nel diritto svizzero* (edited by Borghi), was published. This volume gathers the contributions of both linguists and legal experts and focuses especially on the legal and institutional use of Italian in Ticino. It discusses the peculiarities of different text genres (enactments, judgments, notary deeds) and sheds light on the conservative and at the same time innovative nature of Swiss legal Italian and, more broadly, on its status (and constraints) within a trilingual institutional system.

A few years later, in 2011, a second volume was issued, *Mehrsprachige Gesetzgebung in der Schweiz. Juristisch-linguistische Untersuchungen von mehrsprachigen Rechtstexten des Bundes und der Kantone* (edited by Schweizer and Borghi), which also adopts an interdisciplinary approach to investigating the process of multilingual lawmaking in Switzerland. To this end, all national languages are included in the various analyses. A case study on the “Vocational and

⁵ Egger and Ferrari (2016, 507) confirm that the interest in carrying out investigations into the quality of Swiss legal Italian dates back to mid-2000.

Professional Education and Training Act”⁶ investigates the extent to which correspondence between language versions and comprehensibility are features of Swiss federal legislation. The study also takes into account the implementation of this federal act in multilingual cantons. Some “theses and recommendations” are proposed at the end of the book and some suggestions for improvement to the current process of multilingual lawmaking are put forward.

A third volume, edited in 2013 by Egger, Ferrari and Lala (*Le forme linguistiche dell’ufficialità. L’italiano giuridico e amministrativo della Confederazione Svizzera*) deals specifically with the variety of Italian used in Bern by the federal legislative and executive bodies. Beside law, other text genres that allow for smooth State-to-citizen communication (like press releases and websites) are analyzed, pointing out their strengths and weaknesses.

More recently, in 2019, Egger published his monograph *A norma di (chi) legge. Peculiarità dell’italiano federale*. It is an invaluable reference that offers an unprecedented overview of institutional translation in Switzerland. It focuses on Italian as an official translation language and, in particular, on the specificities of translating the law.

In addition to these four volumes, some conference papers and journal contributions have started to appear over the last two decades. The authors of these publications often have a direct link to the practice: their profile as “federal translators” and researchers allow them to offer a very insightful perspective on the impact of translation on the quality of Swiss legislation (cf. for instance Egger and Grandi 2008 on the translation of the Swiss Criminal Procedure Code and Egger 2015 on the evolution of Swiss legislation in Italian since the Nineties).

As briefly outlined in this section, there is a strong interest in this field and there is no doubt a need for further empirical studies. *Ad hoc* resources are fundamental to this aim. The following section illustrates the key role of corpora in LTS and offers a review of the existing resources that can be used to explore linguistic and translation questions related to Swiss legal and institutional texts.

⁶ Cf. Vocational and Professional Education and Training Act of 13 December 2002 (status as of 1 January 2019, CC 412.10).

4. Corpora in LTS and existing corpora of Swiss legal and institutional texts

Among the research methods in legal translation studies⁷, corpora have gained a central position in recent years (Biel 2018, 34). The LETRINT project and the *Eurolect Observatory Project*, just to mention two of them, confirm this assumption.⁸

The first one, directed by Fernando Prieto Ramos, aims at creating an empirical model to assess quality in legal translation by evaluating products, processes and competences. In the context of this project, different corpora comprising texts issued by the EU, UN and WTO have been compiled.

The second, directed by Laura Mori and currently in its second phase, is concerned with the analysis of EU legislative varieties in 12 languages. For each language, several corpora have been compiled and contrasted, comprising EU directives, their implementation in national legislation and domestic law. Within this large project, which involves 28 researchers from 19 universities, a project on Swiss legislation has been launched by Annarita Felici. It aims at investigating the impact of Eurolects on Swiss legislation through the bilateral agreements between Switzerland and the EU. This corpus-based study will certainly fill a research gap and shed light on unexplored questions related to the Swiss legal languages.

Over the last years, further corpora have been compiled to carry out linguistic research on Swiss legal and institutional texts. The corpus “Bundesblatt / Feuille fédérale / Foglio federale” (cf. Elmiger, Tunger and Schaeffer-Lacroix 2017, 27-32) contains all issues of the Federal Gazette. It covers a very wide time span: from 1849 to 2014 for German and French, and from 1971 to 2014 for Italian, and is available on the platform *CQPweb*. However, it should be mentioned that the Federal Gazette has only been issued in electronic format

⁷ See Biel (2017, 79-82) for a “multi-perspective research framework for legal translation” in which four dimensions (context of production and reception, participants, processes and products) and its different components and research methods are presented.

⁸ Cf. <http://p3.snf.ch/project-157797> for the LETRINT project, and Mori (2018) as well as the website <https://www.unint.eu/en/research/research-projects/33-page/490-eurolect-observatory-project.html> for the Eurolect Observatory Project (accessed on April 16, 2019).

since 1998. All the previous issues have been digitalized by the Swiss Federal Archives (SFA) and are made available both as scanned .pdf files and as output of an OCR processing. The latter version was used to compile this corpus and no cleansing was carried out.

For Swiss legal texts, two further corpora are available (both described in Höfler and Piotrowski 2011), DS21 and SLC. DS21 is a corpus involving a wide range of historical legal texts produced in the territories of today's Swiss Confederation. It covers ten centuries, from the early Middle Ages to the end of the Old Confederation in 1798. It comprises texts in the three official languages, and does not contain any digitalization errors. SLC is a corpus of current Swiss legislation that has been compiled on the basis of the Classified Compilation of Swiss Federal Legislation. The acts comprised in this corpus are also free of errors and contain a series of annotations (such as structural units and POS-tagging). SLC is a trilingual, comparable, aligned corpus.

However, if one wants to carry out micro-diachronic linguistic research on normative acts in recent history and get a fine-grained image of how Swiss legal Italian has evolved, none of these resources seems adequate. This is why LEX.CH.IT has been compiled.

5. Design and compilation of LEX.CH.IT

This section presents the compilation criteria of LEX.CH.IT. For the sake of intelligibility, the description of the corpus is organized in different subsections. First of all, the overall aim and the type of corpus is presented, then the questions of representativeness and size are addressed and a description of the texts included in the corpus is offered. Finally, some technical aspects are tackled.

5.1 Aim and type of corpus

LEX.CH.IT is intended as a resource to identify the evolution of Swiss legal Italian throughout the “new language regime” (cf. section 1). It is a *diachronic* corpus (Baker 2006, 29-30), which covers a time

span of almost five decades (1974-2018). It is *monolingual* and only includes texts in Italian. It has to be understood as a *specialized* corpus: it is not a general-purpose corpus, as it only covers a specific area (Swiss legislation), thereby allowing the characteristics of a language for specific purposes to be investigated.

5.2 Representativeness and size

It is widely recognized that a corpus should be representative of the language variety being investigated and that representativeness “determines the research questions that can be addressed and the generalizability of the results of the research” (Biber, Conrad and Reppen 1998, 246). An interrelated question concerns size. Drawing on the title of a paper by Corpas Pastor and Seghiri (2010), “size matters” in corpus research. However, as the authors themselves (*ibid.*, 114-125) point out, there is still no consensus in the literature as to what size is ideal for a corpus to be representative. Several authors agree on the fact that LSPs can be investigated through “smaller” corpora (cf. for instance Bowker and Pearson 2002, 48). Quantifying the adjective “small”, however, is quite a hard task.

In the case of LEX.CH.IT, both problematic aspects, representativeness and corpus size, have been dealt with by choosing a sample equivalent to the population. In other words, a legal text genre was chosen, i.e. federal acts (for the rationale behind this choice, see the following subsection), and all texts that belong to this genre were collected in the corpus, for a total of 366 acts and over 1.1 million tokens.

5.3 Texts included in the corpus

In Switzerland, three bodies can have legislative functions: the Parliament, the Government and the Federal Court (cf. FOJ 2007, 235-236). For the purposes of this corpus, the choice was limited to the acts issued by the first one, i.e. the Federal Assembly, which is the true legislative body. The Parliament carries out its legislative activity

by means of four types of acts: the Constitution, federal acts, ordinances and federal decrees (*ibid.*, 237-241). The first text genre does not allow any in-depth diachronic investigations, as it is not rewritten from scratch on a regular basis. The choice to focus on federal acts alone (and exclude ordinances and decrees) was made based on the “principle of importance”: “Le critère essentiel pour décider si une règle de droit doit revêtir la forme de la loi fédérale est celui de **l’importance de son contenu**” (*ibid.*, 239).⁹

All the federal acts published in the Official Compilation of Federal Legislation during the chosen time span were therefore included in LEX.CH.IT. In this sense, one could also state that it is a *domain-complete* corpus: it does not only focus on select areas of law.

All the issues of the Official Compilation were scanned manually and only the federal acts were downloaded. The aim was to gather the first version of all federal acts. That is why the acts that amend or repeal acts previously in force were not taken into account.¹⁰ This is the advantage of compiling such a corpus, compared to using a corpus of current legislation: a normative act is often amended, sometimes quite significantly, throughout its life cycle and the only way of preserving the diachronic dimension is to deal with acts as they were first published. Following this principle, each act in the corpus has been scanned and the articles that amend and repeal previous acts (or parts of previous acts) have been excluded.

More generally, only the enacting parts of each federal act have been included in the corpus. As explained in following section, its first application is a study on the level of readability and comprehensibility of Swiss legislation in Italian. Preambles, annexes, date and signatures are not deemed very relevant in this respect and therefore have not been considered. In particular, preambles are usually very short in Swiss normative acts and only include a list of references to other legal sources and documents related to the legislative procedure. Annexes often contain tables of data or amendments and repeals. All those elements could have a negative impact, especially on quantitative analyses aimed, for instance, at

⁹ In other words, in the Swiss legal system, federal acts have the highest rank in the hierarchy of norms after the Constitution.

¹⁰ A federal act that amends a previous act issued, for instance, ten years before, could jeopardize the aim of this corpus, i.e. observing the diachronic evolution of Swiss legal Italian. In fact, the new act might be directly influenced by the one that it amends and therefore not be fully representative of the period in which it was issued.

determining the mean sentence length or the distribution of some specific syntactic features such as subordination vs. coordination.

5.4 On the technical side

Since September 1998, all issues of the Official Compilation of Federal Legislation have been published online in digital format. These texts can be easily downloaded in .pdf format and converted to .txt. The only cleansing that has been done is the removal of headers, footnotes and footers, as well as the parts of text that have been excluded from the corpus (cf. 5.3).

All previous issues of the Official Compilation (from 1974 until August 1998) have been digitalized by the Swiss Federal Archives (as was the case with the Federal Gazette, as discussed in section 4). The output, available in .pdf and .txt format, demanded much more effort to cleanse. All the texts had to be scanned manually in order to correct misrecognized words. Ensuring that all texts making up the corpus are equally clean is key to carrying out comparisons within the corpus and avoiding distorted results. If the texts before 1998 had a higher rate of noise compared to those after 1998, the interpretation of certain results would be much more problematic.

All the texts were saved in a folder as .txt files in UTF8 format. They have not been merged, so that – depending on the research question that one wants to test on the corpus – they can be sorted out and clustered in different groups.

6. Applications of LEX.CH.IT: clarity of Swiss legislation

The first application of LEX.CH.IT is a doctoral research project in which the author intends to investigate the level of clarity¹¹ of Swiss

¹¹ In this paper, the term “clarity” is used to refer to “linguistic clarity” and not to “normative clarity”. For a description of the two concepts, cf. Flückiger and Grodecki (2017).

legislation from a diachronic perspective. This project aims at filling a research gap: until now, few studies have focused on the level of clarity of Swiss legislation – especially when it comes to Italian. The need for empirical evidence in this field has been acknowledged by both linguists (e.g. Berruto 2012, 13-14) and legal experts (e.g. Uhlmann 2014, 177-178). They have encouraged the development of corpus studies to shed light on, respectively, the evolution of the quality of Swiss legal Italian and the overall level of clarity of Swiss legislation. This is the thematic focus of this research project. Subsection 6.1 offers a short description of this ongoing project. Its results, however, will be discussed in details elsewhere, as the main aim of this paper is to draw a profile of LEX.CH.IT.

6.1 Outline of the project

The last decades saw the emergence of several plain language initiatives throughout the world.¹² Meanwhile, as already mentioned in section 1, several measures were adopted in Switzerland to guarantee the quality of legal and institutional texts. At the same time, Switzerland has a long tradition of plain language (Flückiger/Delley 2006, 136-8), as shown, for instance, by the founding father of the Swiss Civil Code, Eugen Huber, who at the beginning of the 20th century introduced the principle of *législation populaire* in the field of legal drafting. Bearing these two elements in mind, the project intends to answer three main research questions:

- RQ1: Is “clarity” a feature of Swiss legal Italian?
- RQ2: What is the evolution of the level of clarity throughout this period?
- RQ3: What is the influence of translation on the level of clarity?

¹² Sweden was a forerunner in the field of legislative plain language and is still very active today (cf. the activity of the Språkrådet, the Language Council of Sweden, <https://www.sprakochfolkminnen.se/sprak/klarsprak/in-english.html>, accessed May 28, 2019). The Anglophone countries have also made a significant contribution (Cutts 2013). At the EU level, the *Fight the Fog* campaign must be mentioned (Wagner, Bech and Martínez 2012, 72-73). See also DGT (2013) for an overview of different plain language initiatives in legal and institutional contexts in both European and non-European countries.

The methodology of this research project has been developed by drawing on the definition of (linguistic) clarity offered by Piemontese (1996, 79-122), which has been widely accepted in the literature. Clarity can be understood as the sum of “readability” and “comprehensibility”. Readability, on one hand, is a quantitative measure, which describes the surface of a text (sentence and word length, lexical and (morpho)syntactic complexity). Comprehensibility, on the other, is a qualitative measure, which describes the deep structure of a text, its logical organization, its cohesion and coherence, its textuality. As such, it is also a “relational feature” that varies according to the (asymmetry between the) actors involved: the producers (i.e. the federal institutions) on one hand and the receivers (i.e. the readers, ranging from legal experts to laypersons/citizens) on the other. The willingness of the producer to express herself in a clear way and the interest of the reader in understanding the text, as well as her knowledge of the subject, determine different levels of comprehensibility.¹³

According to this definition of clarity, this research project has been divided into two methodological phases: the first assesses the level of readability of Swiss normative acts in Italian and the second their comprehensibility.

In the first phase, the author has identified some indicators of readability, especially thanks to a review of the guidelines on how to write clear legal and institutional texts in Italian (Fioritto 1997, Cortelazzo and Pellegrino 2003, Franceschini and Gigli 2003, Raso 2005). Another important source of inspiration were the previous studies that assessed the readability of legal and institutional texts in other italophone contexts (Piemontese 2000, Venturi 2012, Brunato 2014). The indicators of readability that have been identified concern:

- the superficial level of a text (Gulpease index¹⁴, sentence and word length)

¹³ As explained further below, the relational component of comprehensibility cannot be investigated using a corpus-based methodology, and therefore will not be part of this project.

¹⁴ The Gulpease Index (presented in Lucisano and Piemontese 1988) is a readability formula for the Italian language that is similar to the Flesh, the Gunning Fog or LIX indexes. It takes two parameters into account, sentence and word length, to estimate the level of readability of a text for a specific group of readers.

- relevant lexical features (percentage of words belonging to the Basic Italian Vocabulary¹⁵, distribution of some archaic and foreign words, lexical richness)
- morphosyntactic aspects (part of speech distribution, lexical density, noun/verb ratio as a measure of informativeness, archaic morphological features)
- syntactic complexity (structure of the syntactic tree, use of subordination, nominal modification, use of passive)

In order to check the diachronic evolution of these indicators, different text analysis and natural language processing tools are needed. In particular, POS-taggers (Treetagger/TagAnt), concordancers (AntConc), text analysis and text mining tools (Taltac2). A crucial NLP tool is READ-IT, developed by the Institute of Computational Linguistics in Pisa. It performs several automated analyses on an input text (tokenization, lemmatization, POS-analysis, disambiguation and syntactic annotation) and gives information about several readability indicators.¹⁶ The interpretation of the quantitative data relies on different descriptive and inferential statistics that make it possible to compare distributions and assess the significance of specific evolutions. An answer to RQ2 can therefore be provided. Moreover, in this phase, some comparisons with other corpora of legal Italian will be key to understanding to what extent clarity is a feature of Swiss legal Italian, thus also responding to RQ1.

The second phase of this project tackles questions related to comprehensibility.¹⁷ As already mentioned, comprehensibility is a qualitative measure and different research methods are therefore needed. Carrying out qualitative analyses of 366 texts would be impossible. That is why a subcorpus will be created out of LEX.CH.IT and used as basis for the second phase. The selection of the texts to be included in the subcorpus will be done through an “opportunistic sampling”. The aim is to have a number of texts that represent the

¹⁵ Originally developed by De Mauro in the Seventies, it was “updated” in 2016 using a more complete and refined methodology (De Mauro 2016). It includes a list of around 7,000 easy-to-understand words.

¹⁶ Thank you to Dominique Brunato and Giulia Venturi of the ILC (Pisa) for providing the author with a very complete linguistic monitoring of LEX.CH.IT performed through READ-IT.

¹⁷ Some studies on the comprehensibility of Swiss legislation in German are available in the literature (cf. e.g. Höfler 2016). Comparable studies on Swiss legislation in Italian are missing.

whole time span considered in this diachronic study, chosen according to the results of the quantitative analyses carried out in the first phase. In the second phase, a manual annotation will be carried out through a QDA tool and will focus on the textual level and the information structure of the texts selected. This will make it possible to confirm and enrich the answers to RQs 1 and 2. Moreover, the German and French counterparts of the selected acts will be also taken into account, thus allowing for an interlingual comparison, which will help understand if and to what extent the translation process can have an effect on the level of clarity of Swiss legislation in Italian, thus responding to RQ3.

6.2 Perspectives

Aside from this first application of LEX.CH.IT, many other research questions may be addressed. In the future, LEX.CH.IT could be used as a basis for the compilation of parallel and comparable corpora. LEX.CH.DE and LEX.CH.FR, for instance, could be created in order to allow for further interlingual comparisons.

Besides comparisons with other corpora of legal Italian (which have been partially planned in this research project as well), it could be interesting to build corpora of other institutional text genres (press releases, web pages, public competition announcements etc.). In fact, legislation is often referred to as the source of the entire institutional and administrative communication (Cortelazzo 2011, 7). It would be interesting to verify if the level of accessibility varies according to the institutional text genre and to what extent legislation has an influence on other forms of State-to-citizen communication.

Finally, all product-based research carried out on LEX.CH.IT could be enriched with participant-based research. As already mentioned, comprehensibility is also a relational feature and this is why involving producers and receivers could provide new insight and interpretations. In order to triangulate the corpus-based results, focus groups could be organized with some federal translators in order to investigate their stance towards clarity. Moreover, comprehensibility

tests could be carried out on a sample of citizens to further validate (or relativize) the interpretations of the textual analyses.¹⁸

7. Concluding remarks

In terms of the four-fold division of languages presented by McEnery and Hardie (2012, 12)¹⁹, which identifies official majority, official minority, unofficial and endangered languages, one can state that LEX.CH.IT contains texts that fit in the second category. The authors claim that a corpus that focuses on one of the first two categories is easier to compile, thanks to the availability of corpus data, e.g. online. This assumption has been confirmed throughout the compilation of LEX.CH.IT: the availability of digital textual data was of paramount importance for creating the corpus. Yet a lot of work (section 5) was needed to transform unstructured data available online into a structured corpus that can be processed through the most common corpus analysis tools.

When it comes to official minority languages, such as Swiss Italian (section 2), it is often more difficult to find ready-to-use corpora. Although some corpora have started to appear over the last few years, more resources are needed to provide for and further expand on the interest in carrying out research in this field (section 3). In light of the expansion of research involving corpora not only in linguistics and translation and interpreting studies (Bernardini and Russo 2018), but also legal and institutional translation studies (see section 4), creating such a resource is crucial for prolific research into a minority language (variety).

As explained in detail in section 6, LEX.CH.IT was originally compiled to investigate the level of clarity of Swiss federal acts issued over the last five decades. However, the author of this paper, or other researchers interested in investigating the peculiarities of Swiss legal Italian, could use it to develop new research projects without needing

¹⁸ Cf. Viale (2008) for a research project on the aptitude for clarity of Italian civil servants in charge of producing administrative texts and Curtotti et al. (2015) for a study on the level of clarity perceived by an audience reading US legislation.

¹⁹ Drawing on a previous paper by McEnery and Ostler (2000).

to spend a considerable amount of time compiling another corpus. In fact, designing and compiling a corpus is valuable research in itself.

Bibliography

- Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London/New York: Continuum.
- Bernardini, Silvia, and Mariachiara Russo. 2018. "Corpus linguistics, translation and interpreting." In *The Routledge Handbook of Translation Studies and Linguistics*, edited by Kirsten Malmkjær, 342-356. Oxford: Routledge.
- Berruto, Gaetano. 1984. "Appunti sull'italiano elvetico." *Studi linguistici italiani* 10:76-108.
- Berruto, Gaetano. 2012. "L'italiano degli svizzeri." Paper presented at Nuit des Langues, November 8, in Bern, Switzerland. <https://m4.ti.ch/fileadmin/DECS/DCSU/AC/OLSI/documenti/BERRUTO-2012-Italiano-degli-svizzeri-Berna-conferenza.pdf> (accessed May 8, 2019).
- Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biel, Łucja. 2017. "Researching Legal Translation: A Multi-Perspective and Mixed-method Framework for Legal Translation." *Revista de Llengua i Dret* 68:76-88.
- Biel, Łucja. 2018. "Corpora in Institutional Legal Translation: Small Steps and the Big Picture." In *Institutional Translation for International Governance. Enhancing Quality in Multilingual Legal Communication*, edited by Fernando Prieto Ramos, 25-36. London: Bloomsbury.
- Biscossa, Giuseppe. 1968. "Evoluzione della lingua italiana nel Ticino." *Il Veltro* 6:497-541.
- Borghi, Marco, ed. 2005. *Lingua e diritto. La presenza della lingua italiana nel diritto svizzero*. Basel: Helbing & Lichtenhahn.
- Bowker, Lynne, and Jennifer Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

- Brunato, Dominique. 2014. "Complessità necessaria o stereotipi del burocrate? Un'indagine sulla leggibilità del linguaggio amministrativo da una prospettiva linguistico-computazionale." In *La Lingua Variabile nei testi letterari, artistici e funzionali contemporanei: analisi, interpretazione, traduzione, Atti del XIII Congresso della SILFI, Palermo, 22-24 settembre 2014*, edited by Centro di studi filologici e linguistici siciliani, Università degli studi di Palermo.
- Corpas Pastor, Gloria, and Miriam Seghiri. 2010. "Size Matters: A Quantitative Approach To Corpus Representativeness." In *Lengua, traducción, recepción. En honor de Julio César Santoyo/ Language, translation, reception. To honor Julio César Santoyo*, edited by Rosa Rabadán, 111-146. León: Universidad de León, Área de Publicaciones.
- Cortelazzo, Michele A. 2011. Introduction to *Il linguaggio e la qualità delle leggi*, edited by Raffaele Libertini, 7-16. Padua: CLEUP.
- Cortelazzo, Michele A., and Federica Pellegrino. 2003. *Guida alla scrittura istituzionale*. Roma/Bari: Laterza.
- Cutts, Martin. 2013. *Oxford Guide to Plain English*. 4 ed. Oxford: Oxford University Press.
- De Mauro, Tullio. 2016. Il Nuovo vocabolario di base della lingua italiana. *Internazionale*, <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana> (accessed March 18, 2019).
- DGT (Directorate-General for Translation, European Commission). 2013. *Document Quality Control in Public Administrations and International Organisations, Studies on Translation and Multilingualism*. Luxembourg: Publications Office of the European Union.
- Egger, Jean-Luc. 2015. "25 anni di legislazione federale in lingua italiana: alcuni spunti di riflessione." *LeGes* 26 (1):151-171.
- Egger, Jean-Luc. 2019. *A norma di (chi) legge. Peculiarità dell'italiano federale*. Milan: Giuffrè Francis Lefebvre.
- Egger, Jean-Luc, and Angela Ferrari. 2016. "L'italiano federale svizzero: elementi per una ricognizione." *Studi italiani di linguistica teorica e applicata* 45 (3):499-523.
- Egger, Jean-Luc, and Filippo Grandi. 2008. "Il nuovo Codice di procedura penale: un cantiere anche linguistico." *LeGes* 19 (1):31-72.

- Egger, Jean-Luc, and Filippo Grandi. 2013. "Italiano giuridico federale: un dispaccio dal fronte." In *Le forme linguistiche dell'ufficialità. L'italiano giuridico e amministrativo della Confederazione Svizzera*, edited by Jean-Luc Egger, Angela Ferrari and Letizia Lala, 213-242. Bellinzona: Casagrande.
- Egger, Jean-Luc, Angela Ferrari, and Letizia Lala, eds. 2013a. *Le forme linguistiche dell'ufficialità. L'italiano giuridico e amministrativo della Confederazione Svizzera*. Bellinzona: Casagrande.
- Egger, Jean-Luc, Angela Ferrari, and Letizia Lala. 2013b. Introduction to *Le forme linguistiche dell'ufficialità. L'italiano giuridico e amministrativo della Confederazione Svizzera*, edited by Jean-Luc Egger, Angela Ferrari and Letizia Lala, 11-15. Bellinzona: Casagrande.
- Elmiger, Daniel, Verena Tunger, and Eva Schaeffer-Lacroix. 2017. *Geschlechtergerechte Behördentexte. Linguistische Untersuchungen und Stimmen zur Umsetzung in der mehrsprachigen Schweiz. Forschungsbericht*. Geneva.
- Fioritto, Alfredo, ed. 1997. *Manuale di Stile. Strumenti per semplificare il linguaggio delle amministrazioni pubbliche*. Bologna: Il Mulino.
- Flückiger, Alexandre, and Jean-Daniel Delley. 2006. "L'élaboration rationnelle du droit privé : de la codification à la légistique." In *Le législateur et le droit privé, Mélanges en l'honneur de Gilles Petitpierre*, edited by Christine Chappuis, Bénédicte Foëx and Luc Thevenoz, 123-143. Geneva: Schulthess.
- Flückiger, Alexandre, and Stéphane Grodecki. 2017. "La clarté : un nouveau principe constitutionnel ?" *Revue de droit suisse* 136 (1):36-62.
- FOJ (Federal Office of Justice). 2007. *Guide de législation: Guide pour l'élaboration de la législation fédérale*. Bern: FOJ.
- Franceschini, Fabrizio, and Sara Gigli, eds. 2003. *Manuale di scrittura amministrativa*. Rome: Agenzia delle Entrate.
- Höfler, Stefan. 2016. "Die Informationsstruktur von Rechtssätzen und ihre Bedeutung für die Gesetzesredaktion." *LeGes* 27 (2):225-251.
- Höfler, Stefan, and Michael Piotrowski. 2011. "Building corpora for the philological study of Swiss legal texts." *Journal for Language Technology and Computational Linguistics* 26 (2):77-89. doi: 10.5167/uzh-54761.

- Lucisano, Pietro, and Maria Emanuela Piemontese. 1988. "Gulpease. Una formula per la predizione della difficoltà dei testi in lingua italiana." *Scuola e Città* 3:57-68.
- Lurati, Ottavio. 1976. *Dialetto e italiano regionale nella Svizzera italiana*. Lugano: Banca Solari & Blum.
- McEnery, Tony, and Andrew Hardie. 2012. *Corpus linguistics: Method, Theory and Practice*. Cambridge/New York: Cambridge University Press.
- McEnery, Tony, and Nick Ostler. 2000. "A New Agenda for Corpus Linguistics. Working with all of the World's Languages." *Literary and Linguistic Computing* 15 (4):403-419.
- Moretti, Bruno, ed. 2004. *La terza lingua. Volume primo, norma e varietà di lingua in Ticino*. Locarno: Dadò.
- Moretti, Bruno, ed. 2005. *La terza lingua. Volume secondo, dati statistici e «varietà dinamiche»*. Locarno: Dadò.
- Mori, Laura, ed. 2018. *Observing Eurolects. Corpus Analysis of Linguistic Variation*. Amsterdam: Benjamins Publishing House.
- Pandolfi, Elena Maria. 2009. *LIPSI: Lessico di frequenza dell'italiano parlato nella Svizzera italiana*. Bellinzona: Osservatorio Linguistico della Svizzera Italiana.
- Piemontese, Maria Emanuela. 1996. *Capire e farsi capire. Teorie e tecniche della scrittura controllata*. Naples: Tecnodid.
- Piemontese, Maria Emanuela. 2000. "Leggibilità e comprensibilità delle leggi italiane. Alcune osservazioni quantitative e qualitative." In *Linguistica giuridica italiana e tedesca/Rechtlinguistik des Deutschen und Italienischen*, edited by Daniela Veronesi, 103-118. Padua: Unipress.
- Pini, Verio. 2017. *Anche in italiano! 100 anni di lingua italiana nella cultura politica svizzera*. Bellinzona: Casagrande.
- Prieto Ramos, Fernando. 2014. "Legal Translation Studies as Interdiscipline: Scope and Evolution." *Meta* 59 (2):260-277. doi: 10.7202/1027475ar.
- Raso, Tommaso. 2005. *La scrittura burocratica. La lingua e l'organizzazione del testo*. Rome: Carocci.
- Schweizer, Rainer J., and Marco Borghi, eds. 2011. *Mehrsprachige Gesetzgebung in der Schweiz. Juristisch-linguistische Untersuchungen von mehrsprachigen Rechtstexten des Bundes und der Kantone*. Zurich: Dike.

- Schweizer, Rainer J., Marco Borghi, and et al. 2011. “Legislazione plurilingue in Svizzera: tesi e raccomandazioni.” In *Mehrsprachige Gesetzgebung in der Schweiz. Juristisch-linguistische Untersuchungen von mehrsprachigen Rechtstexten des Bundes und der Kantone*, edited by Rainer J. Schweizer and Marco Borghi, 425-440. Zurich: Dike.
- Snozzi, Alfredo. 2005. “L’italiano nella legislazione federale svizzera.” In *Lingua e diritto. La presenza della lingua italiana nel diritto svizzero*, edited by Marco Borghi, 317-329. Basel: Helbing & Lichtenhahn.
- Uhlmann, Felix. 2014. “Qualität der Gesetzgebung: Wünsche an die Empirie.” In *Vom Wert einer guten Gesetzgebung*, edited by Alain Griffel, 171-181. Bern: Stämpfli Verlag.
- Venturi, Giulia. 2012. “Investigating Legal Language Peculiarities across Different Types of Italian Legal Texts: an NLP-based Approach.” In *Bridging the Gap(s) between Language and the Law: Proceedings of the 3rd European Conference of the International Association of Forensic Linguistics, Porto, 15-18 October 2012*, edited by Rui Sousa-Silva, Rita Faria, Núria Gavalda and Belinda Maia, 139-156. Porto: Faculdade de Letras da Universidade do Porto.
- Wagner, Emma, Svend Bech, and Jesús M. Martínez. 2012. *Translating for the European Union Institutions*. 2nd ed. Manchester: St. Jerome.
- Zwicky, Roman, and Daniel Kübler. 2018. *Topkader und Mehrsprachigkeit in der Bundesverwaltung* Vol. 13, *Studienberichte des Zentrums für Demokratie Aarau*. Aarau: ZDA.

Legal texts

- Federal Act on the Compilations of Federal Legislation and the Federal Gazette (Publications Act, PublA) of 18 June 2004 (Status as of 26 November 2018), Classified Compilation 170.512, <https://www.admin.ch/opc/en/classified-compilation/20031819/index.html>.
- Federal Act on the National Languages and Understanding between the Linguistic Communities (Languages Act, LangA) of 5 October 2007 (Status as 1 January 2017), Classified

Compilation 441.1, <https://www.admin.ch/opc/en/classified-compilation/20062545/index.html>.

Federal Act on Vocational and Professional Education and Training (Vocational and Professional Education and Training Act, VPETA) of 13 December 2002 (Status as of 1 January 2019),
Classified Compilation 412.10,
<https://www.admin.ch/opc/en/classified-compilation/20001860/index.html>.

Web pages

Eurolect Observatory Project:
<https://www.unint.eu/en/research/research-projects/33-page/490-eurolect-observatory-project.html> (accessed April 16, 2019).

LETRINT project: <http://p3.snf.ch/project-157797> (accessed April 16, 2019).

Språkrådet (Language Council of Sweden):
<https://www.sprakochfolkminnen.se/sprak/klarsprak/in-english.html> (accessed May 28, 2019).