



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue Moderne per la Comunicazione e la Cooperazione Internazionale
Classe LM-38

Tesi di Laurea

*Design of an Open Source
plug-in in R to Manage a Medical Termbase*

Relatore
Prof. Giorgio Maria Di Nunzio

Laureando
Roberta Rovelli
n° matr.1155437 / LMLCC

Anno Accademico 2018 / 2019

ABSTRACT

This dissertation aims to explore the differences between the main translation approaches available for the medical web sites and to present the design and implementation of an open source plug-in written in R for managing a medical termbase. The existing tools available for translators and localizers are presented, discussing their advantages and disadvantages. Moreover, the peculiarities of scientific, technical and medical translation are analysed, and the plug-in is presented as a new tool for the translation of medical websites and documentation.

ACKNOWLEDGEMENTS

Firstly, I would like Dr Giorgio Maria Di Nunzio, assistant professor at the Department of Information Engineering, whose advice on the creation of my plug-in have been greatly important. His experience with programming languages and termbases has helped me very much.

Moreover, I would like to thank professor Maria Teresa Musacchio of the Department of Linguistic and Literary Studies. Her knowledge in the translation field has been a great aid in the comprehension of the translation of medical texts and of the most used translation tools.

Lastly, I would like to thank Federica Vezzani, the PhD student of the Department of Linguistic and Literary Studies, who has created a medical termbase from which I have extracted terms and terminology useful for my plug-in. Her work has been important for the creation of my dataframes.

TABLE OF CONTENTS

INTRODUCTION	11
CHAPTER 1	15
1. A brief history of translation	15
1.1. In Antiquity	16
1.2. In the Middle Ages	17
1.3. In the 16th century	18
1.4. In the 17-18th centuries	19
1.5. In the 20th century	20
1.6. Today	21
1.6.1. Localization	22
1.6.2. Types of localization	25
1.6.2.1. User interface	25
1.6.2.2. Video games	27
1.6.2.3. Websites	28
1.7. Scientific websites localization	31
CHAPTER 2	35
2. Computer-Assisted Translation	35
2.1. CAT Tools	36
2.2. Terminology	41
2.3. How to create a Termbase	42
CHAPTER 3	47
3. Medical translation	47
3.1. Specialized languages	49
3.2. Medical web sites	50
CHAPTER 4	53
4. The Project	53

4.1. TriMED	54
4.2. The Design of the Plug-in.....	56
CHAPTER 5	61
5. My Plug-In at work	61
CONCLUSION	63
REFERENCES	65
RIASSUNTO IN ITALIANO	73

INTRODUCTION

This dissertation aims to explore the differences between the main approaches available for the localization of web sites and to present a new open source plug-in written in R and designed to manage medical termbases for the translation of medical documents. The existing tools available for translators and localizers are presented, discussing their advantages and disadvantages. The peculiarities of scientific, technical and medical translation are analysed, and the plug-in is presented as a new tool for the translation of medical websites. I have chosen this topic for my dissertation in order to combine my knowledge in Computer Science and my linguistic and translation competences acquired during the two years Master in Languages at the University of Padova. Computer Science has become very important for the translation and it is safe to say that it will keep playing an important role in the future of translation. Today internet, the Web, software and application are used by almost everybody daily both for works and during the free time, therefore a huge quantity of contexts are produced and most of them need to be localized to be suitable for all the users around the world. Since the beginning of the Globalization Era, our lives, the markets and the marketing strategies have changed.

Globalization, or the increased interconnectedness and interdependence of peoples and countries, is generally understood to include two inter-related elements: the opening of international borders to increasingly fast flows of goods, services, finance, people and ideas; and the changes in institutions and policies at national and international levels that facilitate or promote such flows.¹

Therefore, nowadays companies must face a new global market with many different types of clients and each market share requires a different approach. Consequently, using the right language and localize properly the products and texts is essential. However, not only the market field has experienced the influence of the new technologies but also the technical fields and the way information are shared have changed together with the evolution of Internet and the new technologies. For what concerns the scientific fields, many webpages are born with the need to share

¹ <http://www.who.int/topics/globalization/en/>

information and to allow dialogues and discussions between experts (or between experts and lay people) from all around the world. Indeed, nowadays scientific, medical and technical notions can be found on Internet. Pages like *Univadis*², *Sermo*³ and *Figure1*⁴ are designed to put into contact specialists with other experts around the world in order to promote the exchange of ideas between physicians and encourage in this way the progress. Other pages like *PubMed*⁵, *MedScape*⁶ and *UptoDate*⁷ are intended for less specialized users who want to access the knowledge of the professionals and be more informed about health issues and medical discoveries. Although the problem of translation of medicals texts has been existing for a long time and different solutions have been used to tackle this problem, the aim of this dissertation is to illustrate and explain a modern approach to translation of Web contents and, in particular, of the medical ones. Moreover, we present a software tool which I have designed and implemented, and which is intended to be put at disposal of translators and localizers in order to help them with their work.

Chapter One outlines the historical background of translation, focusing first on which type of translations were predominant in the past and which approaches were used to analyse and work with the texts. Translating is an antique art, it dates back to the ancient Egypt and it started with the translation of religious texts. The translation of the scriptures from Hebrew into Greek made around 250-130 BCE in Egypt is regarded as the first major translation project. This translation, also known as the *Septuagint*, is said to have been commissioned by the King of Egypt. Through the centuries many other types of text have been translated from the commercial to the medical ones, from literature books to technical manuals. For each category of texts, different approaches and different methods can be used. Therefore, different schools of thought have been born and they have individually made their contributions to the development of translation as an independent field of studies. For centuries discussions about the different approaches were carried on but only during the 20th century translation reached its independence as academic discipline thanks to James S. Holmes who in

² <https://www.univadis.it/>

³ <http://www.sermo.com/>

⁴ <https://figure1.com/>

⁵ <https://www.ncbi.nlm.nih.gov/pubmed/>

⁶ <https://www.medscape.com/>

⁷ <https://www.uptodate.com/home>

1972 coined the term *Translation Studies*, separating for the first time the translating and interpreting processes and defining more in details the area of competence of translation. From that moment until today, many studies and discussions have been done but, with the advent of internet, the focus has shifted to online contents, the new media and the software localization. Today, localization offers huge opportunities of work for professional translators. Nowadays, the products which need to be localized are countless, but the most important ones are related to Internet and computers like applications, programs, software packages and the websites where all the information and details about on- and offline products and services and all the medical and technical knowledges are hosted. Each typology of media requires a different localization approach. In addition, even on the internet there are different kinds of contents which need specific translation strategies. The localization of medical contents, in particular, demands that the translator has a previous knowledge of the medical field and of its terminology because this type of text usually involved specific vocabulary and terms that need to be translated accurately in order to avoid misunderstandings.

In chapter Two, Computer-Assisted Translation (CAT) and CAT Tools are explained in detail. The advantages and disadvantages of the main existing translation tools are discussed. The second part of the chapter is dedicated to the management of terminology, essential part of the translation process especially when it concerns technical texts. Inside translation environments, the terminology is often contained in termbases, thus it is explained how to create a high quality one.

Chapter Three starts with a brief description of the medical translation history. Then, a definition of *specialized languages* is given, and it is explained what makes them different from the general language. In fact, a *specialized language* is a subset of a language used only by a specific portion of the population because it requires special knowledges and it is related to a field of specialization. The speakers of this language can understand complex concepts using one term because they have already learned that notion and they do not need further explanations. The terms of the specialized languages and the abbreviations are usually two of the most common challenging parts of a text during the translation of a medical website.

Chapter Four presents the *TriMED* Project and the Plug-in. These projects derive from the need to overcome the shortcomings of the available software and from the desire to

offer an innovative tool for the medical translation field. In particular, *TriMED* is a digital library of terminological records designed to satisfy the needs of different categories of users within the healthcare field: patients, language professionals and physicians. Using this application, the users will be able to have a better comprehension of the medical vocabulary because it allows to find definitions, translations and the popular terms equivalent to the technical ones. *TriMED* was realized with the Shiny R package and so was the Plug-in here presented. Starting from the previous project, I have designed a tool specifically intended for medical translators. The application is based on two dataframes: an English one, containing medical technical terms and for each of them the corresponding popular term, the definition, the pronunciation, which part of speech is, the etymology, the synonyms and in which reliable websites has been cited and an Italian dataframe, containing the same pieces of information but translated. These data are essential during a localization process in order to understand how and in which contexts it is appropriate to use the technical term and in which the popular one. Retrieve these term records will be easy for localizers because I have designed a simple and intuitive user interface. Moreover, the application offers the possibility to download a file with *.tbx* format containing all the terminological cards. This format is the most used one for the exchange of termbases among professional translators and it is supported by all the main CAT tools. In addition, the users have the chance to add their own terminological entries to the existing dataframes before downloading the file. The last Chapter intends to demonstrate the advantages and the opportunities offered by the Plug-in during the translation of a medical content using a practical example.

CHAPTER 1

1. A brief history of translation

The art of translation has its roots deep in time. Indeed, the first proofs of texts transposition from a language to another date back to the ancient Egypt. Throughout the Centuries, the society and its priorities have changed. For the most ancient civilizations the most important and discussed topic was religion; therefore, many of the most ancient texts regarded Faith. Clearly the believers felt the need to have access to the original scriptures; however, only few people were enough educated and had the skills needed to read religious texts in their existing version. Consequently, a transposition from a language into another of these works was needed in order to make the texts available for a larger number of readers, and this marked the birth of translation. From that moment until today, the society has experienced a lot of changes and improvements across the various historical eras. The most discussed topics have changed and, as a consequence, the type of texts produced but the importance for the lay people of being able to read the contents has remained the same. The evolution of the civilizations has been also accompanied by the progression of the technologies in every field of our lives and translators have benefited from it. In fact, many tools have been invented throughout the centuries in order to facilitate their work. For instance, new methods for analysing texts, collecting terminology and accessing to previous translation works have been put at disposal of translators. In addition, many approaches to the texts have been tested, evaluated and discussed over the course of time. In fact, many people with dissimilar educational backgrounds and different attitudes have tried to translate different types of texts over the centuries. This has brought to the birth of translation as an independent subject of study and, at the same time, to the development different schools of thoughts.

All things considered, it is possible to claim that the art of translation has managed to evolve together with the progression of the society, remaining important and useful even nowadays. In particular, the modern technologies have a predominant role in the society and in every private and working field. As a result, the translation field has been

influenced by the new media and the Computer Science. Many software tools are now at disposal of translator and a huge share of contents and documents that need a translation is related to Internet and the new media. However, it is still important to know the evolution of translation in order to understand how the modern tools were born and why they are needed.

1.1. In Antiquity

The history of translation started in the ancient times. In the ancient Egypt the need of translation was already felt.

The Pentateuch was in fact translated into Greek in Alexandria during the third century BCE but probably to meet the needs of the Jewish community, in which Hebrew was hardly spoken any longer, rather than to fulfil the wishes of the king. The 200,000-member community, of which 100,000 lived in Alexandria, was a heterogeneous group of mercenaries, merchants, farmers, artisans and intellectuals. They were mainly Greek-speaking and had a special status within the city. They remained attached to their homeland, Palestine, sent taxes and offerings to the Temple of Jerusalem, and continued to make pilgrimages to the Promised Land. They had a closer affinity for Greek culture, however, and had even begun to use the Greek language in some of the synagogue rituals. Translation of the holy books was necessary to replace imperfect oral translations made in the synagogue (Delisle and Woodsworth 2012:157)

The first translations were in fact not related to trade, as it is often assumed but they were related to religion. As a matter of fact, even if the main reason for the contact between the different communities in ancient times was the trading of goods, the cult played already an important role inside the society. For instance, priests have an important role in the ancient Egypt Society. They were treated with respect because they were believed to care for the needs of the gods and usually, they were rich and powerful. Moreover, their education was higher than average. Precisely in Egypt, the first major translation project was composed. This is known as the *Septuagint* and it consisted in the translation of the *Scriptures* from Hebrew into Greek. This translation is said to have been commissioned by the King of Egypt between circa 250 and 130 BCE.

The legend says that a high priest selected seventy (or seventy-two) scholars. [...] Although the translation was originally purported to be a collective effort, a later legend told that the translators were assigned to separate quarters and were not allowed to communicate with each other. Yet they produced identical translations, proof that they were divinely inspired (Delisle and Woodsworth 2012:156).

Therefore, at the time, translation was considered as a tool of God to communicate with all the believers in the most direct way possible.

Even if the art of translation was just born, there were already the first discussions. In particular, during that period the discussion relate to “*sense-for-sense*” translation in opposition to “*word-for-word*” translation was opened for the first time.

There is ample evidence that early scholars of the Romans (Cicero, 46 BC and Horace, 20 BC) and later scholars of the ancient Chinese and Arab cultures seriously contemplated the work of translators and their products and, consequently, realized the ever-existing tug-of-war between form and content or, alternatively, what St Jerome early on (395 AD) called ‘*word-for-word*’ or ‘*sense-for-sense*’ (Farghal 2013:39).

Sense-for-sense translation considers the whole sentence as the unit of the translation, it means the translator should interpret the meaning of each whole sentence before moving on to the next and so on. *Word-for-word* translation, also known as *literal translation*, considers, instead, each word as the unit of translation, and it means that the translator should render each lexical item in sequence in the other language.

1.2. In the Middle Ages

Religion kept playing an important role in the society during the Middle Ages. At the time the Catholic Church had great power on the Western society. It is interesting to notice how the social and cultural structure influences the art of translation in each historical period and how the translated texts play a role in the community. For instance, because of the power of the Church, the main translations produced during the Middle Ages are religion-related.

In the Middle Ages, two centres of power commissioned translations: the State, represented by kings, caliphs, or princes; and the Roman Catholic Church, represented by the Pope himself, or by cardinals and bishops. In the Middle Ages, three main translation enterprises developed under the patronage of these centres of power. The first was the *Baghdad School* [...]. The second enterprise was the so-called “*Toledo School*”. There were two distinct translation movements in medieval Spain. During the first of these two periods, in the twelfth century, the scientific and philosophical heritage of the Greco-Arab world was translated from Arabic into Latin [...]. The third important translation enterprise of the Middle Ages occurred in the second half of the fourteenth century, when a group of scholars translated the *auctoritates*. These central texts of the Greco-Latin tradition were translated from Latin into French (Delisle and Woodsworth 2012:130-131).

The *Baghdad School* was composed by calligraphers, illustrators, transcribers and translators. The peculiarity of this movement is that they worked with non-Arabic

sources trying to produce and translate interesting manuscripts into Arabic. One of the most important works of this school is the translation from Greek into Arabic of the *De Materia Medica*. However, at the beginning of the Middle Ages, Latin was the main lingua franca and many philosophical and religious works were translated into that language from the *Toledo School*. Finally, the third group of scholars focused its work on the translation of the *auctoritates*, texts recognised as creditworthy because they were approved by the Church as witness to what was believed the real Truth.

Late medieval theologians, of course, accepted unquestioningly the fact of the divine authorship of the Bible, a fact which guaranteed its authority. As St Bonaventure put it in his commentary on the Sentences, if we wish to prove that the human *auctores* of Scripture told the truth, the short answer is that we know this to be the case because the Holy Spirit inspired them. It is the faith which the *auctores* received from God which makes their writings authentic. In one of the ordinary disputations which he held at Paris between 1276 and 1292, Henry of Ghent made the same point by reference to artistic skills. Where there is a craftsman (*artifex*) who directs and guides a work and another who works with his hands in accordance with rules conveyed from the craftsman, the latter is not said to be the *auctor* of the work but rather the former. Similarly, although the science of theology is described by men, it is directed and guided by God, who alone can properly be called its *auctor*. The *auctoritas* of this science is reducible to divine *auctoritas*; its sole source was God (Minnis 1984:82).

Holy Scripture was considered the *auctoritas* par excellence. The importance of religion in the society and in the translation field is underlined also by the fact that another noteworthy translation of that period is the *Wycliffe's Bible*, named after John Wycliffe, an English theologian who translated the Bible from Latin into English.

1.3. In the 16th century

Johannes Gutenberg invented in 1439 the printing press. Before that moment, all the books were written by hands; therefore, they were very expensive and it took a lot of time to produce each of them. For these reasons, they were considered luxury goods intended for rich and high-educated people. After the invention of the printing press and the rise of the middle class, books became affordable for many more people and for less educated classes. These new readers were not able to read texts written in Latin and thus they would like to read the works in their own language. Considering that before the press only the richest and educated could read the Holy Scripture, it is not a surprise that the Bible was the first best-seller book and that the two most remarkable translation works of that period are the *Tyndale New Testament* and the *Luther Bible*. The latter

take the name from Martin Luther, a German professor of theology, who translated the Bible from Hebrew and ancient Greek into German. This translation had a huge impact on the history of the Catholic Church, and it revolutionized the religious world in Europe.

Eisenstein argues that, while the medieval Catholic church was a prolific user of printing, the changes it wrought were outside the control of the church. The proliferation of different biblical texts eventually cast into doubt the existence of a single infallible text. This led to alternative interpretations such as Luther's, but the ability to publicize those interpretations by the same means of printing kept them from being crushed as were earlier heresies (Dewar 1998:13).

Therefore, translation had a huge responsibility towards the believers and the Church because different translations can lead to alternative interpretation of God's word and they could be even considered heresies.

Scholars have long recognized the essential role of the press in spreading Protestant doctrine. Luther himself, in fact, claimed that the invention of printing was a gift from God to reform His church. But Eisenstein argues that print did more than spread the Protestant Reformation: in an important sense, print caused the Reformation. Without access to the printed editions of biblical texts and church fathers, and the worrisome variants on crucial dogmatic issues they contain, Luther might never have been stimulated to develop his revolutionary new theology. And without accessibility to print, Luther might never have spread his ideas not only in the Latin of the scholarly community but also in the vernacular German of the lay community (Kingdon 1980:140).

These translations promoted the use of vernacular languages and contributed to the development of modern European languages. In particular, the *Luther Bible* was a turning point for the German language because it has influenced its evolution.

The desire for a unified national language was expressed to varying degrees among the different classes of society of the time, but was especially strong among the rising middle classes. Luther took the new communicative needs into account by using forms of speech that enjoyed widespread regional usage and also had a broad social basis. He was open to all influences, but to those of the Upper German linguistic area in particular (Delisle and Woodsworth 2012:39).

1.4. In the 17-18th centuries

After the Renaissance, the society changed. The Church was no more so powerful and the national movements gained more and more power.

Translators have helped to develop systems of writing. In their efforts to transpose certain fundamental texts from one culture to another, they have also had an impact on the evolution of language itself. [...] Translation is not an isolated phenomenon. Rather, it is associated with

certain major projects – nationalist, ideological and religious in nature – which often had the support of monarchs, aristocrats and institutions. The power of the sponsors, or the critical context in which translation took place, helped provide impetus and, in some cases, ammunition to translators, and gave legitimacy to their work. This in turn made it possible for them to make their mark on their language and culture (Delisle and Woodsworth 2012:21).

Thanks to the spread of the new national languages, there was a large development in the national literature and theatres. Consequently, most of the translation's works were no more related to religion but to books and plays.

Toward the second half of the eighteenth century, drama was rediscovered both as a literary and theatrical art form. Beginning in France, the new interest in theatre gradually spread from one culture to another. Despite strong resistance at first, French translators assumed the role of initiators within their own society. They also became pioneers outside their country. In fact, the history of the translation of Shakespeare can be written as a history of the international dissemination of French versions of Shakespeare, which was later followed by resistance to French models, when the same fate befell translations as original compositions (Delisle and Woodsworth 2012:63).

Once more, the strong bond between society and translation can be noticed. Translation had an influence on the changes of the society through the translated texts and through its influence on the vernacular languages. On the other side, the society required the translation of the genres on trend at the time (i.e. literature books and theatre's plays) and thus, most of the translation of that period are related to this field.

1.5. In the 20th century

Despite its long history, translation became an independent subject of study only in the last Century. In fact, only during the 20th century the discipline called *Translation Studies* was born.

The term "Translation Studies" suggested by Holmes in 1972 for adoption as the standard term for the discipline, is the one now unquestioningly used in English today, albeit implying translation only, as against the discipline of Interpreting Studies which has meanwhile developed beside it (Snell-Hornby 2006:55).

For the first time, translating and interpreting were seen as two separated subjects of study. Interpreting, together with translation, is part of the general concept of linguistic mediation; however, an interpreter is not a translator because these two professionals carry out different tasks, although people often get these professions confused with each other. In outline, translation is the transmission of a message from one language to another by means of a written channel of communication, whereas interpretation uses an oral channel. The translator has the time to write and rewrite the text before the public

can see it and the translation will last for years. On the other hand, the interpreter usually has almost no time to think about the message because the translated speech must be delivered simultaneously or just a little after the source speech. There are different types of interpretation (the main are simultaneous or consecutive) and of translation. New institutions and researches on the topic were developed. But other years gone by before a proper definition of the discipline and of its camps was established.

If it was the pragmatic turn of the 1970s that made the emergence of Translation Studies as an independent discipline possible, it was what later became known as the “cultural turn” of the 1980s that largely established its basic profile. The “cultural turn” is a name later given to a development that several of the various camps of the now generally (if grudgingly) accepted band of translation scholars like to claim as their own (Snell-Hornby 2006:60).

From that moment on, translation became a wide field of studies. New techniques and new approaches to different types of texts have been developed and discussed. Moreover, the main arguments of the translations are no more only religion, literature or theatre but many other fields, such as science, technical manuals, medicine, politics and business.

1.6. Today

Translation is nowadays considered as an independent academic discipline that involves many fields of study such as linguistics, semiotics, terminology and comparative literature.

Translation studies is the new academic discipline related to the study of the theory and phenomena of translation. By its nature it is multilingual and also interdisciplinary, encompassing languages, linguistics, communication studies, philosophy and a range of types of cultural studies (Munday 2001:1)

Moreover, these days translators choose a field of specialization like law, economic, medicine or literary translation and follow specialized courses to be trained accordingly to chosen field. Legal translation, for instance, has a legal value, and thus the translator is legally responsible for eventual mistakes in the interpretation of the source text and the eventual legal consequences. The translator who decides to operate in this field has to have a good knowledge of the law of the country in which he or she is working and of the country to which the text is destined to. Moreover, an excellent familiarity with

the legal terms is needed.

It is a paradox of the age of globalization that the demand for translation has grown despite the spread and dominance of English. This is especially true of specialized translation. Broadly speaking, specialized translation (*Fachübersetzen, traduction spécialisée, traduzione specializzata*) covers the specialist subject fields falling under non-literary translation, the best known of which include science and technology, economics, marketing, law, politics, medicine and mass media, most of which are dealt with in this volume as well as lesser researched areas such as maritime navigation, archaeology, geography and nutrigenomics (Sarcevic 2006:9).

Therefore, specialized courses have been designed. Literary translation, instead, requires prosodic skills and the ability to transmit the message being faithful to the intensity and aims of the original author, in a way that readers of the translated text should feel the same emotions and picture the same mental images of the readers of the original text. However, today there are new fields in which text localization is needed. A main revolution in the history of translation has indeed been the birth of Internet which has stimulated a worldwide production of new types of texts and media that need to be localized.

The most momentous and sudden changes in the world of translation and interpreting took place in the last decade of the 20th century. The 'machine translation systems' came of age, and developed in ways that could be put to practical use around 1990. The large-scale use of computers (and the complementary electronic tools including translation memories) by all the 3,000 translators and some of the nearly 1,000 interpreters at the European Union institutions meant that enormous corpora of translations became available to language professionals at the EU institutions. Finally, machine translation took to the Internet in 1997 and 1998 (first in Babelfish and then Systran itself) (Dollerup 2004:11).

Moreover, a new market of translation services, language localization and translation software is born. Many tools have been invented like the MT (machine translation) and CAT (computer-assisted translation) tools that are supposed to speed up the translation process. However, it is important to say that the work of specialized translators, especially in technical fields, is still needed.

1.6.1. Localization

Since the Internet was invented, people have had many new ways to share and retrieve information. This invention has had effects on many aspects of our daily life and in almost every field. The business sector is probably the one that has changed the most

after the arrival of the Web and it keeps changing together with the evolution of the technologies.

In a society where private firms are the main source of wealth creation it should come as no surprise that, once the technology of the Internet became available in the 1990s, the fastest, most comprehensive diffusion of its uses took place in the realm of business. The Internet is transforming business practice in its relation to suppliers and customers, in its management, in its production process, in its cooperation with other firms, in its financing, and in the valuation of stocks in financial markets. The proper uses of the Internet have become a key source of productivity and competitiveness for all kinds of business (Castell 2001:64).

The strategies chosen by the companies to communicate with their clients, suppliers and with the other companies have evolved simultaneously to the global spread of the Internet. At the same time, new affordable means of transportation and new technologies that facilitate the moving of people and goods and the spread of information have been invented and made available to the masses. All these factors and the international agreements have marked the beginning of the globalization Era.

If trains and ships facilitate the movement of manufactured products from one part of the world to another, the internet traverses and trades information and communication instantly across the globe via fibre optics, satellites and cloud computing. From telegraph to telephone, radio to television, computers and now the mobile Internet, international communication has been shaped by technological innovation. Combined with the extraordinary advances in computing capacity and concomitant reduction in costs, the convergence of computing and communication technologies has created unprecedented global interconnectedness in what Business Week hailed in 1999 as the new *Internet age* (Thussu 2006:225).

From the small family business to the big multinational company, from the lay people to the scientists, everybody has experienced transformations in the communication with the others. For instance, since old local markets have become saturated, entrepreneurs have searched new clients on a global scale and, as any good business man knows, speaking the language of the customer is a key factor in selling. Therefore, knowing how to adapt products and their promotion to a specific market share has become crucial nowadays. This process does not involve only the translation of the product's characteristics and advertisement in the language of the target country, but also the adjustment of all the features to the culture of the chosen market. This is the reason why today the localisation of the products, websites and all the other online tools is becoming more and more important. The *LISA (Localization Industry Standards Association)* has stated that *localization* is:

1. The process of modifying products or services to account for differences in distinct markets.
2. the process of adapting software for a particular geographical region (locale). Translation of the user interface, system messages, and documentation is a large part (but not all) of the localization process. Localization is often abbreviated as L10N. This abbreviation is formed using the first and last letters of the word (L, N) and the number 10, which specifies the number of letters between the L and the N.
3. The process of modifying products or services to account for differences in distinct markets.⁸

These days products which need localization are countless, but the main ones are related to Internet and computers like applications, programs, software packages and the websites where all the information and details about on- and offline products are hosted. Indeed, for what concerns all these goods a mere translation is not enough. If a company wants to sell a program or a software, not only the program and the user interface have to be translated and localized but also the whole package and support materials such as online guide, paper documentation, marketing material, the product's packaging and the related web pages. In computer-related localizations many disciplines which require both translation and Computer science skills are involved, among the others the automatic translation, terminology management and machine assisted translation. The process of adaptation of the products has today become a natural part of the products life.

Il ciclo di vita di un prodotto sviluppato in un'unica lingua e per un unico mercato può essere riassunto schematicamente in cinque fasi: definizione dei requisiti, progettazione, sviluppo del *software*, testing e infine il rilascio del prodotto. Se invece si decide di localizzare, il processo diventa più complesso, soprattutto se le versioni in lingue diverse verranno rilasciate contemporaneamente al prodotto originale (Musacchio and Sostero 2007:60).⁹

Considering that localization in the field of software tools and Computer Science is a growing sector, the professional perspectives in this area are huge. However, in addition to computer and translation skills, a good understanding of the target culture is essential to adapt the product to the target consumers. Knowing the geopolitical and religious situation is crucial to not offend the market. For example, where there are territorial disputes or in territories where different religious cults coexist.

In the era of globalization, English is still the main language for the trade and the

⁸ <https://web.archive.org/web/20100315061742/http://www.lisa.org:80/Glossary.108.0.html?tid=1>

⁹ Normally the life cycle of a product design in one language and for one market has five phases: definition of the requirements, planning, development of the software, testing and, in the end, the release of the product. But if one decides to localize the process become more complicated, especially the versions in the different languages will be release at the same time as the original product.

exchange of scientific findings. Therefore, most of the software maintain their name in English even in the foreign countries. Yet translating all the other parts is indispensable to make a program accessible to the non-expertise users.

Diversi anni fa, l'ex cancelliere tedesco Willy Brandt, premio Nobel per la pace 1971, dettò quello che oggi è noto come il primo postulato del commercio globale: "*If I am selling to you, I speak your language. If I am buying, dann müssen Sie Deutsch sprechen*". E' infondo uno die principi della globalizzazione, intesa come l'azione positiva dell'integrazione degli scambi e dell'interdipendenza delle nazioni (Musacchio and Sostiero 2007:69).

It is important to underline that not all the localization processes require the same skills and that different types of texts necessitate different translating approaches; consequently, diverse software require different types of localization.

1.6.2. Types of localization

User interfaces, video games, marketing contents, mobile applications and websites are only some of the online materials that usually require localization. Each one of these different sources need different types of localization approaches and different skills. The translator who decide to operate in the field of the online localization has usually to have good computers skills besides the linguistic ones.

1.6.2.1. User interface

The user interface, also called "GUI", is what allows the human-computer interaction, especially for the less expert users who do not know programming languages and what is behind the creation of a program, a software, an operation system or an application.

The *user interface* to an interactive product such as software can be defined as the languages through which the user and the product communicate with one another. In the case of software applications, this usually means the way displays and feedback are designed (the application-to-user language) and the way users indicate to the application what they want to do next through interactions with display elements via input devices such as a mouse or keyboard (the user-to-application language). As far as users are concerned, the user interface is the product. Just about their entire experience with the product is their experience with its user interface (Mayhew 1999:1).

Therefore, an intuitive user interface is essential, and it is one of the factors that can determinate the success of a software. Consequently, without a correct localization the average user would not be able to use even the simplest program.

Some of the most popular programs in the world are surely the ones included in the Microsoft Office package.

Ovviamente, dello sviluppo di mercati globali hanno goduto innanzitutto le multinazionali, cui vai riconosciuto il merito di aver saputo favorire e sfruttare la diffusione dei mezzi di comunicazione. *Microsoft*, per esempio, trae oltre il 50% dei propri profitti dalle vendite in paesi di lingua e cultura diversi dalla propria ed è un perfetto esempio di azienda globale (Musacchio and Sostiero 2007:69).

Their programs are available in many countries and, for this reason, they have been localized. The menus and toolbars of the applications in the different countries are the most blatant example of how a program should be adjusted according to the final users.



Figure 1 Word Office 2010 – German version

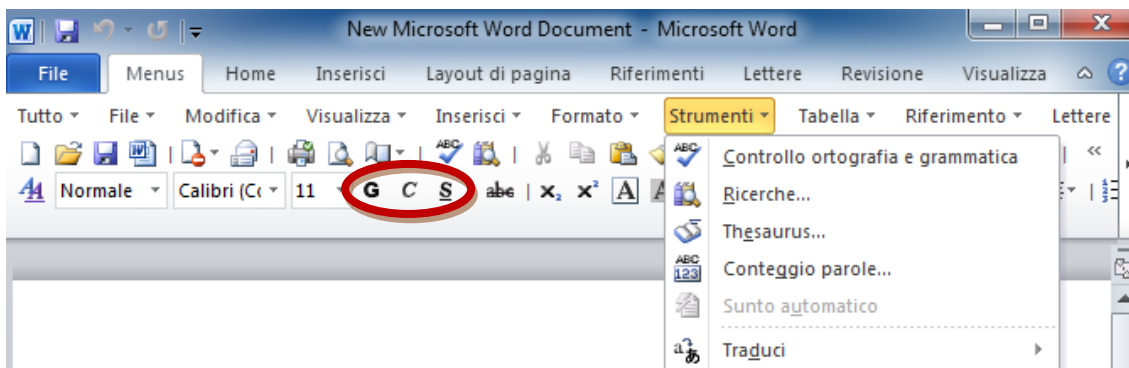


Figure 2 Word Office 2010 – Italian version

In the Figure 1 and 2, the Italian and German versions of *Word Office 2010* are shown respectively. After a general observation, it is possible to notice the signs of the localization. In the English version, for instance, the symbols for bold, italics and underline are “B”, “I” and “U”, in other words they are the initials of the words that indicate what will happen when the user clicks on them. In the German version they have been adapted to “F”, “K” and “U” which are the initials of *Fettudruck*, *Kursiv* and

Unterstrichen whereas in the Italian version the user will find “G”, “C” and “S” which correspond to *Grassetto*, *Corsivo*, *Sottolineato*. These adaptations were made because it is more intuitive for the user, even for a beginner, to understand how the program works if the buttons refer to his or her own language. Moreover, in the English version there is a drop-down menu called “Tools” which has been translated with the word “Extras” in the German version, that is not a literal translation whereas in Italian has been kept the exact translation “Strumenti”. This decision probably has been made because of the complexity of the German language. In German, there is no short word for translating the concept of “tools” in this type of context, therefore during the localization a shorter but still intuitive word (“extras”) was considered more appropriate.

Translating a user interface is not like translating a plain text because the text is inserted in a program and it should be adapted to the code and the graphic of the program. For this reason, the localization of user interfaces requires also good Computer Science skills and at least a general understanding of the programming language.

1.6.2.2. Video games

A video game is a complex environment because it involves not only the written text but also the images, the music and the voice of the characters. The video games’ market is huge and it could offer many possibilities to people who works with localization.

Si tratta di un processo complesso per il quale sono necessarie amplissime competenze, coinvolgendo di fatto tutti i media con intervento artistico e culturale sui contenuti. Si va, infatti, dalla localizzazione classica delle risorse, grafica a animazioni alla traduzione dei “copioni” utilizzati per i dialoghi dei personaggi e per l’interazione dei giocatori, che presenta per questo alcune delle caratteristiche tipiche della traduzione per il doppiaggio, a cominciare dal *lip sync*, per finire al trattamento dei componenti multimediali che, a sua volta, va alla manipolazione delle tracce al doppiaggio e al *voice over* (Musacchio and Sostero 2007:80).¹⁰

Consequently, this type of localization entails a good ability to handle multimedia files. Moreover, many video games involve a plot and characters with strong personalities

¹⁰ It is a complex process that require wide knowledges because it involves all the media with an artistic and cultural intervention on the contents. It goes from the classic localization of the resources, graphics and animation to the translation of the “screenplays” used for the characters’ dialogues and for the interactions of the players, which includes some characteristics typical of the translation for the dubbing, like the *lip sync*, the management of the multimedia components (ex. Tracks manipulation) and the *voice over*.

and both in the story or in the dialogues there could be references to cultural, historic, religious or political events or symbols. During the localization it is important to take into consideration all these factors because they could not be appreciated, appropriate or even understood from the final public.

Sometimes even an attempt to accurately portray a historical scenario can go awry. In the original *Age of Empires*®, a scenario was created in which the Yamato armies of Japan invaded the Korean peninsula and effectively overwhelmed the Chosen people of Korea. Historians tell us that this is generally what occurred and so the game designers diligently replicated reality. However, the government of South Korea sees history differently and disputed the scenario's accuracy. A downloadable patch was developed that slightly changed the scenario so that the Yamato invasion wasn't quite as overwhelming, i.e., the Chosen armies were given a fighting chance in the game's world (Maxwell Chandler and O'Malley Deming 2012:23).

When this type of issues occurs, the dialogues are adjusted too. Thus, paying attention to the cultural context and knowing what could offend or not being appreciated from the final public is crucial. In this type of localization not only translators but also interpreters could be involved in order to perform the *voice over*. In addition, graphic and Computer Science experts play a huge role in videogames localization because even the images could be linked to a culture or a philosophy or way of thinking and they must be adjusted too.

1.6.2.3. Websites

Today, internet and the world wide web are more and more used in the everyday life. Millions of persons search information online on daily basis and it is more important than ever to be reachable from as many users as possible. After the globalization of markets, most of the big companies have decide to operate on a multinational level. Considering that the web site is like a welcome card for the enterprises, it has become necessary for the companies to provide users with information about their business, the products and the after selling assistance in their own language because the consumer like to feel that the company care about him/her and, on the other side, the company wants to offer the best buying experience in order to convince the client to choose its products or services also in the future. The management of the corporate website contents is part of the marketing and communication strategies of a company.

Web sites are a critical component of the rapidly growing phenomenon of eCommerce. Worldwide, Internet retail sales have grown from \$18.23 billion in the fourth-quarter of 2000 to \$25.29 billion in the fourth-quarter of 2001. Web sites play a significant role in the overall

marketing communication mix—they complement direct selling activities, present supplemental material to consumers, project a corporate image, and provide basic company information to customers (Loiacono, Watson and Goodhue 2002:4).

This type of localization is different from the others because a website contains different texts in different forms and formats. In addition, it usually hosts other media like photos or videos. Moreover, one of the characteristics of web-based texts is that they are often linked with other texts through hyperlinks which must be updated during the localization process as well.

The technological aspects contribute to the complexity of web textuality: since language plays a fundamental role in the source code as well, this latter includes “linguistic hints” about the content of the web pages in the form of keywords, key phrases, and full sentences or short texts. These elements contribute to determine the website ranking in the results of search engines and consequently the visibility of the information presented. For this reason, each web page should be considered as a harmonious whole, since it is a sort of container from where the reader picks up information s/he needs, and therefore artificially separating what is considered to be the main body from the rest is an arbitrary operation and it would not make sense in many cases. In sum, web pages tend to be more complex and more mixed than traditional paper or electronic documents (Cappelli 2008:5).

Furthermore, web pages keep changing because the information should always be updated, consequently the localization of a website could be a long-term project. While translating a website, attention to details is essential. There are indeed many stylistic and cultural particulars hidden in the texts. Date and time formats, as well as calendar settings, change, for example, between USA and Europe. While most of the European texts tend to use the *dd.mm.yyyy* format, the American ones tend to use the *m/d/yyyy* or *yyyy-mm-dd* formats. In that websites where products are sold (i.e. e-commerce platforms) other details are important like currency formats, number formats (thousand separator), address formats and units of measure.

It is important not to forget about the graphics details as well. Colours play an important role in marketing campaigns but in different countries they could have different meaning and make people feel in different ways (e.g. the colour *red* means danger in European cultures, but good fortune in Chinese cultures).

A glaring and well-known example of website localization which includes all these details is the website of the worldwide colossus *Amazon*. The American e-commerce platform has different homepages for the different countries in which it operates because it adjusts not only its deals and advertisement to the location but its graphic and all the above-mentioned details as well.



Figure 3 Amazon Home page in China

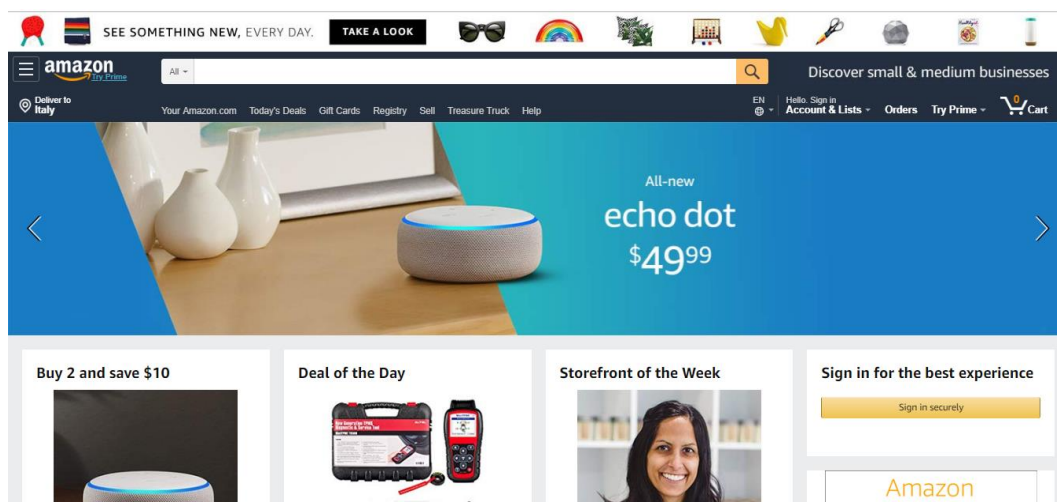


Figure 4 Amazon Home page in the USA

In the figure 3 and 4 the Chinese and North American homepages of Amazon are displayed. They are not different only because of the translation but also for many other smaller or bigger adaptation details.

In the Chinese homepage the background colour is white, there is a side menu on the left, the currency is the Yen, the cart is located on the left of the “wish list” and the logo is under its Chinese translation. The first products that the user see are clothes and items for the everyday life. In the American version the background colour is dark blue, there is no side menu on the left, the currency is the dollar, the *wish list* is on the left and the cart on the right and the logo is slightly different. In this version of the website the promoted products are electronics devices designed for the house.

All these particulars are the results of the localization, but they derive from market and users' studies as well because the company adjusts itself not only to the culture but also to the clients buying habits in that area.

1.7. Scientific websites localization

There are many other types of websites besides the business and commercial ones. Each kind of them require a specific approach and special knowledges. Two other main categories of websites which need to be localized are the ones related to science and scientific and technical fields. In order to translate these websites categories, the translator need to have acquired notions concerning the specific discipline besides the Computer Science and languages skills.

The contents of these webpages need to be localized not in order to sell good but in order to share information and to allow dialogues and discusses between experts (or between experts and lay people) all around the world. Indeed, nowadays scientific, medical and technical notions can be found on Internet. Some pages are designed to put into contact specialist with other experts around the world in order to promote their collaboration and encourage developments and breakthrough in the field. On these websites technical words are used because it is assumed that the reader has already a good knowledge of the technical concepts. Other pages, instead, are designed to put into contact specialists and lay people with the purpose to promote, stimulate and disseminate access to their studies or with an educational purpose. In this case the experts try to use a vocabulary that lay people can understand and difficult concept are explained in an easier way because the readers are probably people who wants to get a general idea of the concepts and they have not enough background knowledges to understand technical concepts explained in a technical way.

Even if English is since years the main language for the technical and scientific papers, there are still professionals who prefer to publish their works and studies in their own languages and if the other specialists do not know that language, it could be a problem. In addition, the websites of hospitals or other institutions could need to be localized in order to allow foreign patients or researchers to access to all the information.

At present, one of the most reliable terminological resource providing patient-oriented information is SNOMED CT, that is an international standard for the interoperability in Digital Health; it can be used to represent clinically relevant information consistently and to support the development of comprehensive high-quality clinical content in health records (Vezzani and Di Nunzio 2019:4).

A patient-oriented website is the one of the “Fondazione Poliambulanza”, one of the most well-known hospital in Lombardy.

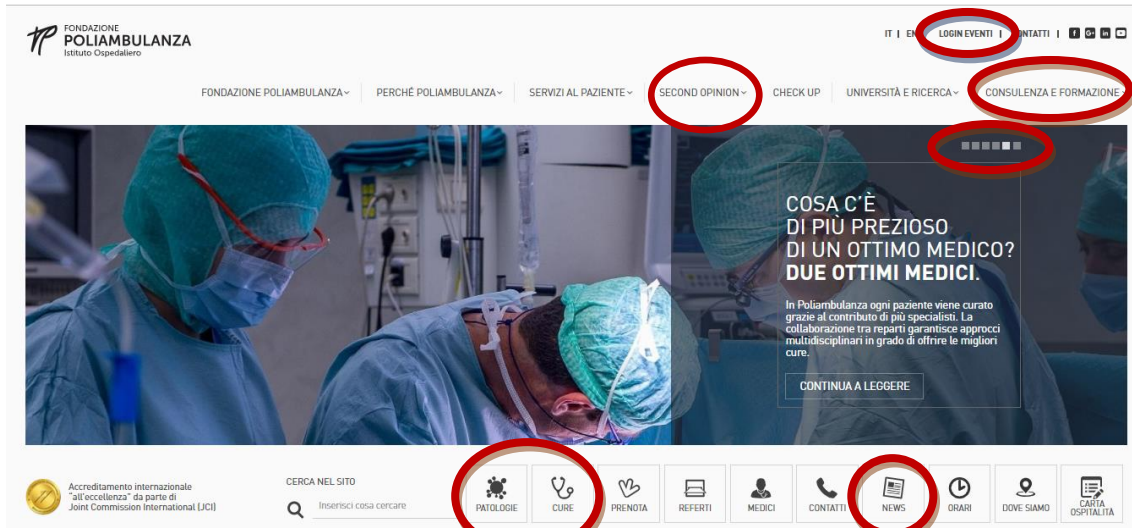


Figure 5 Fondazione Poliambulanza - Italian homepage

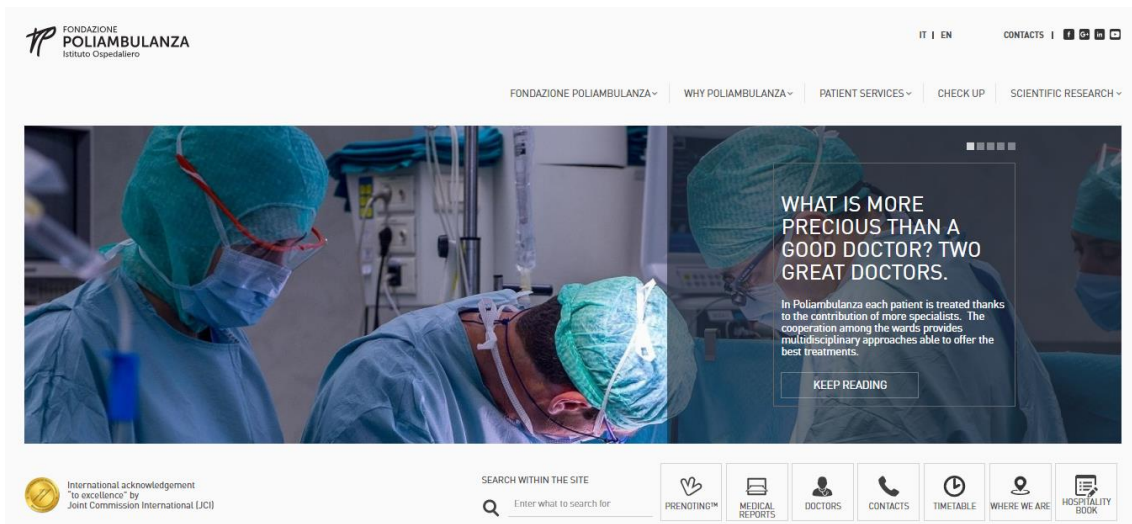


Figure 6 Fondazione Poliambulanza – English homepage

In the English version some pages of the website have been translated only the useful pages of the website. The foreign user can find all the information about how to book a medical examination, the services offered by the hospital and the Scientific Research but not all the news about local events and nurse school owned by the clinic.

There are websites designed for both experts and everyone who wants to be well-

informed about health care, for instance, *Cochrane Italia*, *Health Search* and *AIFA* (*Agenzia Italiana del Farmaco*).

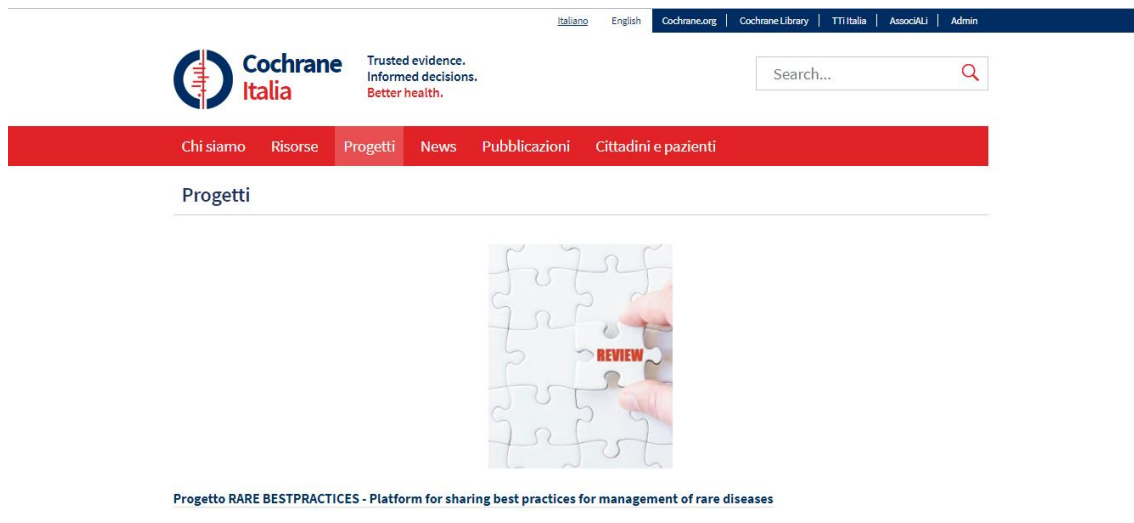


Figure 7 Cochran - Italian version

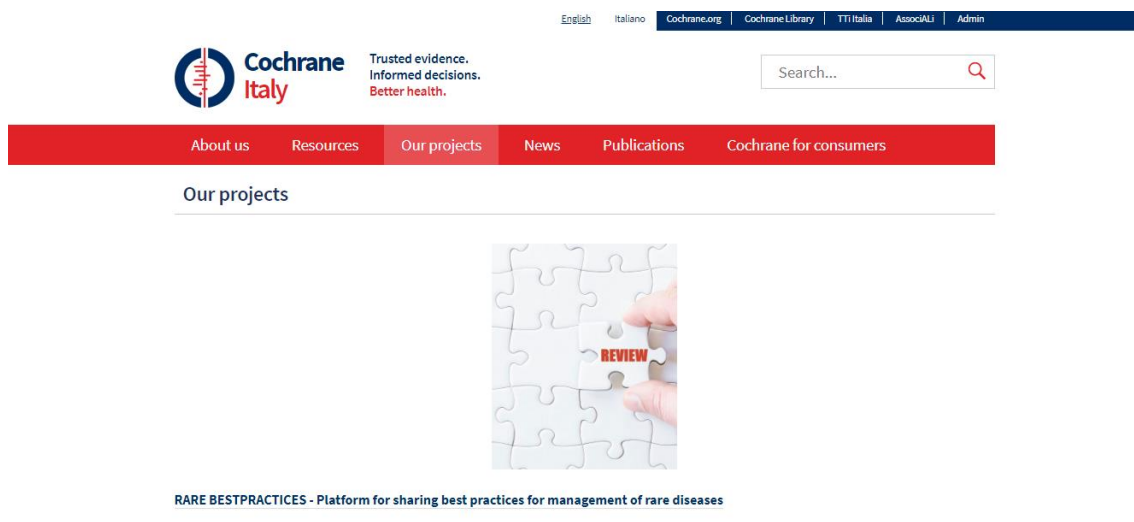


Figure 8 Cochran - English version

Cochrane is a global network that aims to promote evidence-informed health decision making and to share medical researches throughout different countries. In each of them a platform with the main information is available in the local languages but the Cochrane Library site is offered only in the English and Spanish version. Moreover, most of the collaboration are conducted in English.

Health Search - Istituto di Ricerca della Società italiana di Medicina Generale (SIGM)

website has been created to help the researchers to collaborate with each other and to share papers. In this case only the newsletter is available just in Italian, all the other pages have an English version as well. The terms used are more technical, therefore it is supposed that the user has previous medical knowledges.

The portal of *AIFA – Agenzia Italiana del Farmaco* is the official portal of the Italian Medicines Agency, the authority responsible for drugs regulation in Italy. In this case, technical terms both from the medical and the legislative field are used.

Today translators have at their disposal many tools that help them with the localization of technical, scientific and medical web pages both the ones with educational aims and the experts-oriented ones.

CHAPTER 2

2. Computer-Assisted Translation

Today, a wide range of software for translators is available. Depending on the use of these tools, it is possible to identify three different types of translations methods: machine-aided human translation (MAHT), human-aided machine translation (HAMT) and fully automated machine translation (FAMT). In particular, for the aims of this dissertation, we are taking into account the MAHT, which is also called *Computer-Assisted Translation* (CAT) because the computer is like an assistant for the professional who should put a great effort in the work even if the software tools give a great help.

CAT tools have been incorporated into the translation process, generally speaking, for the following reasons: a) they can save a significant amount of time compared to translating without CAT tools; b) they facilitate managing projects that involve translating a source text into various languages; c) they allow previous translations to be re-used as a language resource for new translations; d) they simplify the handling of files in different formats and facilitate page layout or DTP and final publication of the translations; e) they systematise the translation process so that it can be standardised and protocols can be implemented; and they establish different professional roles within the translation process, and thus encourage specialisation among translation professionals (Gea Valor 2010:255)

In fact, these days being specialized in a sector is very important for translators because in the medical, scientific and technical fields a deep knowledge of the subject is essential for a high-quality translation. When it comes to localize websites, in particular the specialized ones, the machine-aided human translation is the most used method because one of the main problems during this type of translation is to find the most appropriate terminology and the CAT tools are mainly designed to help the translator overcome this issue. However, most of the existing software tools were originally intended for the translation of plain texts though today the professionals must handle more complex source texts. As a matter of fact, in the past most of the material which required translation were manuals, papers, articles and books. Nowadays a huge quantity of translation works is instead related to Computer Science or multimedia

(websites, videogames, software, application etc.). These sources are more complex. For instance, during the localization of a website, the localizer needs to cope with a complex environment which includes different types files and media in different formats. Therefore, specific tools and Computer skills are needed.

Since the late 1980s the most dynamic sector of the translation profession has been that linked to translating digital content—translating for the screen, not for the printer; translating for localisation, not for publishing. Localisation, in its classic late nineties’ definition, means the linguistic and cultural adaptation of a product or service into another language or locale. It has translation at its core, but equally involves associated engineering and managerial tasks. From the nineties onward, this shift went hand-in-hand with increasing demand, as the Information Technology industry realised that the task of translating user interfaces, user assistance, web pages, video games etc, far exceeded the capacities of its bilingual staff. This was the age of the Language Services Providers (LSP) that employed translation technology, the internet, and pools of professional translators and revisers to process large jobs efficiently and competently. Without professional intervention, the industry would have choked in the linguistic mess that the amateurs had been creating (Garcia 2009:2).

It is important for anyone who aspire to work in the field of translation knowing which software tools are available because Computer Science could be very useful for translators. Moreover, computing is nowadays essential for collaborative translation project because it allows to share and receive information and translation memories from other co-workers.

2.1. CAT Tools

A wide range of CAT tools are available for translators and localizers but some of them are more used by professionals than others because they offer particular useful functions. As previously said, the CAT tools were originally designed for the translation of plain texts but now they can come in handy also for the localization of more complex sources of translation such as websites, software and applications. In this chapter, we will show and analyse the advantages and disadvantages of some of the most used CAT tools.

*OmegaT*¹¹

OmegaT is “a free translation memory application written in Java. It is a tool intended for professional translators”¹². This program can work with more than 30 file formats

¹¹ <http://omegat.org/>

among which the XHTML and HTML formats. These types of file are those used to write a Web page and are therefore, they are useful when it comes to translate a web site. Nowadays, the web pages are becoming more and more complex and with a lot more than simple plain text and even the written part is its own formatting. The disadvantage of *OmegaT* is that the source text must be extrapolated in one of the available formats and after the translation has to be reintegrate with the other parts of the website.

In this software, the text is divided into segments on the left of the working space, the segments that have already been translated are highlighted in yellow and on the right suggested translations are displayed with a percentage of concordance with inserted the translation memories and the previous segments.

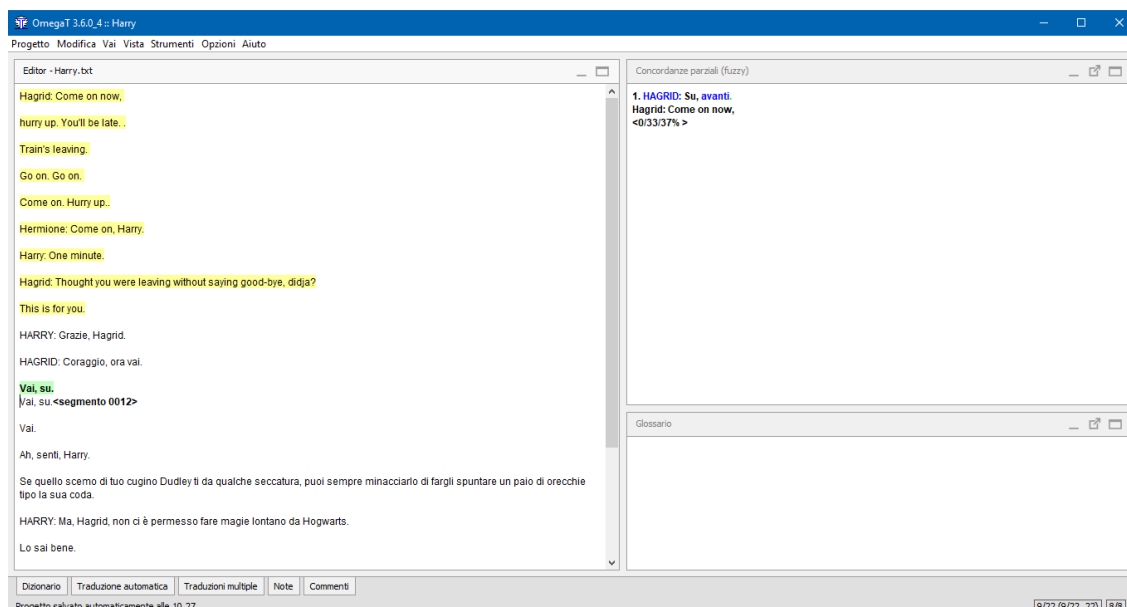


Figure 9 Example of OmegaT project

*SDL Trados Studio*¹³

According to its website “SDL Trados Studio 2019 is a complete software environment utilizing a translation memory. It allows language professionals to edit or review translations, manage translation projects, organize terminology and connect to machine translation”¹⁴. A disadvantage of this software is that it is a paid software, thus the

¹² <http://omegat.org/>

¹³ <https://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/>

¹⁴ <https://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/faqs.html>

translator must buy it. However, an advantage is that *SDL Trados Studio* supports more than 70 file formats. Nevertheless, neither this application was originally designed for website translations and it has the same disadvantage of OmegaT.

In this software, the segments of the source text are visible on the left and the corresponding translation on the right, the suggestions are available on the top of the screen and there is a rich toolbar.

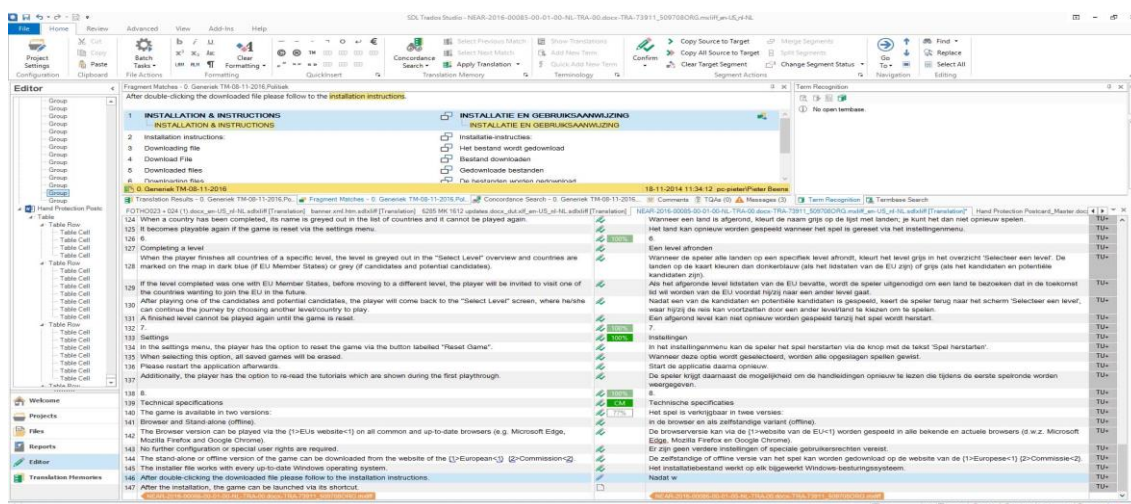


Figure 10 Example of *SDL Trados Studio* project

Besides these CAT Tools, other more useful programs for website localization are nowadays available. For instance, *site translator*, *SDL Passolo* and *Poontoon*.

*Site translator*¹⁵

Site translator is a paid software designed specifically for website translations. “Site Translator is designed for web site owners and web developers who need a simple, do-it-yourself solution to web site language translation. With a few mouse clicks, you can translate virtually any site designed in HTML, ASP.NET, PHP, PHP3, or PHTML into multiple languages”. Therefore, the text can be uploaded both as text in one of the available formats or by means of a URL. Moreover, additional translation memories can be added, and unwanted words and tags can be left out. In additions it has an intuitive graphics.

¹⁵ <https://www.deskshare.com/translation-software.aspx>

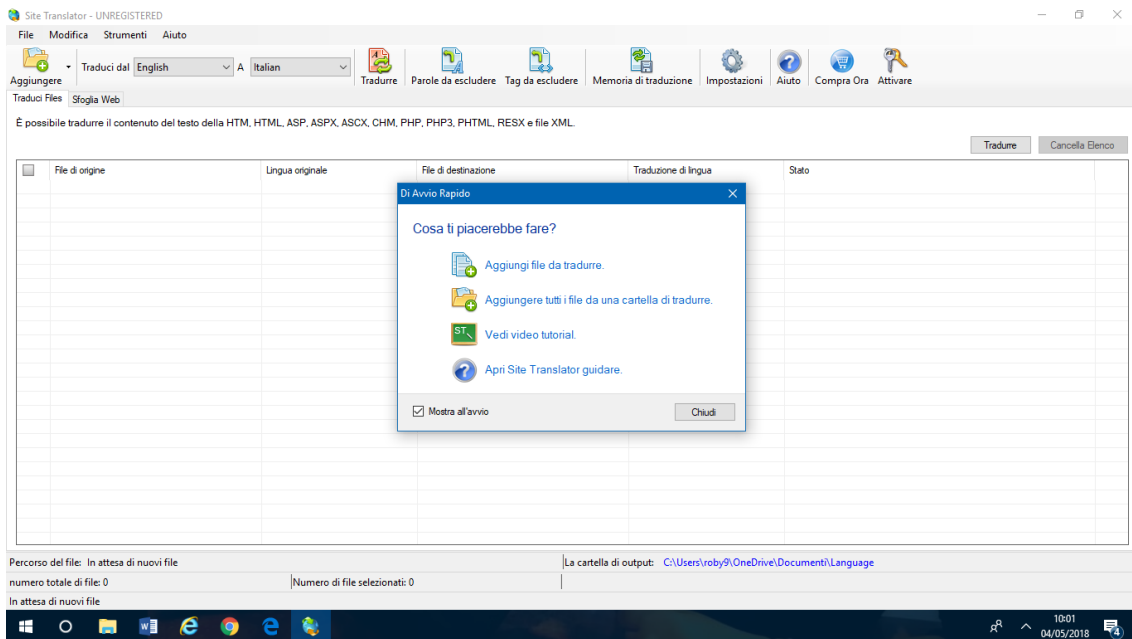


Figure 11 Example of Site Translator Studio project – Part1

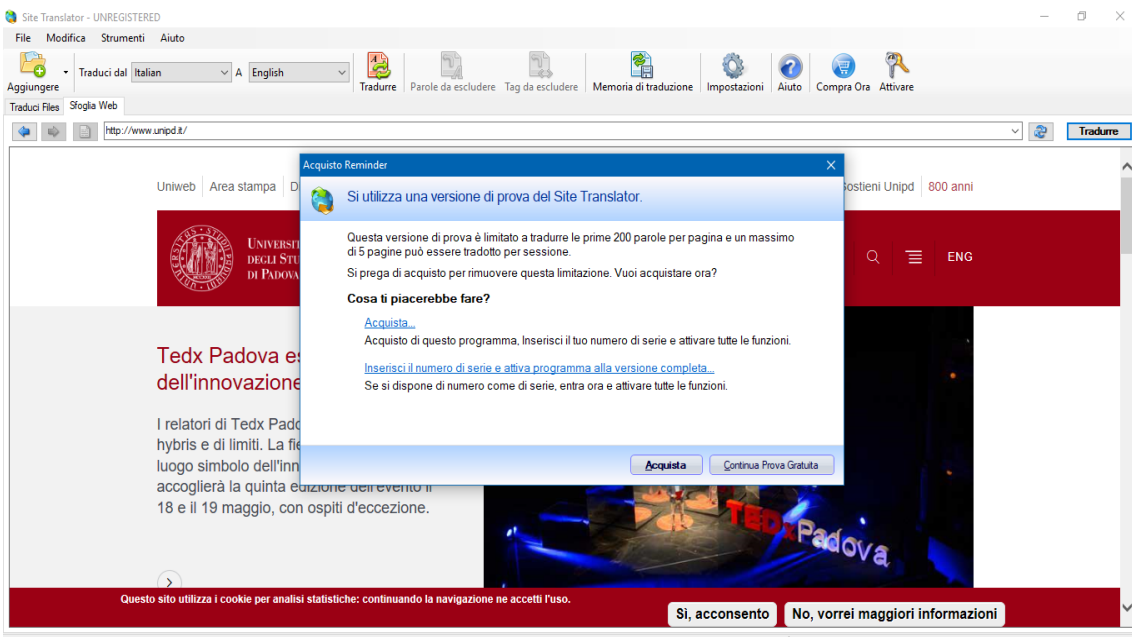


Figure 12 Example of Site Translator Studio project – Part2

SDL Passolo¹⁶

This paid software created by the same company of *SDL Trados Studio* is designed specifically for software localization. Its interface helps even the translators who has no

¹⁶ <https://www.sdl.com/software-and-services/translation-software/software-localization/sdl-passolo/>

programming experience because it allows user to work in WYSIWYG (What You See Is What You Get) mode. Moreover, using *Passolo* it is possible to export a list of string in the SDLXLIFF (XML Localization Interchange File Format) format and, once it has been translated, to set it up again. In addition, almost every year a new updated version is released to create the ideal software localization environment and maybe in the future a too specifically designed for websites localization.

*Poontoon*¹⁷

Poontoon is a huge localization project promote by Mozilla and which has the aim making Firefox available in as many languages as possible thanks to the work of a global community of localizers from all over the world. One of the most remarkable features of Pontoon is localizing web content in-place, on the web page itself. The moment the user moves the mouse over a text, the translation replaces the original text in the web page.

Despite a wide range of tools for the MAHT (machine-aided human translation) is available, “over 200 million users every day”¹⁸ still rely on the fully automated machine translation, albeit it is hardly possible to obtain a high-quality translation of truly natural language in this way. However, it could be an option for lay people or to have a general idea about the content of the page. This type of translation is offered by free tools (e.g. *Google Translate*, *Bing Translate*) or paid tools (e.g. *Idiomax*) and Browsers’ plug-ins (e.g. *Instant Translate* for Google Chrome and *Translate* for Safari). Nevertheless, these are not professional translation and often they are not accurate, especially when it comes to translate technical sites. If the owner of a company decides to localize its website to reach a new market share, he should not use these types of tools but rely on a professional translation because the fully automated translations are often not grammatically accurate and do not take into account the culture of the target audience.

¹⁷ <https://pontoon.mozilla.org>

¹⁸ <https://www.languageinsight.com/blog/2017/google-translate-vs-professional-translation/>

2.2. Terminology

One of the most important functions in a CAT tool is the one that help the translator in the management of the terminology. Dictionaries, glossaries, lexicons have always played an important role in the translation processes because finding the most accurate and appropriate terms for that context make the difference between a good translation and a high-quality translation.

Terminologies are becoming more important to modern day society as technology and science continue to grow at an accelerating rate in a globalized environment. Agreeing upon which terms should be used to represent which concepts and how those terms should be translated into different languages is important if we wish to be able to communicate with as little confusion and misunderstandings as possible. Since the 1990s, an increasing amount of terminology research has been devoted to facilitating and augmenting terminology-related tasks by using computers and computational methods. One focus for this research is Automatic Term Extraction (ATE) (Foo 2012: 6).

In particular, terminology plays an important role in the technical, scientific and medical translations. Most of the CAT tools allow the translators to insert pre-existing Translation Memories (TM) before starting the translation.

When you study Terminology, a differentiation is always made in its definition because the word refers to three concepts: terminology as the study of terms, terminology as the practical aspect of doing terminographical work, and terminology as a set of specialized terms (Brenes 2017 :14).

The field of terminology is so wide that today some translators have become full-time terminologists, that is to say that they have chosen terminology as a professional career. Even the European Union has created the *Termcoord* Project, which is a terminology coordination unit that has the aim “to assist translators with their day-to-day tasks and facilitate terminology research and terminology management in the translation units, as well to increase the EP’s contribution to the EU terminology database IATE”¹⁹.

Terminologists specialize in identifying the terms that are used by subject field experts when writing documents and for oral communication within individual disciplines. They create and disseminate terminological resources by recording terms or using specialized software to ‘extract’ them from texts. They document terms and related concepts by crafting definitions, locating meaningful contexts, and providing guidance for usage. Frequently, terminologists play a role in naming new concepts or new products. They are also experts in using a variety of terminology management software solutions (DeCamp, Karsch, Kristy, Muegge & Wright, 2016: 3).

¹⁹ <http://termcoord.eu/what-we-do/>

Therefore, the work of a terminologist is fundamental to have a high-quality translation especially when it comes to technical documentation in which specific terms have a specific meaning and it is important to give the exact same meaning, otherwise the field's experts will not find the translation appropriate and useful or it will cause misunderstandings.

The history of terminology is ancient. Indeed, it is possible to claim that it was born in year 2600 AC when thematic monolingual dictionaries were made by the Sumerians on clay bricks. It continues with the Egyptian and Greeks who compiled dictionaries as well. From the first until the nineteenth Century, the focus was more on dictionaries and glossaries based on medical terms. Even if some studies started during the Renaissance, the term *terminologie* was coined by the German professor Christian Gottfried Schütz in 1770. Nonetheless, the modern history of terminology as an independent discipline began with the works of Eugene Wüster, who belonged to the *Vienna School*, and of Dmitrij Semënovič Lotte, who belonged to the *Russian school*. In fact, over the years, terminology has evolved, and different schools and different approaches were born.

The classical schools of Terminology are the Vienna (Austrian) School, the Soviet School, and the Czech (Prague) school, all of which emerged from the work done by these experts. According to Maria Teresa Cabré, three approaches were taken by these schools:

- A first approach that considers terminology to be an interdisciplinary but autonomous subject at the service of scientific and technical disciplines.
- A second approach focusing on philosophy, which is primarily interested in the logical classification of concept systems and the organization of knowledge.
- A third approach focusing on linguistics, which considers terminology a subcomponent of a language's lexicon and special languages as subsystems of general language (Brenes 2017 :28).

The advancement of Computer Science has brought a change in the structure and creation of databases and currently there are many tools for creating glossaries and terminological termbases.

2.3. How to create a Termbase

A high-quality terminology database or glossary is essential for a good translation, especially if it is a sector-based one. Today, the terminologists have at their disposal many software tools that facilitate the creation of database and glossary. In addition,

nowadays it is very simple to share glossary all over the world and to cooperate in the creation of new ones.

In any case, it is fundamental to know how to create a new glossary from scratch. Hellmut Riediger, translation and terminology professor, has defined three main phases for the creation of a high-quality terminology database.

Before describing how a glossary is structured, Mr. Riediger identifies three phases a terminologist has to go through while writing it. During the preliminary phase the work must be organised, the domain must be defined and relevant documents related to the topic have to be analysed. During the following phase, the main one, the terms are selected and validated in each language, and the terminological cards are filled out. Lastly, during the conclusive phase, the glossary is finished and ready to be published.

Starting from the domain (medicine, economy, chemistry...) and the subdomain ("psychiatry" could be, for instance, a subdomain of "medical and biological sciences"), Mr. Riediger lists all the fields a terminological card could contain: term, variants, definition (and related source), context (and related source), synonyms, antonyms, notes, equivalents, reliability.²⁰

Therefore, creating a good glossary is not as simple as it may seem, but it is a long and hard work, and it could take some time to be finished. Considering that the following translations will be depending on it, all the time spent to create a good terminology database is worth the effort. Above all, it is important to extract the terms from reliable sources. Nowadays many software tools are available for the terminologists as well. WordSmith Tools, TextStat, AntConc and WebCorp10 are some of them.

*WordSmith Tools*²¹

It is a package of programs (*WordList*, *Concord*, *Keywords*) for looking at how words behave in texts and they are used in corpora. In particular, *WordList* tool create a list, in alphabetical or frequency order, of all the words or word-clusters in a text. *Concord* is useful to understand the concordances of words because it allows to see any word or phrase in context. *KeyWords* comes in handy when the aim is to find the key words in a text.

The WordSmith Tools have been used by Oxford University Press for their own lexicographic work in preparing dictionaries, by language teachers and students, and by researchers investigating language patterns in lots of different languages in many countries world-wide²².

²⁰ <http://termcoord.eu/2015/11/what-is-terminology-and-how-to-write-a-glossary/>

²¹ <https://www.lexically.net/wordsmith/>

²² <https://lexically.net/downloads/version7/HTML/index.html>

*TextStat*²³

It is another program specifically designed to find and analyse word combinations and frequency from text corpora. The advantage of this program is that it supports any amount of text, therefore it is possible to create a huge corpus. Moreover, *TextSTAT* is available in three languages (German, English and French).

*AntConc*²⁴

This is an additional tool that could prove helpful during the construction of a terminology database.

AntConc is a freeware, multi-platform, multi-purpose corpus analysis toolkit, designed specifically for use in the classroom. It hosts a comprehensive set of tools including a powerful concordancer, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot (Anthony 2004:7).

*WebCorp10*²⁵

This program is different from the others for different reasons. First of all, it is completely online, it does not imply the download of a software. In addition, it does not require to build a corpus of texts because it uses the Web as a corpus. Another peculiarity consists in the fact that it is possible to decide if the research should start from a word or from a text. If the research starts from a word, the program will analyse the corpora and will create a list of contexts where the word has been used. If the research starts from a specific website or text, the program will analyse the text and will return a list of words in order of frequency.

Once a list of terms has been created and their accuracy has been verified, the terminologist should proceed to the creation of the real database which could then be insert in a CAT tool and used during a translation or localization process. A *termbase* is even better than a simply standard glossary because it contains not only the terminology but also all the related information which are called *metadata*. They are ordered and recorded as data categories, also called *terminological fields* which describe the term including, among others, the entry term, subject field, part of speech, preferred term, admitted term, examples, definition, usage comments. The list “ISO 12620” classifies

²³ <http://neon.niederlandistik.fu-berlin.de/static/textstat/TextSTAT-Doku-EN.html>

²⁴ <http://www.laurenceanthony.net/software/antconc/>

²⁵ <http://www.webcorp.org.uk/live/>

almost 200 possible data categories for a single terminological entry. However, ISO 12616 lists only three of these data categories (term, source and date) as mandatory²⁶.

It is important to underline that most *termbases* are multilingual and contain metadata for each term entry in a range of different languages. These data are really useful during a translation and computer Science helps terminologists creating termbases as well. Like for the CAT tools, today there is a wide range of tools at disposal of terminologists. The most famous and used software used in the creation of termbase is *SDL MultiTerm* but there are also other possibilities like *TermStar NXT* or *LF aligner*. *SDL MultiTerm* is a plug-in of *SDL Trados* and one has to buy it separately. However, it is the most used by the professionals and therefore it is important to know how it works.



Figure 13 Example of Termbase Entry – Multiterm 2009

As it is possible to see from the image, a *termbase* is composed by many entries, each of them could be in two or more languages and each translation could be followed by other information, such as the definition in that language and more formal or informal version of the term.

Once the *termbase* is ready and well organized, it could be used to create a Translation Memory. In fact, CAT tools allow the use of TMs which are databases containing previous translations. They automatically learn from every word that has been translated during the localization of a product. Translation Memories learn in real time, segment by segment, which means that the translation of a segment will be

²⁶ <http://inmyownterms.com/a-termbase-what-you-should-know/>

considered for the translation of the following segment. When translators work on similar texts the TMs suggest translations based on similarity and they inform the localizer if the match is fuzzy or exact matches. *Fuzzy matches* are suggestions from previously translation where texts match approximately. They could be a starting point for the translator and they help to speed up the work. *Exact matches*, instead, indicates that the text of a past translation is completely the same of the new translation. Working with a translation memory has many benefits among which the reduction of translation costs, the reduction of translation time and an increasing text quality. Having a good termbase as a start for the creation of a translation memory eliminate ambiguity, accelerate translation speed and guarantees correct usage and spelling of sector terminology.

The Translation Memories and the *Termbases* can be shared with translators around the world and can be open by different CAT Tools but, as previously explained, not all the tools support the same formats. The most common standards are *.tmx* (Translation Memory eXchange), *.tbx* (TermBase eXchange) and *.xliff* (Localisation Interchange File Format).

CHAPTER 3

3. Medical translation

History of medical translation is ancient.

Unlike most technical fields of translation, medicine has had a very long history of writing; it almost rivals the written tradition of law. [...] The distinctly scientific method that characterizes modern medicine in Europe and the Americas is traceable to a Greek civilization (500-30 B.C.) that succeeded in passing on its tradition first to the Roman Empire (100 B.C.-400 A.D.) and then to Medieval Europe (1200-1500 A.D.). In the process it created the core of the contemporary Western medical writing system. Why would one language area adopt the language of another [...]? It is axiomatic that dominance in knowledge, customs or technology has major repercussions upon language relationships. What is seen as superior tends to flow into what is seen as inferior; one may view the process in terms of either push (imposition) or pull (borrowing). Whoever leads the field gets to create the words that capture the emerging concepts and products (Fischbach 1998:13).

It is possible to claim that the medical history in the Western Europe started with Hippocrates' (c. 460 – c. 370 BC) works.

Hippocrates is considered to be the father of modern medicine because in his books, which are more than 70 (Grammaticos, Diamantis 2008:1).

Hippocratic Corpus was translated into Aramaic, Latin and Arabic. Not much later, between 50 and 70 CE, Dioscorides, a Greek pharmacologist and botanist, wrote *De Materia Medica*, an important volume about plants and medications. This work was translated into Latin Syriac and Arabic during the first century after Christ. The works of Galen, a prolific author with particular interest in pharmacy, are dated back to the second century AD. Many of Galen's works were translated into Syriac and Arabic, Latin too.

During the first centuries AD, Greek texts were translated above all in the Middle East. The works of the doctor Sergius of Reshaina (d. 536), who translated many Galen's works into Syriac in the 6th century, are only an example.

Sergius' Syriac translation of Galen's *On Simple Drugs* — along with the so-called Syriac Epidemics — is undoubtedly the most significant medical text in Syriac in terms of bulk and influence that remains unedited (Bhayro, Hawley, Kessel and Pormann 2013: 146).

Another important translator from that period is Hunayn ibn Ishaq (d.873).

Once he had felt was an authentic version, Hunanyn's translations were precise but not overly literal. The quality of these translation was such that Hunanyn was paid for them by their weight in gold. Within fifty years Hunanyn and his students completed the monumental task of rendering in Arabic and Syriac all the most important Greek medical texts written over a millennium – including all of Hippocrates, Aristotle, Galen, Discorides and their important commentators, from Oribasius to Paul of Aeginata (Tschanz 2003:39).

From that moment, the translation of Greek books into Arabic continues for the following centuries and in the 12th century the *Baghdad School* was founded with the aim of translating interesting non-Arabic works into Arabic. During the same period, the so-called *Toledo School of Translators* was founded. This group of translators was, instead, specialized in the translation from classical Arabic into Latin and Castilian. Their works contributed to the foundation of the modern Spanish language and the birth of new medical terms.

The labor done by the *Toledo School of Translators* is imprinted not only in anatomical language but also in everyday speech in many instances, and indeed in anatomical knowledge itself. To this, the translation of Avicenna's Canon made by Gerard of Cremona in Toledo in the last third of the 12th Century was especially important. Such translations impregnated the texts of Albertus Magnus, Mondino de Luzzi and Guy de Chauliac – which in turn were the major suppliers of anatomical Latin names in the Middle Ages and therefore those who made Gerard of Cremona the vector of Ancient Greek culture into anatomical science (Arráez-Aybar, Bueno-López and Raio 2015:13).

In 1439 Guttenberg's printing press revolutionized not only the publication field but the medical field as well because it facilitated the spreading of the medical knowledges. Moreover, after this important and revolutionary invention, translating started to become a real profession. However, several new medical branches appeared only between the 17th and the 19th centuries. Their birth produced a huge amount of new terminology because, even if many terms were borrowed from other sciences, several new ones were coined.

A new wave of significant breakthroughs in the field of medicine as a scientific activity was experienced in the 19th and in the 20th centuries. Simultaneously, English has become the *lingua franca* for most of the technical fields all over the world. In the era of the Internet and the World Wide Web, English has become the dominant language of international scientific communication as well, which implies multiple linguistic contacts between English and other vernacular languages.

The use of English as an international language of science (EILS) is by now well documented; depending on one's orientation, English may be seen as a neutral lingua franca or it may be seen more insidiously as a dominating and overpowering force (Tardy 2004:1).

Medical websites, online publications, videos and e-conference use English as the standard language. Therefore, today is essential for a translator, who wants to focus on the medical field, to understand English technical and popular medical terms.

3.1. Specialized languages

Medical terminology is part of the so called *special* or *specialized languages*.

Per lingua speciale si intende una varietà funzionale di una lingua naturale, dipendente da un settore di conoscenze o da una sfera di attività specialistici, utilizzata, nella sua interezza, da un gruppo di parlanti più ristretto della totalità dei parlanti la lingua in cui quella speciale è una varietà, per soddisfare i bisogni comunicativi (in primo luogo quelli referenziali) di quel settore specialistico; la lingua speciale è costituita a livello lessicale da una serie di corrispondenze aggiuntive rispetto a quelle generali e comuni della lingua e a quello morfosintattico da un insieme di selezioni, ricorrenti con regolarità, all'interno dell'inventario di forme disponibili nella lingua (Cortelazzo 1994: 7).

Therefore, a *specialized language* is a subset of a language used only by a specific portion of the population because it requires special knowledge and it is related to a field of specialization. The speakers of this language are able to understand complex concepts using one term because they have already learned the notion and they do not need further explanation. For instance, the medical term *pyrexia* indicates the rise in the body's core temperature as mechanism developed by the immune system to reduce the severity of illness by preventing bacteria and viruses from multiplying. This concept could be obvious for those who are expert in the medical fields and call to their mind all the related notions on how and in which circumstances this phenomenon take place. However, *pyrexia* is probably a murky concept for lay people and in order to express quite the same concept the popular term *fever* will be used. Yet, it is important to notice that the effect on the mental image is not the same. In fact, the word *fever* for lay people indicates only the rise of the temperature without implying technical explanations about the phenomena.

Inside the common language, there are many terms related to the medical fields and to other specialized sectors.

Caratteristica della medicina, rispetto ad altre scienze, è la forte proliferazione terminologica, che la allontana da scienze tipicamente “dure” come la matematica e la fisica, in cui è elevato il grado di formalizzazione non verbale (formule, grafici, ecc.), avvicinandola all'area umanistica. È facile rendersi conto di questo dato, verificando la quota di lemmi di ambito medico registrati da un dizionario dell'uso (Serianni 2005:114)

Therefore, many terms related to medicine are included in the everyday language, but these words are not to be considered part of the specialized language. In fact, the words that constitute a specialized language imply bigger and technical concepts that only those who are expert in that particular field can understand. Using these terms, the professional avoids to constantly repeat notions that are already known both from the speaker and the receiver and they facilitate the explanation of more complex and technical notions. For instance, the definition of *neoplasm* given by NIH²⁷ could be taken as an example: “Malignant neoplasms show a greater degree of anaplasia and have the properties of invasion and metastasis, compared to benign neoplasms”. In this explanation, three terms from the medical language are used (*neoplasm*, *anaplasia*, *metastasis*). While for an expert this could be a short but efficient explanation for the term *neoplasm*, for a lay person this definition requires the research of the explanations of the other two technical terms or of a longer, but easier and more understandable description written without specialized terms.

3.2. Medical web sites

There are two main types of medical websites: experts-oriented and lay people-oriented. The latter type usually tries to explain in an easy and simple way difficult terms and concepts to non-professionals, who for study purposes or personal interest decide to examine in depth medical issues in order to have a better comprehension in the field.

For what concern patient-oriented websites, the most reliable ones should be the websites of the Departments of Health. In this case, the government usually use popular terms in order to explain to the citizens the most common diseases and medical practise, besides explaining how the National Health System works. Taking as example the website of the *U.S. Department of Health & Human Services*²⁸, it is possible to notice that information about the most popular issues related to health (vaccines, screenings, nutrition and fitness etc.) are available. The American Department has even created a website appositely dedicated only to vaccines²⁹ offering specific notions about them and giving the citizens the possibility to have a deep understanding on this matter. In this

²⁷ <https://meshb.nlm.nih.gov/record/ui?ui=D009369>

²⁸ <https://www.hhs.gov/programs/prevention-and-wellness/index.html>

²⁹ <https://www.vaccines.gov/>

case, common terms are used in order to explain in the easiest way possible the health issues to the lay people. For instance, on this website the definition of *vaccines* is the following: *A vaccine is made from very small amounts of weak or dead germs that can cause diseases — for example, viruses, bacteria, or toxins. It prepares your body to fight the disease faster and more effectively, so you won't get sick*³⁰. This explanation is obviously intended for someone who does not have a deep medical knowledge but wants to have a better understanding of the issue.

However, the definition of the same concept on a experts-oriented website is usually different because another type of language, a specialized one, will be used. For instance, the *Medscape* website, a leading medical resource for physicians, defines vaccination as *the introduction of killed or attenuated microorganisms (bacteria, viruses, fungi, protozoa, or rickettsia), antigenic proteins derived from them, or synthetic constructs, administered for the prevention, amelioration, or treatment of infectious and other diseases. Vaccination can be administered orally, nasally, or via injection*³¹. In this definition there are various terms (i.e. antigenic proteins, protozoa and rickettsia) that could be not immediately clear to lay people. These terms require a further investigation by non-expertise, but they are immediately understandable for a physician.

Another peculiarity of the medical language are the abbreviations and acronyms. They are an important part of this specialized language, but they could be a problem during the translation because from one language to another they change according to which words they stand for.

Abbreviations are another problem. Those not commonly used by the medical profession as a whole should be avoided, however obvious they may be to the specialist. If for reasons of space it is desirable to use abbreviations, these should be fully identified the first time they occur. Although most would recognize BUN as the abbreviation of *blood urea nitrogen*, how many would know what SGCT is? The field is hematology. It stands for *serum glutamic-oxaloacetic transaminase*. In a list of hospital equipment items we were recently asked to translate for overseas distribution, the term IPP respirator appeared. None of the hospital administrators we consulted knew the answer, so we called the client and finally found out that the abbreviation stood for intermittent *positive pressure respirator* (Fischbach 1962:471).

Acronyms and abbreviations are widely used elements in both written and spoken medical communication. Many terms, such as names of disease entities (e.g. *autosomal dominant polycystic kidney disease*), names of chemical compounds (e.g. *6-pyruvoyltetrahydropterin synthase*), or therapies (e.g. *transurethral laser-induced*

³⁰ <https://www.vaccines.gov/basics/index.html>

³¹ <https://search.medscape.com/search/?q=vaccines>

prostatectomy) hardly ever appear in their full form because it would compromise the efficiency of the communication. The popularity of abbreviations in the medical language is due to the economy in space and time that is really needed in many medical emergency situations. In addition, it enables medical professionals to communicate in a fast, simply way complex concepts facilitating the exchange of ideas in order to promote new discoveries. Therefore, the translator should have a deep understanding the medical field in order to recognize the proper meaning and aims of the acronyms in a specific context.

In patient-oriented websites, the aim is not always to share and promote information and awareness with less expert people but also to advertise products, medicines, private clinics or collect money for new studies and researches.

Another difficulty the conscientious medical translator encounter is how to tone down the claims which are the mainstay of American advertising copy. We must realize that the doctor in Amsterdam or Annam is not necessarily accustomed to the same phraseology or form-and form plays a greater role in many countries overseas than we sometimes realize. Our plain and direct approach is often interpreted as abruptness. To write "*following thorough experimentation and after extensive clinical trials, we have come to believe that our antibiotic is highly effective in urinary tract infections and probably second to none,*" will sound more convincing in Oslo and Dakar than an assurance to the effect that "*for urinary tract infections, more clinicians prescribe our antibiotic than any other*" or "*our antibiotic is more effective than any other in clinical use today*" (Fischbach 1962:471).

As a result, the translation and localization of medical websites could be tricky not only for the technical terms but because sometimes it required the reformulation of entire sentences. In addition, nowadays medical terminology is growing thanks to the several new medical innovations. Therefore, it is important to be always updated about new terms.

CHAPTER 4

4. The Project

After analysing all the main existing tools available, it is safe to claim that nowadays Computer Science gives a huge help to terminologists, localizers and translators. However, it is important to make some observations.

First of all, most of the tools available for terminologists are free of charge, therefore it is possible to create good terminology databases without paying. However, if the final aim is to use the terminology inside a translation environment, the glossaries should be transformed into Translation Memories (TM) or in a format supported by the main CAT tools. Considering that the majors software used by professional translators are accessible only under the payment of a licence and that usually the files produced using them have proprietary extensions, there is the need of a free open-source software which allows the creation of terminological termbases in a format supported by all the principals CAT tools without a proprietary extension and for free. For instance, most of the professionals use *SDL Trados* for work and the files generated with this software are saved with the extensions *.sdlxliff*, *.sdltm*, *.sdltb* which are based on the open standards but they also contain the proprietary extension *sdl*, therefore, they can be usually opened only by those who have bought *SDL Trados*. This could be a problem because it does not permit to share information with those who usually use other CAT Tools and it makes complicate to use translation memories produced with other software tools. However, it is to say that unfortunately using open-source software usually means to have less functionalities at disposal.

In addition, there are a lot of tools for translators but most of them are not specifically designed for the translation of technical, scientific or medical websites. That is to say, none of them is exclusively intend for a specific field and consequently they do not consider the specific difficulties that could be encountered during the translation process of that particular type of text. In medical texts, for instance, acronyms are often used but from one language to another they change. Moreover, to the technical term usually corresponds a popular term. For a translator, it is important to know which of these two

terms is the more appropriate in the context of the translation in order to adapt the text accurately on the base of the final readers. If the readers are experts, the translator should choose the technical term, but if the translation concerns an educational text, the popular term might be more appropriate. Additional information could be useful to understand how to transport a concept not only from a language to another but also from a culture to another. For this reason, it is useful to have a terminological database that contains not only the translation of the term but also other information that could be useful for translators and localizers.

Another useful functionality would be the possibility to add new terms and all the related information to an already existing dataframe in an easy way and to have the chance to download the final and complete dataframe in a format compatible with all the main translator software.

A free open-source program that combined all these useful functionalities and designed for the needs of translators and localizers has so far not been available.

4.1. TriMED

The desire to fill these gaps in the field of the software translation and the idea of creating a useful tool for the medical field have brought the PhD Student Federica Vezzani in collaboration with the assistant professor Giorgio Maria Di Nunzio to create the *TriMED* project.

TriMED, is a digital library of terminological records designed to satisfy the information needs of different categories of users within the healthcare field: patients, language professionals and physicians. *TriMED* offers a wide range of information for the purpose of simplification of medical language in terms of understandability and readability inspired (Vezzani and Di Nunzio 2019:1).

This new resource is now available in three languages (English, Italian and French) and it can be visualized in three ways depending on the user.

There is a visualization designed for the patients which works as an Information Retrieval tool, that is to say that allows the users to find useful information about technical terms. For lay people, it is often problematic to understand medical terminology, therefore *TriMED* provides the user with the English definition of the

technical term and suggests which is the corresponding popular term. In this way, it should be easier for the patient to understand medical expressions.

The second visualization is intended for language professionals.

TriMED is actually designed for technical-scientific translators who have the difficult task to decode and then transcode medical information from a source language into a target language (Di Nunzio and Vezzani 2018:4).

In this case, *TriMED* works like an automatic translator providing the translation of the term, the corresponding popular term, the definition and an analysis of the term in both languages. These pieces of information help the translator to understand in which contexts the term is used and if it is more appropriate to use the technical or the popular term.

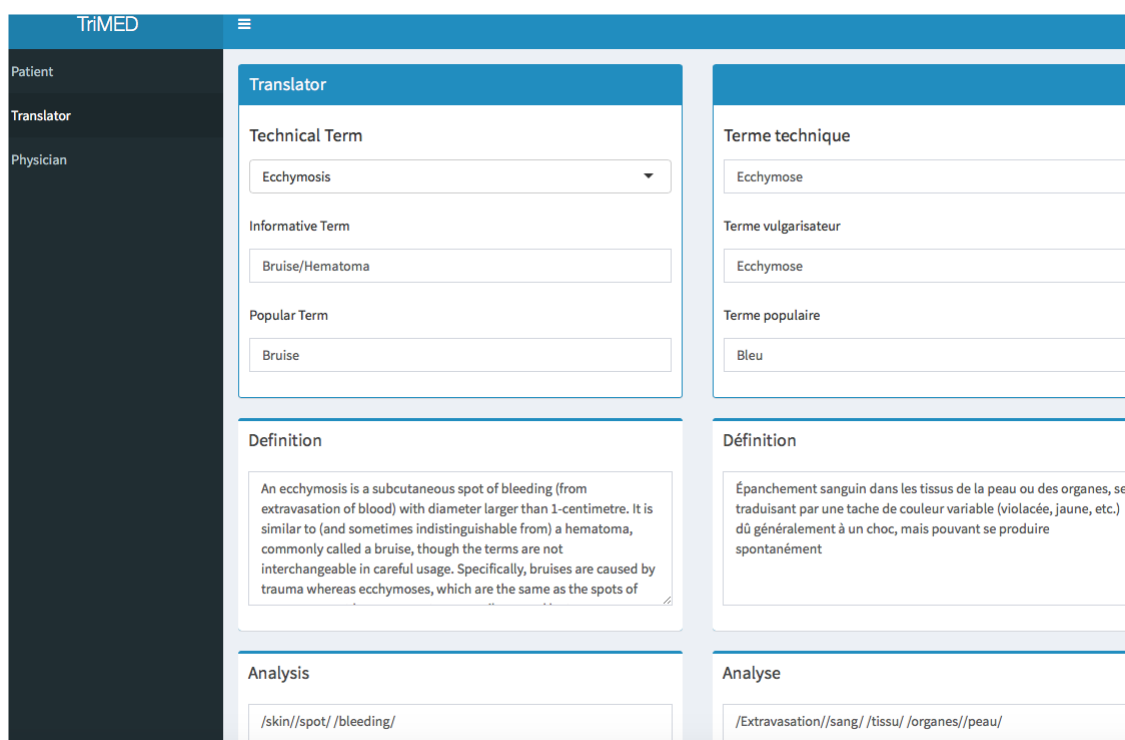


Figure 14 Translator Visualization – *TriMED*

The last visualization is designed for physicians. At an international level, most of the scientific texts are produced and released in English. Sharing the information around the world is essential to keep all the specialists updated about new health care protocols and scientific discoveries but the language could be a barrier in the communication among medical specialists speaking different languages. *TriMED* aims to help researchers by

offering a regularly updated and reliable translations of the medical terms together with their informative e popular equivalents.

The terminological database has been created extracting manually the terms from a corpus of reliable documents concerning the oncology field and, in particular, breast cancer treatments. The aim is to provide always updated information and the most precisely possible translations both of the technical and the popular terms and of the all related information.

Currently, the eHealth resource contains about 328 terminological records that are under review; a demo of the application can be consulted online at the website and it was realized with the Shiny R package showing the interaction among the three levels of communication (Vezzani and Di Nunzio 2019:6).

However, *TriMED* has not been designed only to analyse and simplify the medical terminology but also to facilitate the linguistic analysis of the literature domain. In fact, during the construction of terminological records some of the English technical terms have been extracted in a semi-automatic way from literary texts. This innovative approach shows that *TriMED* can support not only specialized documents but also for literary corpora.

4.2. The Design of the Plug-in

Taking inspiration from the *TriMED* project, I have developed my own plug-in with the aim to offer a new translation tool that combines the functionalities of the existing ones with some implementations. In addition, after analysing the existing CAT tools, I have identified of the most common disadvantages and I tried to overcome them.

In particular, after analysing the existing tools, we have decided to create a plug-in written in R using the Shiny package. *R* is a programming language and environment usually used for statistical and graphics but taking advantage of the *Shiny* package offered by *RStudio* it is possible to develop interactive web applications and plug-ins.

A plug-in is a (sometimes essential) piece of software code that enables an application or program to do something it couldn't by itself. One of the more common plug-ins is Adobe Flash Player. Without Flash Player you won't, for example, be able to view BBC News bulletins embedded into web pages. Other plug-ins are available for different things. There are plug-ins

for social media networking, foreign language alphabets and many other things.³²

This new plug-in is designed to manage medical termbases and to be used as a complement of a CAT tool. Its construction has started with the creation of two dataframes.

The concept of a data frame comes from the world of statistical software used in empirical research; it generally refers to "tabular" data: a data structure representing cases (rows), each of which consists of a number of observations or measurements (columns). Alternatively, each row may be treated as a single observation of multiple "variables". In any case, each row and each column has the same data type, but the row ("record") datatype may be heterogenous (a tuple of different types), while the column datatype must be homogenous.³³

The R language is mainly used in the statistical field but in this case, I have created two dataframes with different variables: medical technical term, the corresponding popular term, the definition, the pronunciation, which part of speech is, the etymology, the synonyms and in which reliable websites has been cited. The first dataframe has all the entries and all the related information in English and the other one the same data but in Italian. These pieces of information are essential during a localization process in order to understand how and in which context use the term. For instance, it is important to know in which contexts using the technical term is better than using the popular term and viceversa. Moreover, the definition helps the user to understand better the term both in the source language and in the target language. Another important piece of information is the *part of speech* because it helps the translator to create a grammatically correct phrase. The terms I have used for the plug-in have been provided from the work of the PhD Student Federica Vezzani, who is specialized in the construction of terminological record and has designed and implemented the *TriMED* project. For each of the terms, reliable web pages, in which they are cited in, are listed. In this way, translators can retrieve additional information and be sure of the authority of the source. Moreover, my plug-in allows the users to insert and save new terminological records. This option could be very useful to translator and terminologists, if they want to add new terms and information before downloading the *.tbx* file and using it inside a translation environment. I will explain later in this chapter in details how to add new terminological entries with my plug-in.

³² <http://www.bbc.co.uk/webwise/guides/about-plugins>

³³ <https://github.com/mobileink/data.frame/wiki/What-is-a-Data-Frame%3F>

```

11 # Definition of vectors
12 English_term <- c("Pathology", "Leucocyte", "Erythrocyte", "Pyrexia", "Neoplasm",
13 Popular_term <- c("illness", "white blood cell", "Red blood cell", "Temperature", "1
14 Definition <- c("Used as a count noun, a pathology, pl. pathologies, can also refe
15 "Leukocytes are the cells of the immune system that are involved i
16 "Erythrocytes are the most common type of blood cell and the verte
17 "Pyrexia is defined as having a temperature above the normal range
18 "An abnormal mass of tissue that results when cells divide more th
19 "An ecchymosis is a subcutaneous spot of bleeding from extravasati
20 Partofspeech <- c("Noun", "Noun", "Noun", "Noun", "Noun", "Noun")
21 Pronunciation <- c( "", "", "", "", "", "", "" )
22 Etymology <- c("Derivation and composition: From French pathologie, from Ancient c
23 "Via French leucocyte, from leuco-, a Latinized combining form of c
24 "From Ancient Greek through New Latin erythro- red + -cyte cell.",
25 "From post-classical Latin pyrexia, from Ancient Greek ??????????????
26 "From neo- + -plasm, after German Neoplasma.",
27 "From New Latin, from Ancient Greek ekkhis, from ?????????????? o, I po
28 Synonym <- c("Medical science",
29 "white blood cells or leucocytes",
30

```

Figure 15 Part of the R code of the English Dataframe

After the creation of the dataframes, a simple but intuitive interface has been created. On the left side, different labels have been created in order to allow the creation in the future of different user interfaces. Now, we are going to analyse the *formal features* interface. This interface gives the opportunity to the users to select an English technical term from a drop-down menu and immediately after the selection all the pieces of information related to that technical term will be display on the screen in English and Italian.

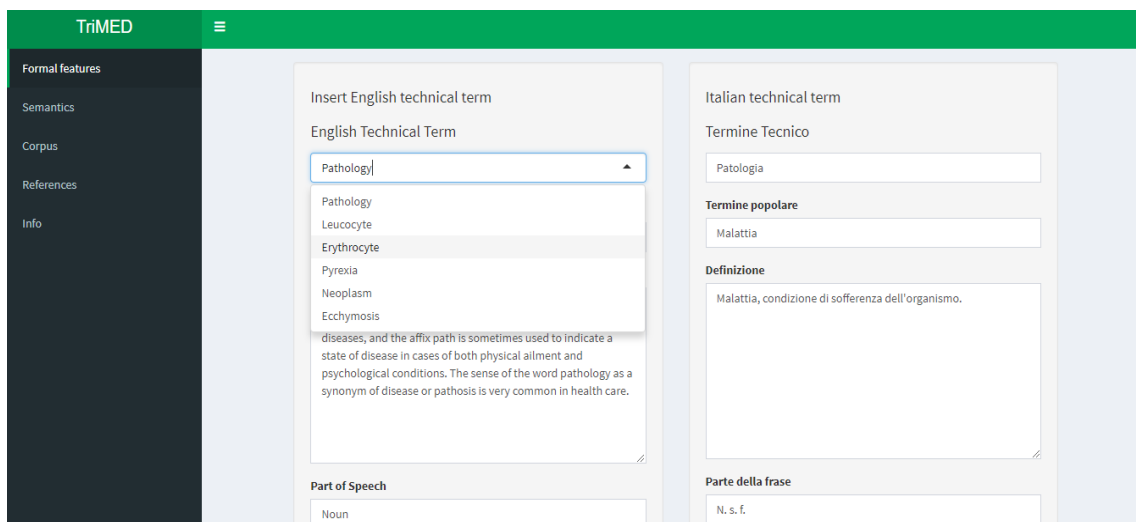


Figure 16 User Interface – Drop-down menu

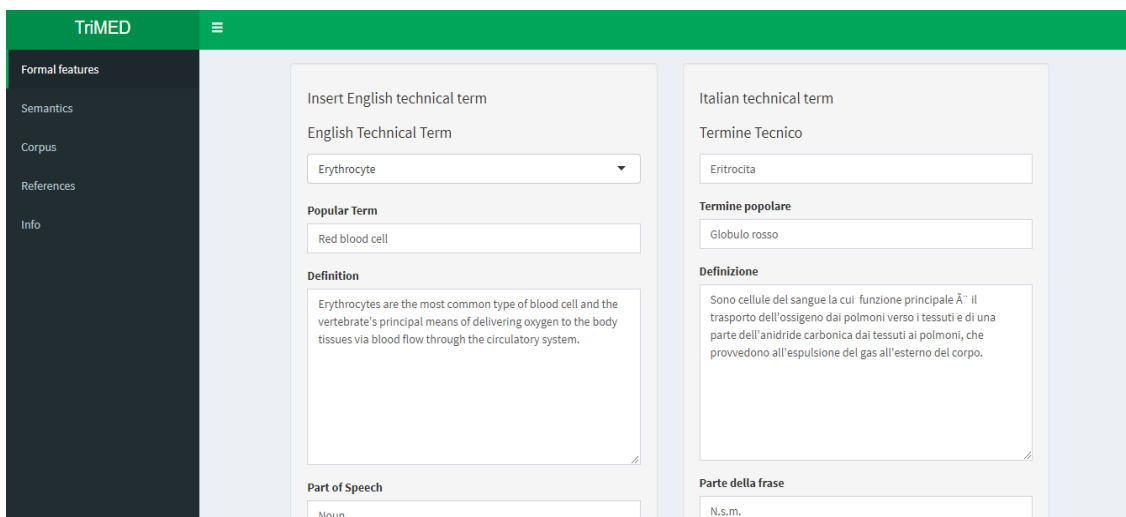


Figure 17 User Interface

As you can see from figure 16 and 17, if the term “Erythrocyte” is chosen, all the other fields automatically update themselves with all the related information.

Another peculiarity of my plug-in is that it allows the user to download for free the terms and the relative information in a file with *.tbx* format. Other plug-ins (e.g. *SDL Multiterm*) allow to create terminological records and to download them but usually they are paid plug-ins or use proprietary extension. This format is based on the XML one, it is usually used for the creation and the exchange of the *Translation Memories* (TM). Moreover, it is supported by the main paid and open source CAT tools. This functionality of my plug-in can simplify the exchange of data between translators.

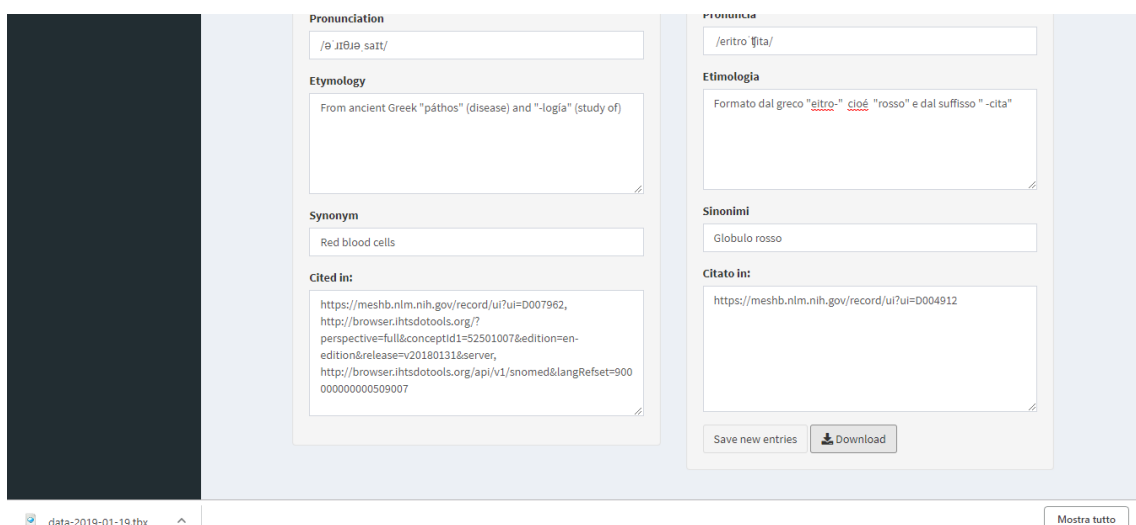


Figure 18 Download button

At the bottom of the page there is the “Download” button. If the user clicks on it a TBX file named with the date of the download will be saved in the computer. In order to open the file, a program which can handle TBX file is need. For instance, *Notepad++*.

```

1 <tbx>
2 <termEntry>
3 <langSet xmlang="en-GB">
4 <English_term>Pathology</English_term>
5 <Popular_term>Illness</Popular_term>
6 <Definition>Used as a count noun, a pathology, pl. pathologies, can also refer to the predicted or actual progression of particular diseases, and the affix
7 <Partofspeech>Noun</Partofspeech>
8 <Pronunciation></Pronunciation>
9 <Etymology>Derivation and composition: From French pathologie, from Ancient Greek páthos, disease and -logĀ-a, study of.</Etymology>
10 <Synonym>Medical science</Synonym>
11 <Cited><https://meshb.nlm.nih.gov/record/ui?ui=D010336, http://browser.ihtsdotools.org/?perspective=full&conceptIdl=384595002&edition=en-edition&
12 </langSet>
13 <langSet xmlang="it-IT">
14 <Terminetaliano>Patologia</Terminetaliano>
15 <Terminetpopolare>Malattia</Terminetpopolare>
16 <Definizione>Malattia, condizione di sofferenza dell'organismo.</Definizione>
17 <Partedellafraze>N. s. f.</Partedellafraze>
18 <Pronuncia></Pronuncia>
19 <Etimologia>comp. di pato- e -logia</Etimologia>
20 <Sinonimi>Morbo, malessere, malore</Sinonimi>
21 <Citazioni><https://meshb.nlm.nih.gov/record/ui?ui=D010336</Citazioni>
22 </langSet>
23 </termEntry>
24 <termEntry>
25 <langSet xmlang="en-GB">
26 <English_term>Leucocyte</English_term>
27 <Popular_term>White blood cell</Popular_term>
28 <Definition>Leukocytes are the cells of the immune system that are involved in protecting the body against both infectious disease and foreign invaders.</D
29 <Partofspeech>Noun</Partofspeech>
30 <Pronunciation></Pronunciation>
31 <Etymology>Via French leucocyte, from leuco-, a Latinized combining form of Greek leukos white, clear, from PIE root leuk- light, brightness + -cyte cell.<
32 <Synonym>White blood cells or leucocytes</Synonym>
33 <Cited><https://meshb.nlm.nih.gov/record/ui?ui=D007962, http://browser.ihtsdotools.org/?perspective=full&conceptIdl=82501007&edition=en-edition&
34 http://browser.ihtsdotools.org/api/v1/lanoms&langRefset=9800000000058907</Cited>
35 </langSet>
36 <langSet xmlang="it-IT">
37 <Terminetaliano>Leucocita</Terminetaliano>

```

Figure 19 Notepad++

Figure 19 shows the structure of the TBX file. This type of file is composed by nodes. The root node is `<tbx>` and it includes all the terms and the related information. The main child node is `<termEntry>` because it marks the beginning and the end of a term entry and this means that it includes all the information in English and Italian related to one of the technical terms. A child of the node `<termEntry>` is the node `<langSet>`, which provide information about the language. For instance, the node `<langSet xmlang="en-GB">` means that the following information will be in English. The last children nodes indicate which type of information can be find in that node (e.g. `<Definition>`, `<Partofspeech>`, `<Pronunciation>`).

Besides downloading the data, my plug-in allows to enter new entries and save them in the dataframes by means of another button located at the bottom of the page near the downloading button. Therefore, it is possible to make the dataframe broader before downloading it.

Therefore, my open-source plug-in offers to translators and terminologists useful functionalities for free one the contrary of the existing tools.

CHAPTER 5

5. The R Plug-In at work

The peculiarities and difficulties of the medical translation have been discussed in the second and third chapters and all the main available tools for the computer-assisted translation have been presented and analysed in the first one. The aim of this chapter is to explain in detail how this new plug-in would be a useful tool during a medical translation and how it differs from the other existing tools through a practical example. Let's suppose that we want to translate into Italian a short text, like the following abstract of a medical paper.

Abstract

Leukocytes can't perform phagocytosis in blood stream. Blood velocity prevents phagocytosis because there is no time for leukocyte to recognize and catch bacteria. Bloodstream clearance from pathogens is performed by erythrocytes. During motion in bloodstream erythrocytes become charged by triboelectric effect. This charge attracts bacteria and fixes them on the surface of erythrocyte, then bacteria are engulfed and killed by haemoglobin oxygen. In bloodstream, leukocyte thin-wrinkled elastic membrane can't be charged by triboelectric effect and so leukocyte can't catch bacteria by means of electrostatic attraction force. Leukocytes engulf and kill bacteria out of blood circulatory system: in tissues, lymph nodes, slow velocity lymph, etc. Erythrocyte and leukocyte are bactericidal partners: the first kills bacteria in bloodstream, the second kills them locally, out of blood circulation.³⁴

As it is possible to see, in the text there are two terms (*Leukocytes* and *Erythrocytes*) which are included in my Plug-in as well. At this point, we have different options. If we are already quite experts in the field, we can look up in the plug-in the definitions only to be sure to have understood correctly the text. If we are not so expert and we want to use a CAT tool, we can download the *.txb* file and use it as a glossary inside a translation environment. Before downloading it, it would be also possible to add other useful terms with all the related information. For instance, we could create glossary entries for the following terms: *phagocytosis*, *haemoglobin*, *lymph nodes*. In fact, these terms are not

³⁴ <https://www.ncbi.nlm.nih.gov/pubmed/25259410>

immediately so clear and they may need to be analysed in order to translate them in the proper way.

The screenshot shows the TriMED interface with a sidebar on the left containing 'Formal features', 'Semantics', 'Corpus', 'References', and 'Info'. The main area is divided into two columns. The left column is titled 'Insert English technical term' and contains:

- 'English Technical Term' dropdown menu with 'Phagocytosis' selected.
- 'Popular Term' text input field.
- 'Definition' text area containing: 'The ingestion of bacteria or other material by phagocytes and amoeboid protozoans.'
- 'Part of Speech' dropdown menu with 'Noun' selected.

 The right column is titled 'Italian technical term' and contains:

- 'Termine Tecnico' text input field with 'Fagocitosi' entered.
- 'Termine popolare' text input field.
- 'Definizione' text area containing: 'La fagocitosi è la capacità posseduta da diverse cellule di ingerire materiali estranei e di distruggerli.'
- 'Parte della frase' text input field with 'N. s. f.' entered.

Figure 20 Example of a new term entry

Once the file has been downloaded, it can be used in a CAT Tool. For instance, if we choose to use *OmegaT* the file should be put in the “glossary” folder, in this way the terms contained in the file will appear in the glossary area and they will help during the translation of the text.

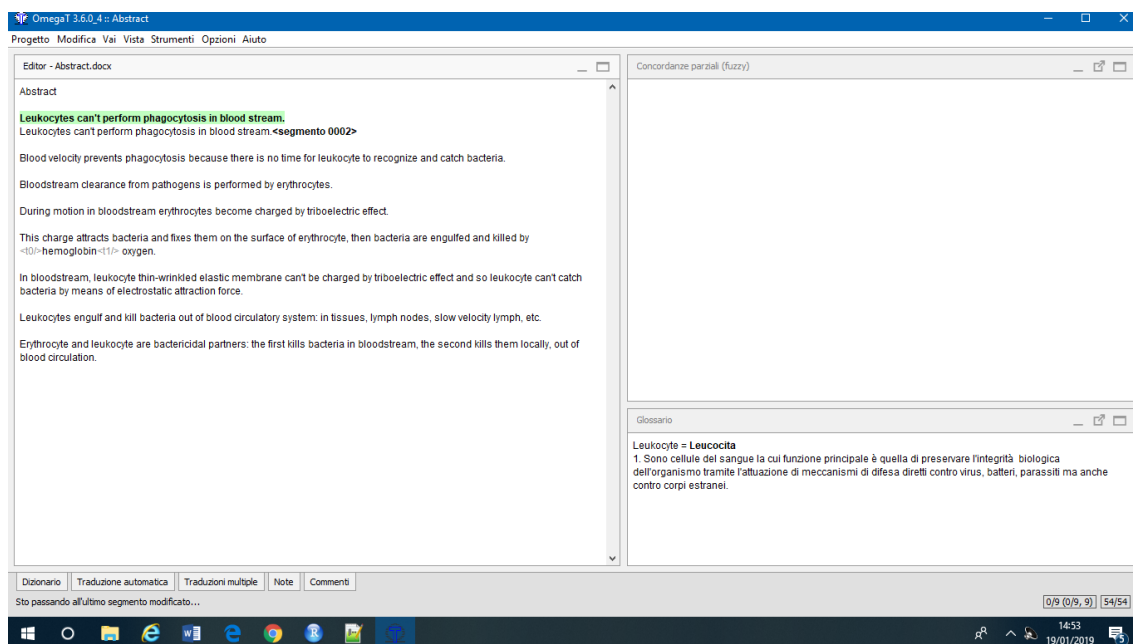


Figure 21 Translation in OmegaT

CONCLUSION

This dissertation explores the differences between the main approaches available for the localization of web sites and presents a new own open source plug-in written in R and design to manage a medical termbase for the translation of medical documents.

Chapter One outlines the long historical background of translation. The Computer-Assisted Translation (CAT) and CAT Tools Chapter Two are explained in detail in Chapter Two. Chapter Three starts with a brief description of the medical translation history and then it gives a definition of *specialized languages* and it explains what makes them different from the general language. Chapter Four presents the *TriMED* Project and my Plug-in. The last Chapter intends to demonstrate the advantages and the opportunities offered by my Plug-in during the translation of a medical website using a practical example.

This work led to the conclusions that Computer Science has become really important for translation, but the perfect CAT tool has not been invented yet, even if there are many tools at translator's disposal. Moreover, medical translation has always been challenging for different reasons and its peculiarities could cause problems to translators and therefore an appropriate tool is needed. However, in this context my plug-in is presented as a new innovative tool that can be really useful to localizers and translators because it helps them in the management of terminological *termbases*.

During the translation of medical documents, the correct translation of the terms is an important part. Therefore, I have created two dataframes, one in English and one in Italian but not only with the medical technical term and its translation but also with other important information that can come in handy to a translator. In fact, for each entry, I have inserted also the corresponding popular term, the definition, the pronunciation, which part of speech is, the etymology, the synonyms and in which reliable websites has been cited. The first dataframe has all the entries and all the related information in English and the other one the same data but in Italian. It is important to know in which contexts using the technical term is better than using the popular term and vice versa. Besides the terms that I have entered, my plug-in allows the users to insert and save new terminological records. After the creation of the dataframes, a

simple but intuitive interface has been created. This interface gives the opportunity to the users to select an English technical term from a drop-down menu and immediately after seeing all the relative information in English and Italian. Another peculiarity of my plug-in is that it allows the user to download for free the terms and the relative information in a file with *.tbx* format. Other plug-ins (e.g. *SDL Multiterm*) allow to create terminological records and to download them but usually they are paid plug-ins or use proprietary extension. This format is based on the XML markup language, it is usually used for the creation and the exchange of the *Translation Memories (TM)*. Moreover, it supported by the main paid and open source CAT tools. This functionality of my plug-in can simplify the exchange of data between translators.

Other implementations and tests should be carried out to understand which kind of functions could be more useful to the final users. For instance, my tool could be tested by professional translators in order to ask them what they would improve or add to the plug-in. The choice to make it open-source is not casual and it allows others after me to work on it, but I think that is already a really good and useful tool.

REFERENCES

Bibliography

Anthony, L. 2004 *Proceedings of IWLeL 2004: An Interactive Workshop on Language E-learning 2004* Tokyo: Waseda University.

Arráez-Aybar, L.A. & Bueno-López, J.L. & Raio, N. 2015 *Toledo School of Translators and their influence on anatomical terminology* Amsterdam: Elsevier.

Bhayro, S. & Hawley, R. & Kessel, G. & Pormann P.E. 2013 *The Syriac Galen Palimpsest: Progress, Prospects And Problems* Journal of Semitic Studies LVIII/1 Spring 2013 Oxford University Press: 131-148.

Cappelli, G. 2008 *The translation of tourism-related websites and localization: problems and perspectives* Pisa: Università di Pisa.

Castells, M. 2001 *The internet galaxy* Oxford: Oxford University Press.

Cortelazzo, M. 1994. *Lingue speciali* Padova: UNIPRESS.

DeCamp, J. & Karsch, B.I. & Kristy, J. K. & Muegge, U. & Wright, S.E. 2016 *T&I Descriptions* Atlanta: NAJIT.

Delisle J. & Woodsworth, J. 2012 *Translators through history* Amsterdam/Philadelphia: John Benjamins B.V.

Dewar, J.A. 1998 *The information age and the printing press: looking backward to see ahead* California: Rand Papers.

Di Nunzio, G.M. & Vezzani, F. 2018 *Using R Markdown for Replicable Experiments in Evidence Based Medicine* Padova: Università di Padova.

Dollerup, C. 2004 *Translation today: a global view* Denmark: University of Copenhagen.

Farghal, M. 2013 *'Word-for-word' or 'sense-for-sense' translation: ruminating the age-long polemics* Kuwait: Kuwait University.

Fischbach, H. 1962 *Problems of Medical Translation*. Bulletin of the Medical Library Association. July 1962; Vol. 50, No. 3, pp 462–472.

Fischbach, H 1998 *Translation and Medicine* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Foo, J. 2012 *Computational Terminology: Exploring Bilingual and Monolingual Term Extraction* Linköping: Linköping University.

Garcia, I. 2009 *Beyond Translation Memory: Computers and the Professional Translator* University of Western Sydney The Journal of Specialised Translation Issue 12 July.

Gea-Valor, M. L. & García-Izquierdo, I. & Esteve, M.J. 2010 *Linguistic and Translation Studies in Scientific Communication* Bern: Peter Lang ag.

Grammaticos, P.C. & Diamantis, A. 2008 *Useful known and unknown views of the father of modern medicine, Hippocrates and his teacher Democritus* Thessaloniki: HSNM.

Kingdon, R 1980 *Review of The Printing Press as an Agent of Change*, Library Quarterly, Vol. 50, no. 1, p. 140.

Loiacono, E.T. & Watson, R.T. & Goodhue D.L. 2002 *WebQual™: A Measure of Web Site Quality* Worcester: Worcester Polytechnic Institute.

Maxwell Chandler, H. & O'Malley Deming, S. 2012 *The Game Localization Handbook* 2nd ed. Ontario: Jones & Bartlett Learning.

Mayhew, D. J. 1999 *The usability engineering lifecycle: a practitioner's handbook for user interface design* Burlington: Morgan Kaufmann.

Minnis, A. J. 2010 *Medieval Theory of Authorship: Scholastic Literary Attitudes in the Later Middle Ages*, 2nd edition. Philadelphia: University of Pennsylvania Press.

Munday, J. 2001 *Introducing Translation Studies Theories and applications* London: Routledge.

Musacchio, M. T. & Sostero, G.H. 2007 *Tradurre: formazione e professione* Padova: CLEUP.

Sarcevic, S. 2006 *Insights Into Specialized Translation* Bern: Peter Lang.

Serianni, L. 2005 *Un treno di sintomi* Milano: Garzanti Libri.

Snell-Hornby, M. 2006 *The Turns of Translation Studies* Amsterdam/Philadelphia: John Benjamins B.V.

Thussu, D.K. 2006 *International Communication: Continuity and Change*, 2nd edition. London: Bloomsbury Academic.

Tschanz, David W. 2003 *Hunayn bin Ishaq: The Great Translator*. Journal of the International Society for the History of Islamic Medicine: 39–40.

Vezzani, F. & Di Nunzio, G.M. 2019 *Computational Terminology in eHealth* Padova: Università di Padova.

Tardy, C. 2004 *The role of English in scientific communication: Lingua franca or Tyrannosaurus rex?* Journal of English for Academic Purposes, Vol. 3, Issue 3, p. 247-269.

Ebook

Brenes, P. 2017 *Terminology for beginners*: <http://inmyownterms.com/>

Sitography

BBC. *What are plug-ins?* Available at <http://www.bbc.co.uk/webwise/guides/about-plugins> (accessed 3 January 2019).

Deskshare, *Site Translator*. Available at <https://www.deskshare.com/translation-software.aspx> (accessed 30 October 2018)

European Parliament, *Terminology Coordination*. Available at <http://termcoord.eu/what-we-do/> (accessed 27 October 2018).

European Parliament, *What is terminology and how to write a glossary* 2015. Available at <http://termcoord.eu/2015/11/what-is-terminology-and-how-to-write-a-glossary/> (accessed 2 November 2018).

Figure1, *Expand your clinic knowledge*. Available at <https://figure1.com/> (accessed 3 January 2019).

GitHub. *What is a Data Frame?* Available at <https://github.com/mobileink/data.frame/wiki/What-is-a-Data-Frame%3F> (accessed 3 January 2019).

In my own terms by Patricia Brenes, *A termbase: What you should know*. Available at <http://inmyownterms.com/a-termbase-what-you-should-know/> (accessed 30 October 2018).

Language Insight, Google Translate VS Professional Translation. Available at <https://www.languageinsight.com/blog/2017/google-translate-vs-professional-translation/> (accessed 27 October 2018).

Laurence Anthony, *AntConc*. Available at <http://www.laurenceanthony.net/software/antconc/> (accessed 28 October 2018).

Lexical Analysis Software Ltd, *WordSmith Tools*. Available at <https://www.lexically.net/wordsmith/> (accessed 28 October 2018).

Matthias Hüning, *TextSTAT*. Available at <http://neon.niederlandistik.fu-berlin.de/static/textstat/TextSTAT-Doku-EN.html> (accessed 28 October 2018).

Medscape, *Vaccines*. Available at <https://search.medscape.com/search/?q=vaccines> (accessed 13 November 2018).

Mozilla, *Pontoon*. Available at <https://pontoon.mozilla.org> (accessed 29 October 2018)

Omega T, *Omega T a free translation memory tool*. Available at <http://omegat.org/> (accessed 28 October 2018).

PubMed, *Erythrocyte and leukocyte: two partners in bacteria killing*. Available at <https://www.ncbi.nlm.nih.gov/pubmed/25259410> (accessed 27 October 2018).

Research and Development Unit for English Studies, Birmingham City University, *WebCorp*. Available at <http://www.webcorp.org.uk/live/> (accessed 28 October 2018).

SERMO. Available at <http://www.sermo.com/> (accessed 3 January 2019).

SDL, *SDL Trados Studio 2018 Professional*. Available at <https://www.sdl.com/software-and-services/translation-software/sdl-trados-studio/faqs.html> (accessed 28 October 2018).

Univadis, *Il portale di riferimento per il professionista della salute*. Available at <https://www.univadis.it/> (accessed 3 January 2019).

UpToDate, Inc. Available at <https://www.uptodate.com/home> (accessed 3 January 2019).

U.S. Department of Health and Human Services (HHS), *Vaccines basics*. Available at <https://www.vaccines.gov/basics/index.html> (accessed 13 November 2018).

U. S. National Library of Medicine, *Neoplasms*. Available at <https://meshb.nlm.nih.gov/record/ui?ui=D009369> (accessed 12 November 2018).

WHO, *Globalization*. Available at <http://www.who.int/topics/globalization/en/> (accessed 27 October 2018).

RIASSUNTO IN ITALIANO

Questa tesi di laurea magistrale ha come obiettivo quello di esplorare le differenze tra i diversi approcci disponibili per la localizzazione di siti Web e la successiva presentazione del plug-in da me sviluppato nel linguaggio di programmazione R e ideato per facilitare la gestione di glossari terminologici medici utili durante la traduzione di documenti medici. La decisione di dedicare il mio elaborato proprio a questo tema è nata dal fatto che gli strumenti informatici e le nuove tecnologie hanno influenzato e stanno caratterizzando anche il campo della traduzione e ritengo che sia importante per i traduttori, localizzatori e lessicologi sapere quali strumenti ci sono a loro disposizione e come utilizzarli. Inoltre, il campo della localizzazione e quello degli strumenti informatici in generale offrono molto materiale di lavoro per i traduttori e sono dei campi in continua crescita ed evoluzione, per questo motivo credo siano degni di un'adeguata analisi. Infine, dopo aver analizzato i CAT Tool e gli strumenti attualmente disponibili, credo che il plug-in da me sviluppato possa essere uno strumento utile e innovativo.

Tuttavia, per comprendere il mondo della traduzione di oggi, è utile ripercorrere il suo percorso dall'antichità ai giorni nostri per vedere come si è arrivati dalle prime traduzioni su pergamena alla traduzione assistita dai computer.

L'arte della traduzione, infatti, ha origini molto antiche. Le prime testimonianze di trasposizioni di testi da una lingua ad un'altra risalgono all'antico Egitto. Nel corso dei secoli la società e le sue priorità sono cambiate. Per la maggior parte delle civiltà più antiche l'argomento più importante e maggiormente discusso era la religione e proprio per questo motivo la maggior parte degli scritti più antichi sono dedicati proprio a questo argomento. Infatti, anche se il motivo di maggior incontro e confronto tra le varie culture e comunità era il commercio, il culto giocava un ruolo molto significativo all'interno della società. Per esempio, nell'antico Egitto i preti erano trattati con molto rispetto e reverenza perché si credeva che si prendessero cura dei bisogni degli dei e per lo stesso motivo erano spesso delle persone ricche e potenti e con un'educazione al di sopra della media. Proprio in Egitto e in questo contesto nacque il primo grande progetto di traduzione noto con il nome di *Septuagint*. Il suo obiettivo era la traduzione

delle Sacre Scritture dall'ebraico al greco e si crede che questa traduzione fu commissionata dal re d'Egitto in persona tra il 250 e il 130 A.C.

Insieme alla nascita della traduzione, nacquero anche le prime discussioni a riguardo. In particolare, durante questo primo periodo il dibattito riguardava la contrapposizione tra la corrente di pensiero che sostiene che la traduzione deve essere fatta "parola per parola" e quella che invece crede che una traduzione "senso per senso" sia più appropriata e significativa.

La religione ha continuato a giocare un ruolo importante all'interno della società anche durante il Medioevo. In quel periodo la Chiesa Cattolica aveva un enorme potere nelle società occidentali. È interessante notare come la struttura sociale e culturale abbia influenzato l'arte della traduzione in ogni periodo storico e come i testi tradotti abbiano a loro volta avuto un ruolo nelle comunità. Ad esempio, visto che la Chiesa aveva un enorme potere durante il medioevo e molte delle traduzioni di quell'epoca sono legate alla religione. Tuttavia, c'era anche un altro grande centro di potere che commissionava delle traduzioni, ovvero lo Stato.

Sotto questi due grandi poteri nacquero tre scuole di traduzione: la scuola di Baghdad, la scuola di Toledo e un gruppo di letterati che nella seconda metà del quattordicesimo secolo tradusse le *auctoritates*.

La *scuola di Baghdad* era composta da calligrafi, illustratori, trascrittori e traduttori. La particolarità di questo movimento è che lavorò partendo da testi non arabi provando poi a produrre e tradurre interessanti manoscritti in arabo. Una delle più importanti opere di questa scuola fu la traduzione dal Greco all'arabo del *De Materia Medica*.

Tuttavia, all'inizio del Medioevo il latino era la principale lingua franca e molte opere filosofiche e religiose furono tradotte in questa lingua dalla *Scuola di Toledo*.

Infine, il terzo Gruppo di intellettuali si dedicò alla traduzione delle *auctoritates*, testi riconosciuti come autorevoli perché approvati dalla Chiesa come testimonianza di quella che si credeva essere la Verità. Le *Sacre Scritture* erano considerate l'*auctoritas* per eccellenza.

Nel 1439 l'invenzione della stampa moderna da parte di Johannes Gutenberg rivoluzionò il mondo dell'editoria. Prima di quel momento tutti i libri erano scritti a mano perciò erano molto costosi e serviva molto tempo per produrre ognuno di essi. Per questa ragione venivano considerati beni di lusso pensati per persone ricche e con una

certa educazione. Dopo l'invenzione della stampa i libri diventarono alla portata di molte più persone anche di classi meno istruite. Questi nuovi lettori non erano capaci di leggere i testi scritti in latino e quindi avrebbero volute poter leggere i testi nella propria lingua. Considerando che prima della stampa solo i più ricchi e colti erano in grado e potevano permettersi di leggere le Sacre Scritture, non sorprende il fatto che la Bibbia tradotta fu il primo libro ad essere stampato con questa nuova tecnica e il primo best-seller della storia. Inoltre, visto gli sconvolgimenti che stava subendo la Chiesa in quel periodo non sorprende neanche che due delle più rinomate traduzioni quel periodo siano il *Nuovo Testamento di Tyndale* e la *Bibbia di Lutero*.

Queste traduzioni portavano con sé anche grandi responsabilità sia verso i credenti che verso la Chiesa perché diverse traduzioni potevano portare a diverse interpretazioni della parola di Dio e quindi potevano addirittura essere considerate eretiche.

Tuttavia, queste traduzioni promossero anche l'uso delle lingue popolari contribuendo alla nascita e sviluppo delle principali lingue moderne europee.

Ancora una volta è possibile notare lo stretto legame tra traduzione e società.

E lo si nota ancora dopo il Rinascimento, quando la società cambiò. La Chiesa non era più così potente mentre i movimenti nazionali guadagnavano sempre più consensi. In questo periodo in cui si stavano sviluppando le identità nazionali, le nuove tecnologie e la diffusione di queste nuove lingue portarono anche allo sviluppo della letteratura e delle opere teatrali nazionali. Di conseguenza molti lavori di traduzione di quel periodo non riguardano più la religione bensì libri e opere teatrali.

Nonostante questa sua lunga storia, la traduzione diventò una materia di studio indipendente solo durante lo scorso secolo. Infatti, fu solo durante il ventesimo secolo che nacque la disciplina chiamata *Translation Studies*.

Per la prima volta, traduzione e interpretariato furono viste come due materie di studio differenti. L'interpretazione infatti è legata al processo di traduzione orale di discorsi mentre la traduzione è più legata alla parte scritta.

Da questo momento in più la traduzione è diventato un ampio settore di studi. Nuove tecniche e approcci a diversi tipi di testi sono stati sviluppati, testati e discussi. Inoltre, i principali settori della traduzione non sono più solo religione, letteratura e teatro ma la traduzione ha ampiamente esteso i propri confini anche a molti altri settori quali la scienza, i manuali tecnici, la medicina, la politica e il commercio.

La traduzione oggi è considerata come una materia accademica indipendente che coinvolge diverse altre discipline quali la linguistica, la semiotica, la terminologia e la letteratura comparativa.

Inoltre, oggi i traduttori generalmente scelgono un campo di traduzione come, ad esempio, il campo legislativo, economico, medico o letterario e seguono corsi di specializzazione in base al settore da loro scelto. La traduzione giurata, ad esempio, ha valore legale e quindi il traduttore è legalmente responsabile per gli eventuali errori di interpretazione del testo di partenza e le eventuali conseguenze legali. Dunque, il traduttore che decide di operare in questo campo dovrebbe avere una buona conoscenza della legge del Paese in cui opera e di quello da cui o per cui viene richiesta la traduzione. Inoltre, dovrebbe avere un'ottima conoscenza della terminologia legale.

Un'altra grande rivoluzione che ha cambiato e influenzato la storia della traduzione è stata la nascita di internet, che tra le altre cose, ha stimolato una produzione a livello mondiale di testi che necessitano di essere localizzati.

Inoltre, è nato un nuovo mercato di servizi di traduzione, localizzazione e di software per la traduzione. Sono stati inventati anche molti strumenti per aiutare e velocizzare il processo di traduzione ed è proprio questo campo che ho deciso di approfondire nell'ambito questa dissertazione.

Sulla base di come vengono utilizzati questi strumenti è possibile identificare tre diversi approcci alla traduzione: *machine-aided human translation* (MAHT), *human-aided machine translation* (HAMT) e *fully automated machine translation* (FAMT). Per gli obiettivi di questo elaborato prendiamo in considerazione principalmente la MAHT, che è anche chiamata *Computer-Assisted Translation* (CAT) perché il computer è come un assistente per il traduttore professionista che comunque svolge una gran parte del lavoro anche se il software è di grande aiuto.

Quando si tratta di localizzare siti web, in particolare quelli specializzati in un settore, la *machine-aided human translation* è il metodo più utilizzato perché uno dei principali problemi durante questo tipo di traduzione è trovare la terminologia più appropriata e i CAT tools sono pensati per aiutare i traduttori ad affrontare questo problema. Tuttavia, la maggior parte dei software esistenti furono originariamente creati per la traduzione di testi semplici mentre oggi i professionisti devono occuparsi di fonti molto più complesse. Infatti, in passato i materiali che richiedevano traduzioni erano

principalmente manuali, ricerche scientifiche, articoli e libri. Invece oggi una grande quantità di traduzioni è invece legata all'informatica e ai nuovi media (siti web, videogiochi, software, applicazioni ecc.). Queste fonti sono molto più complesse di quelle del passato. Per esempio, durante la localizzazione di un sito web, il traduttore deve gestire un ambiente complesso che include diversi tipi di file in diversi formati. Quindi è necessario che il localizzatore sappia usare gli strumenti adeguati e che abbia buone conoscenze informatiche.

È importante per chiunque aspiri a lavorare nel campo della traduzione conoscere i software attualmente disponibili perché l'informatica può offrire un grande aiuto ai traduttori. Inoltre, le capacità informatiche sono oggi giorno fondamentali per i progetti di traduzione collaborative perché attraverso i software e internet è possibile condividere e ricevere informazioni e memorie di traduzione da altri professionisti e dai collaboratori.

Anche se è disponibile una vasta gamma di CAT tools, alcuni sono più utilizzati di altri perché mettono a disposizione alcune funzioni particolari. Di seguito l'analisi dei vantaggi e degli svantaggi di alcuni dei principali CAT tools.

OmegaT

Questo programma gratuito può lavorare con 30 formati di file diversi, tra cui quelli utilizzati per scrivere le pagine web quindi può tornare molto utile durante la localizzazione di un sito internet.

SDL Trados Studio

Questo è forse il software più utilizzato dai professionisti, può gestire più di 70 tipi di formati. Questo programma però, al contrario di *OmegaT*, è a pagamento.

Site translator

È un software a pagamento specificatamente pensati per a traduzione di siti web.

SDL Passolo

Questo software a pagamento appartiene alla stessa famiglia di *SDL Trados Studio* ma è stato ideato specificatamente per la localizzazione di software.

Poontoon

Poontoon è un enorme progetto di localizzazione promosso da Mozilla che ha come obiettivo quello di rendere Firefox disponibile in più lingue possibile grazie al lavoro di una comunità globale di localizzatori provenienti da tutto il mondo.

A questo punto è importante notare che nonostante ci siano così tanti strumenti per la traduzione professionale, milioni di persone continuano ad affidarsi a strumenti per la traduzione automatica (es. *Google Translate*, *Bing Translate*) anche se difficilmente si ottiene una traduzione di elevata qualità utilizzando questi metodi, soprattutto quando si tratta della traduzione di testi tecnici.

Detto questo, una delle più importanti funzioni di un CAT tool è in che modo aiuta il traduttore nella gestione della terminologia perché è uno dei principali aspetti che determina la qualità di una traduzione. La storia della terminologia è antica quasi quanto quella della traduzione e alcuni traduttori oggi sono diventati terminologi a tempo pieno.

La terminologia gioca un ruolo particolarmente importante nelle traduzioni tecniche, scientifiche e mediche. Per questo motivo la maggior parte dei CAT tools consente ai traduttori di inserire Memorie di Traduzione e glossari preesistenti prima di iniziare la traduzione. Ovviamente l'avanzamento degli strumenti informatici ha portato un cambiamento anche nella creazione e nella struttura delle banche terminologiche e oggi abbiamo a disposizione molti strumenti per la creazione di glossari e di database terminologici. Inoltre, oggi è possibile condividere e ricevere facilmente glossari da tutto il mondo oppure collaborare alla creazione di nuovi.

Un buon glossario o una buona banca dati terminologica sono essenziali al fine di una buona traduzione, specialmente per quelle settoriali. Quindi è importante sapere come creare un glossario da zero. Hellmut Riediger, professore di traduzione e terminologia, ha definite tre fasi per la creazione di un database terminologico di buona qualità. Durante la prima fase il lavoro deve essere organizzato, va definito il dominio per cui si vuole creare il glossario e bisogna analizzare un corpora di documenti legati all'argomento. Nella fase successiva vanno selezionati i termini e convalidati in ogni lingua del glossario e bisogna compilare le schede terminologiche. Nell'ultima fase il glossario va analizzato e rivisto prima di essere pubblicato e utilizzato. Dunque la

creazione di un buon glossario richiede molto lavoro ma poi questo verrà ripagato durante lo svolgimento della traduzione.

Inoltre, oggi sono disponibili molti strumenti informatici che aiutano nella creazione dei glossari e alcuni di essi sono *WordSmith Tools*, *TextStat*, *AntConc* e *WebCorp10*.

Creare un *termbase* è anche meglio di creare un semplice glossario perché contiene non solo la terminologia ma anche tutte delle informazioni collegate ad ogni singolo termine, che vengono chiamate *metadati*. Queste informazioni sono contenute in più lingue per ogni voce del *termbase* e sono molto utili durante la traduzione.

Avere un buon *termbase* elimina le ambiguità accelerando la traduzione e garantendo l'utilizzo corretto del linguaggio settoriale. I formati più comuni per la creazione e condivisione di *termbase* sono *.tmx* (Translation Memory eXchange), *.tbx* (TermBase eXchange) e *.xliff* (Localisation Interchange File Format).

Come già detto, i glossari e le memorie di traduzione sono molto importanti quando si tratta di traduzioni tecniche, scientifiche e mediche. Nell'ambito di questa tesi abbiamo deciso di concentrarci in particolare sull'ambito medico.

La storia della traduzione è molto antica, infatti è possibile affermare che nell'Europa occidentale iniziò con i lavori di Ippocrate (460-370 A.C.). Tuttavia, la medicina è una scienza in continua evoluzione e innumerevoli progressi scientifici sono stati fatti durante il diciannovesimo e ventesimo secolo. In quello stesso periodo l'Inglese è diventato la *lingua franca* per la maggior parte dei settori tecnici e scientifici e anche nel campo medico. Nell'era di Internet, l'inglese è diventato la lingua dominante nella comunicazione scientifica e non ma proprio per questo motivo molti contenuti necessitano di essere tradotti per renderli fruibili da tutti gli utenti.

Anche nei siti web medici, nelle pubblicazioni online, nei video e nelle conferenze online, l'inglese è diventata la lingua standard. Quindi, al giorno d'oggi è necessario che un traduttore che si vuole specializzare nel campo medico sappia bene l'inglese ma soprattutto i termini tecnici.

La terminologia medica infatti è parte delle cosiddette *lingue speciali*.

Per lingua speciale si intende una varietà funzionale di una lingua naturale, dipendente da un settore di conoscenze o da una sfera di attività specialistici, utilizzata, nella sua interezza, da un gruppo di parlanti più ristretto della totalità dei parlanti la lingua in cui quella speciale è una varietà, per soddisfare i bisogni comunicativi (in primo luogo quelli referenziali) di quel settore specialistico; la lingua speciale è costituita a livello lessicale da una serie di corrispondenze aggiuntive rispetto a quelle generali e comuni della lingua e a quello morfosintattico da un

insieme di selezioni, ricorrenti con regolarità, all'interno dell'inventario di forme disponibili nella lingua (Cortelazzo 1994: 7).

Quindi chi è in grado di parlare questa lingua può capire da un solo termine, dei concetti complessi perché conosce già quelle nozioni e non ha bisogno di ulteriori spiegazioni. Per esempio, il termine medico *piressia* indica l'innalzamento della temperatura corporea come meccanismo sviluppato dal sistema immunitario per ridurre la gravità di una malattia prevenendo che virus e batteri si moltiplichino. Questo termine può sembrare scontato a coloro che sono esperti nel campo e medico e può richiamare nella loro mente tutte le nozioni relative a questo termine e le circostanze in cui si verifica questo fenomeno. Tuttavia, per le persone comuni *piressia* potrebbe suonare come un termine complesso dal significato "oscuro" e per esprimere il concetto praticamente simile utilizzerebbero il termine *febbre*. In ogni caso sarebbe importante notare che l'immagine mentale prodotta non sarebbe la stessa. Infatti, la parola febbre per le persone comuni evocerebbe solo il concetto di innalzamento della temperatura senza per forza implicarne le cause e le spiegazioni tecniche del fenomeno.

All'interno della lingua comune sono contenuti legami abitati tecnici e scientifici. Quindi sono contenuti anche molti termini legati alla medicina ma queste parole non sono considerate parte delle *lingue speciali*. Infatti, le parole che costituiscono le lingue speciali nella maggior parte dei casi implicano concetti più ampi e/o tecnici che solo gli esperti del settore possono comprendere.

Se prendiamo in considerazione i materiali medici presenti online, possiamo trovare principalmente due tipi di siti web medici: quelli pensati per utenti già esperti in materia e quelli per i più inesperti. Il secondo tipo è spesso dedicato a chi non ha competenze mediche ma vuole approfondire argomenti di natura medica per interessi personali, per motivi di studio o per comprendere meglio alcuni concetti. Mentre i siti più specialistici sono dedicati a professionisti ed esperti utilizzano un linguaggio più tecnico e spiegano concetti più complessi al fine di confrontarsi e magari portare all'avanzamento scientifico.

Questi siti sono i più complessi da localizzare e non esiste ancora uno strumento informatico in grado di soddisfare tutte le necessità dei traduttori durante questo tipo di lavoro. Il desiderio di colmare questa lacuna ha portato il professor Giorgio Maria Di Nunzio in collaborazione con la dottoranda Federica Vezzani a creare il progetto *TriMED*.

TriMED, is a digital library of terminological records designed to satisfy the information needs of different categories of users within the healthcare field: patients, language professionals and physicians. *TriMED* offers a wide range of information for the purpose of simplification of medical language in terms of understandability and readability inspired (Vezzani and Di Nunzio 2019:1).

Questa nuova risorsa informatica è al momento disponibile in tre (inglese, italiano e francese) e può essere visualizzata in tre differenti modalità in base alle esigenze dell'utente.

C'è una visualizzazione pensata per i pazienti che lavora con uno strumento di recupero informazione, ovvero consenta agli utenti di trovare informazioni utili riguardanti termini tecnici. La seconda visualizzazione è pensata per traduttori e altri esperti in materia linguistica. In questo caso *TriMED* lavora come un traduttore automatico fornendo la traduzione del termine tecnico, il corrispondente termine popolare, la definizione e un'analisi del termine in entrambe le lingue di lavoro.

L'ultima visualizzazione è pensata per i medici professionisti. *TriMED* ha l'obiettivo di aiutare i ricercatori offrendo traduzioni aggiornate, regolari e affidabili dei termini medici insieme alle relative informazioni e termini popolari equivalenti.

Prendendo ispirazione dal progetto *TriMED*, ho sviluppato un mio plug-in con l'obiettivo di offrire un nuovo strumento di traduzione che combini le funzioni degli strumenti attualmente esistenti con alcuni miglioramenti.

Dopo avere analizzato gli strumenti attualmente disponibili e i loro punti di forza e di debolezza, ho deciso di creare un plug-in scritto in R usando il pacchetto Shiny. R è un linguaggio e ambiente di programmazione solitamente utilizzato per nell'ambito della statistica e dei grafici ma sfruttando il Shiny offerto RStudio è possibile sviluppare applicazioni web e plug-in. In particolare, il mio plug-in è stato pensato per gestire le schede terminologiche mediche e per essere usato come complemento dei CAT tool.

La creazione di questo nuovo strumento è iniziata con la creazione di due dataframe, uno contenente le informazioni in inglese e uno contenente le corrispondenti informazioni in italiano. In particolare, nel dataframe inglese è contenuto un termine medico tecnico inglese, il corrispondente termine popolare, la definizione, la pronuncia, quale è la sua funzione grammaticale, l'etimologia, i sinonimi e in quali fonti autorevoli è stato citato. Nel dataframe italiano sono contenute la traduzione del termine e tutte le altre informazioni in italiano.

Tutte queste informazioni sono molto utili durante il processo di localizzazione

soprattutto per comprendere come e in che contesto utilizzare un determinato termine. Ad esempio, è importante capire se in un contest è meglio utilizzare il termine tecnico o quello popolare.

Oltre alle schede terminologiche da me inserite, il mio plug-in consente di inserirne delle nuove, opzione che può tornare molto utile a traduttori e terminologi.

Dopo aver creato i due dataframe ho creato un'interfaccia semplice e intuitiva. Sulla sinistra dello schermo sono state create delle etichette che in future daranno la possibilità di accedere a diverse visualizzazioni dell'applicazione. Al momento mi sono occupata però in particolare della costruzione dell'interfaccia denominata *formal features*. Questa interfaccia offre l'opportunità agli utenti di selezionare un termine tecnico medico inglese da un menu a tendina e, subito dopo la selezione, di visualizzare tutte le informazioni in italiano e in inglese relative a quel termine.

Un altro punto di forza del mio plug-in è che consente agli utenti di scaricare gratuitamente le schede terminologiche in formato *.tbx* cliccando semplicemente sul bottone "Download". È possibile aggiungere anche nuove schede terminologiche prima del download in modo da rendere il database più ampio prima di utilizzarlo. Inoltre, questo formato consente di usare le schede come glossari all'interno di ambienti di traduzione.

Altri plug-in (ed. *SDL Multiterm*) consentono di creare schede terminologiche e di scaricarle ma generalmente questi programmi sono a pagamento o utilizzano un formato proprietario. Il formato TBX invece è basato sul formato XML, che è quello solitamente utilizzato per la creazione e lo scambio delle memorie di traduzione. Un altro vantaggio del formato TBX è che viene supportato dalla maggior parte dei CAT tools, sia a pagamento che non. Dunque questa funzionalità del plug-in qui proposto può semplificare lo scambio di dati tra traduttori e localizzatori.

RINGRAZIAMENTI

Vorrei ringraziare di cuore i miei genitori, per avermi permesso di portare a termine questo percorso di studi.

Grazie anche a mio fratello Luca, a cui voglio un bene dell'anima e che sa sempre farmi tornare il sorriso.

Grazie a mia sorella Barbara, che è sempre al mio fianco e sa sempre supportarmi e sopportarmi.

Grazie alle fondamentali compagne di corso e avventure Genny e Noemi per il sostegno e le risate. Grazie per la vostra immensa pazienza nei momenti di ansia e panico, senza di voi sicuramente non avrei raggiunto questo traguardo.

Danke Johanna für deine Hilfe und die *crazy nights*. Dank dir ist mein Deutsch viel besser geworden und ich freue mich, dich als Freund zu haben.

Grazie Giacinta, al mio fianco fin dall'infanzia. Abbiamo affrontato insieme momenti difficili ma anche condiviso attimi e periodi di felicità e spero che continuerà ad essere così.

Grazie Talita perché, anche se il percorso a volte è accidentato, sei rimasta al mio fianco. Spero di poter condividere anche altri traguardi con te.

Grazie Mariarita, sai che stimo te e il tuo modo di affrontare le difficoltà della vita. Grazie per sapermi ascoltare e esserci anche se lontana.

Grazie a Fabio e al piccolo Michele, a Anna, Desy, Luca, Ilenia, Sara, Salvatore, a Stefano e Erika e agli altri amici della Ceccarelli per avermi supportata e aver condiviso con me momenti bellissimi.

Infine grazie a chi ha creduto e crede in me e non mi ha mai permesso di perdere le speranze.