An Automatic Translator from the Florentine Vernacular Language to Modern Italian Language

Luca Pavan

Institute of Foreign Languages, Faculty of Philology, Vilnius University, Vilnius, Lithuania

Email address:
pavan@panservice.it

Abstract: Along several centuries hundreds of books were written in Florentine vernacular language. The latter, however, is not easy to understand even for native Italian speakers. In 2016-2018 the author created a PC software which provides the possibility to automatically translate entire texts from Florentine vernacular language, as it is found in the literature, into modern Italian language. In this article the author intends to describe the phases of the realization of this software as well as the results of its use. The software in its dictionary currently includes about 25 000 definitions of the vernacular language (where "definitions" mean the presence of terms or phrases in vernacular literature that are replaced in the respective terms and phrases of modern Italian). Numerous studies over the years have demonstrated the limitations of machine translation, often using the error rates of translation software. Even translators that use complex algorithms created with statistical methods frequently end up generating unreliable results, sometimes questioning the very usefulness of translation software. However, if used with the necessary precautions, automatic translators can simplify a job or can help in understanding of texts even for those who know little or no foreign languages. The translator described in this article, although not immune from the defects of machine translation, can be useful both to scholars of Italian literature of past centuries, as well as to those who, while knowing Italian, want to approach texts that cannot be fully understood without the support of footnotes.

Keywords: Italian Literature, Florentine Vernacular Language, Machine Translation, Translation

1. Introduction

The idea of the automatic translator from an ancient language to the corresponding modern language stems from the need to simplify the study or reading of vernacular literature. Although like Italian, the vernacular literature cannot be fully understood by its users, even if native speakers, unless they are scholars of literature or researchers. Without footnotes that accompany the texts in the vernacular, there is no way to understand such texts on a lexical and syntactical level. The problems of comprehension are various: terms and verbs fallen into disuse, complex syntactic constructs different from modern Italian, differences in the use of pronominal particles and articles, Latinisms, truncations, etc.

The author intends to answer to the question, whether this translator should have been defined more properly as a "converter", since it compares an ancient language with the same language in a modern version. Even if the translation from the vernacular language may appear a task simplified by the proximity with the Italian language (merely because of the similarity between those two languages) many difficulties remain. For this reason, it was decided to keep the wording "translator". Moreover, it is easier to classify this software in the category of "translators".

Criticism of machine translation is essentially due to the high error rates that are produced in some cases. Although the progress of Machine Translation (MT) has been made, many scholars think it is impossible to obtain in all cases high quality translations for various reasons. This does not, however, mean that the use of computers is useless for translations, even if computers do not seem to be able to compete with a human translator [1]. Language is always in continuous transformation and even the most complex of algorithms does not have the ability to process in every situation the implicit meaning of communication in the linguistic phenomena. Some scholars, already starting from the 1940's, said that the automatic translation is rather

impossible because of language continuous changes [2]. A good translation is the main goal of automatic translators. The most recent developments of Machine Learning (ML) showed that now there are improvements in translations using statistical methods and artificial intelligence. This will be elaborated further.

2. Method

This article describes the realization of an automatic translator for Personal Computer along with the results obtained using it in the translation of texts from the vernacular language to modern Italian language. The author uses a comparative method when analyzing different book editions of Florentine vernacular literature and the output of the automatic translator.

3. The Automatic Translator

3.1. Main Features

The translator is written for PC in C programming language and it currently works on Microsoft Windows operating systems. It is, apparently, the first and currently the only translation program from Florentine vernacular literature into a modern Italian language.

The structure is simple, and the translator uses the so-called "direct translation system", *i.e.* without the contribution of syntactic-grammatical rules [3]. The program searches for those terms and phrases which appear in "definitions" and replaces them throughout the text, writing another file that is the "translation" into Italian. The terms and phrases to be replaced (*definitions*) are read by another file, which presents them in alphabetical order with their translation. The program first orders the definitions from the longest to the shortest, so that the longest definitions are replaced before the others: this is done to avoid conflicts in the substitutions. The latter may occur in the cases of replacing a term in a sentence, that if modified, would no longer be recognized.

A statistics file is also generated, indicating which replacements have been made and in which line number of the original file.

The program uses the writing of temporary files to make replacements. Each definition is compared with the text to be translated, thus you have to wait for a relatively long time to translate entire works. However, there are no limitations and you can translate works of any length. If you want to translate a short text extracted from a work, you can do it by forming a text file, and then read by the translator.

This approach does not use sophisticated algorithms and therefore may seem simple, but in the case of the vernacular language, it is quite effective, despite several encumbrances that we will be examined later in this chapter. However, this approach quite differs from the first computational models of a few decades ago, when such models were based on grammatical and syntactical rules analyzed by the computer. In many cases newer machine translation programs avoid

setting grammar rules inside them since they are very complex to manage. This translator, being designed for use on a single computer, does not make use of the statistical methods adopted by today's online translation services. The latter often uses parallel calculation, which can generate the most likely translation for a given term or a given incoming phrase in a very short time examining a very large number of *corpora* (*i.e.* monolingual or multilingual texts from which to extract statistical data). Since 1989 the automatic translators based on grammatical and syntactical rules became less popular due to the spread of new corpus-based translators [4]. At the same time, experiments to combine statistical methods with artificial neural networks continue [5]. ML showed remarkable results in solving those translation problems, "surpassing by far the performances obtained, and obtainable, from every system based on rules" [6]. The *corpus*, in this sense, is nothing more than "the representative sample of a language" [7].

The translator created by the author does not use the ML and corpora. The main drawback of the translator described herein is slowness of execution. While automatic translators on the Internet can produce a medium-length text translated almost in real time, this translator needs more time. An advantage, however, is that it is possible to translate texts of any length. A relatively large number of texts in the vernacular literature, already digitized, are present on Internet, so it is easy to obtain and use these texts for translations.

The choice of definitions to be used with the translator has of the utmost importance. The author has tried to choose such works in Florentine vernacular language which served as example for terms and phrases used in the translation process.

The works on which the translator is currently based are two editions of Petrarca's *Canzoniere*, an edition of Boccaccio's *Decameron* and an edition of the *Vocabolario della Crusca*. Recently, an edition of Petrarca's *Triumphs* has also been used. In the next chapter, we will see why the choice fell on these texts. However, in the future, the translator can be expanded by adding definitions from other works. The more definitions there are, the more accurate and precise the translations will be. The advantage is that the vernacular is a dead language, so it is not susceptible to continuous transformations and the accuracy of the translator can only increase proportionally according to the number of texts used to obtain the definitions.

From the description just made, it is clear (though not obvious) that from time to time the list of definitions can be enlarged. The list proceeds from the longest to the shortest definition and to add a new definition the following steps should be taken: (i) choosing a text to be used for the additions; (ii) translating it with the definitions that already exist; and then (iii) using the translated text to insert new missing definitions. The definitions can be arranged in any order. The program then automatically arranges the definitions from the longest to the shortest, giving an order of precedence for replacements. However, the definitions have been sorted alphabetically by the author, so that it is more convenient to search for a word or phrase in the file. There is in fact a sort of "dictionary" that can also be used by those who study works in

the vernacular.

The program could be easily adapted in the realization of other automatic translators for other languages, inserting the appropriate definitions. The approach described above can be extended, for example, to other ancient languages, allowing the creation of software with minimal means without the aid of complex statistical algorithms that seem to be indispensable for modern languages in use today.

The choice of texts has a decisive influence on the behavior of the automatic translator, as is the case with online translation corpora.

3.2. Choice of Texts

The author's choice of the basic texts from which the definitions are extrapolated was due to reasons of diffusion of the Florentine vernacular literature. A text that conveys the vernacular more than any other outside of Tuscany is certainly the Petrarca's *Canzoniere*, whose success is enormous over the centuries. The *Canzoniere* is the model *par excellence* of vernacular poetry. The phenomenon of *Petrarchism* provoked a general process of imitation among most of the authors who followed Petrarca. The lyric poetry of Petrarca became a model so imitated that some scholars have not hesitated to define this process of imitation as plagiarism.

The presence of Petrarca in Florentine vernacular literature was a very advantageous condition in the realization of the translator. Many terms and phrases of *Canzoniere*, which appear in the list of definitions, are "imitated" in a myriad of subsequent authors and are therefore translated by the program into current Italian. Petrarca's influence can also be seen in opera librettos[1]. Since the birth of the opera in the sixteenth century and beyond, the influence of Petrarca can be felt. Thus, with a help of translator it is also possible to translate them into current Italian. There are several Internet sites containing opera librettos in digital form from every era, although most of those works are no longer performed.

The *Canzoniere* is therefore a fundamental text inserted in the translation program. Which edition of it to choose? There are in fact different editions of the Canzoniere in modern Italian: some editions have a more philological character, others are closer to the modern language. A "digitalized" version of the *Canzoniere* found on the web is the famous philological version by Gianfranco Contini. Although this edition is not exempt from some criticism [8], it shows, for example, that the copulative conjunctions are written as *et* ("and"). In other more recent editions this does not happen. Therefore, the author decided to use two different editions of *Canzoniere*: the one of Bettarini (closer to that of Contini) and another of Vecchi Galli, who even states, "the time seems right to take the distance from the "antiquarian" taste of the Contini's *vulgata*" [9]. In this way, the translator can be at ease even with slightly different texts.

Another work from which the definitions have been extrapolated is Boccaccio's *Decameron*. The success of the

Decameron at the European level was immediately very vast, and from the sixteenth century the book, thanks to Bembo, became, "the supreme model of prose" [10]. Some personal names and geographical names, which frequently can be found in the Decameron are also included in the translator's definitions list[2].

Vocabolario della Crusca (1686 edition) was chosen by the author as another text for the extrapolation of definitions to be included in of the translator. The choice is obligatory here: in addition to the terms used by the vernacular language the vocabulary includes examples of dozens of authors from different periods. Moreover, the numerous examples that this vocabulary offers, at least partially solve the problem of using words in different contexts[3]. The definitions deriving from the *Vocabolario della Crusca*, makes use of a corpus. In fact, it is the first vocabulary representing a large corpus of linguistic data (*i.e.* lemma in alphabetical order, definition and example) to which all subsequent vocabularies will refer [7].

Finally, in a more recent version of the author's program, new definitions from an edition of Petrarca's *Trionfi*, edited by Guido Bezzola, have been added. Bezzola himself underlines the importance of this text as a "demonstration of a new taste and culture towards the *Comedy*", and of how "Dante created the premises" [13].

The possibility of choosing other texts to insert new definitions is not excluded in the future, although it increases the risk to slow the translation process.

4. Didactic Use of the Translator

4.1. Use of the Translator for the Italian Language Students

One of the translator's main aims is to facilitate philological studies of the vernacular language, simplifying the understanding of such texts. The translator was already used for teaching purposes at the University of Vilnius, providing a group of Italian language students with a teaching exercise that included the original versions of texts in vernacular and those translated by the program. The aim was to observe if the level of understanding of the translated texts was higher than the original texts. The choice of the texts to be read by students was not accidental: the text with obsolete terms is more difficult to understand especially for foreign students. Therefore, considering the average level of preparation of the students, not too much difficult texts were chosen. Short texts in the vernacular of Petrarca and Boccaccio were distributed, followed by a test with twelve multiple-choice answers. The test required to choose the right meaning of some words or phrases, proposing the choice between three different meanings. Then, once the errors had been corrected, the original texts were compared and discussed with the version provided by the translator to

---

[1] "Rinuccini and his way of treating Petrarca later became a model for Alessandro Striggio's Favola d'Orfeo" [11].

[2] In the drafting of the *Vocabolario della Crusca* geographical names were omitted because "it seemed from the beginning that they no longer taught language" [12].
[3] "One of the main problems facing an automatic translation system is polysemy of words, first and foremost verbal polysemy" [2].

see if the level of understanding had increased. Many errors would have shown that the translator is certainly a useful tool to help students, or just readers, understand the vernacular. The results, obtained by a group of six students, are summarized in a graph (Figure 1).
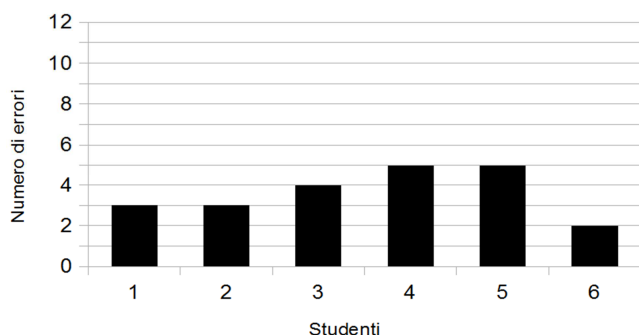

Figure 1. Student's mistakes in translation (average = 3,6).

The x-axis shows the number of students, the y-axis the number of mistakes they did in translation. The average error, although not very high, shows a marked degree of difficulty in understanding the vernacular language. The most frequent errors concern certain constructs such as *v'aggio* ("vi ho") or *tre anella* ("tre anelli"), or even words as *fiata* ("volta"). On the other hand, other words have not created difficulties: for example, the term *humanitade* ("umanità") has always been correctly understood, perhaps through the reminiscence of Latin. No student was able to run the test without errors. By proposing the translated version of the texts to the students, the level of understanding has improved, even if it is still about literature and a certain "training" is needed to understand what is read. The effectiveness of the translator has not been questioned by the students, who have readily understood the help of the computer in understanding the vernacular language.

4.2. Some Statistical Data

The Table 1 below shows a list of authors translated by the program. It contains the name of the author, the translated work and the number of substitutions made in the translated text. The authors have been chosen from different periods, even outside of Tuscany, and their works obtained in digital copy and in text format from some Internet sites. The number of substitutions shows that the translator is effective, although each translated text should be examined in detail to understand if and where the program makes mistakes or where it does not translate at all. However, the high number of replacements suggests that the translator's basic texts are a valuable support for the translation process. Moreover, from the ratio between the length of the texts and the number of substitutions it can be deduced which authors, after the scholars of the fourteenth century, use a more standardized vernacular language. A work that has a high number of substitutions in relation to its length could finally indicate that Florentine vernacular language used is more distant from the modern Italian.

Table 1. List of authors with number of substitutions.

Author	Work	Number of substitutions
F. Petrarca	Canzoniere	17901
G. Boccaccio	Decameron	36343
D. Alighieri	Commedia	17981
F. Sacchetti	Trecentonovelle	19234
N. Machiavelli	Il principe	2960
L. Ariosto	Orlando furioso	41946
L. Pulci	Morgante	29559
T. Tasso	Gerusalemme liberata	19471
A. Tassoni	La secchia rapita	8025
P. Metastasio	Achille in Sciro	979

The Table 1 shows the high number of substitutions in Ariosto's *Orlando Furioso*, "being a large part of the Furioso's vocabulary of petrarchist imprint" [16]. Ariosto, moreover, "shows the ways in which the petrarchist amorous phenomenology is unfolded" until it constitutes a parody of the love itself.

It is observed, that many other authors are also petrarchists. For example, Pulci: "the number of times in which Petrarca is recalled in the Morgante is huge" [16]. The same can be said of Tasso, who "identifies with greater precision and lucidity the serious, thematic and stylistic vein of Petrarca's production" [17].

The choice of the translator's basic texts is therefore well founded at the same time remembering that in the future the list of definitions may be expanded by adding new definitions.

5. Results and Discussion

One of the advantages of translation using the program is that only terms or phrases which are considered difficult to understand or are no longer in use should be included in the definitions. However, to simplify the text under examination, many terms not common in the modern Italian language have also been included in the list of definitions. At the same time the author faced several problems.

One of the biggest problems the author encountered was the presence of *truncation*. The program does not recognize whether a term is truncated or not, so the author preferred, where possible, to include in the definitions also the truncated version of words. Often the infinitive form of verbs is truncated, for example, instead of "acconciare" it is used "acconciar", "addivenir" instead of "addivenire" etc. All the verbs always have the version with truncation in their definitions. However, many other words can be truncated and in these cases the translator can do very little, except when such definitions are taken from basic texts.

Elisions are another major problem: if possible, they have been removed from the translated text to simplify the language.

Another problem derives from the use of the article *il* or its equivalent *lo* in the vernacular literature. For example, if the result of translation is *lo*, this can be confused with the pronoun *lo* in the modern Italian, making it impossible for the

translator to understand what the function of the particle is. One specific example can be found in a novel of the Decameron, the fourth novel of the eighth day: Boccaccio writes "il proposto" (in modern Italian "prevosto"), but also "lo proposto [4]". Nouns which use the article *lo* in the *Decameron* are placed as definitions together with the article itself and then again inserted in the definitions without article. In this way, they are translated with the current article *il* into Italian. But the problem occurs when the translator encounters nouns not present in the basic texts. This aspect of the vernacular literature is quite frequent in Boccaccio, less so in Petrarca, where the use of articles is closer to modern Italian. One of such examples, among the infinite possible ones. can be taken from *Canzoniere*'s canzone XXII, verse 22:

et non mi stancha primo sonno od alba:
ché, bench'i' sia mortal corpo di terra,
lo mio fermo desir vien da le stelle.
The program, with ten replacements, translates as follows:
e non mi stanca primo sonno od alba:
perché, benché io sia mortale corpo di terra,
il mio fermo desiderio viene dalle stelle.

In the author's view, the translation or, better to call it, the "conversion" is acceptable.

A more complex example from the sonnet LXXIX, verse 1:
S'al principio risponde il fine e 'l mezzo
del quartodecimo anno ch'io sospiro,
piú non mi po' scampar l'aura né 'l rezzo,
sí crescer sento 'l mio ardente desiro.
The program translates with nine replacements:
S'al principio risponde il fine e il mezzo
del quattordicesimo anno che io sospiro,
più non mi può scampare l'aria né l'ombra,
sí crescer sento il mio ardente desiderio.

The translation is quite accurate, although obviously lacking the sense of allusion to Laura (*scampar l'aura*) and her "shadow which protects the poet" (*rezzo*) [15]..Another difficulty for the translator is the presence of verbs which are no longer used in the modern Italian. The number of definitions would increase in an unmanageable way if they include all the verb tenses. So, the author choses to rely on a compromise: in addition to infinitive, the third person singular and plural of the present simple are included as well as past participles and possibly other verb tenses, if found in the examples of the *Vocabolario della Crusca*. In this way, there is a good chance that the program will be able to translate a substantial part of the verbs.

Unfortunately, the translator contains one unresolvable problem: the combination in a word of verb and pronoun, for example "recargli", where "recare" is a verb and "gli" a pronoun. If the translator finds a verb used only in the vernacular literature to which a pronoun is attached, it can translate such word only if such combination is included in the definitions. This problem becomes deeper due to the high

number of combinations and it can be partly solved by inserting the examples from the *Vocabolario della Crusca*.

Although roughness described above exists, the translator gives the best results for translations of those authors who imitate the models of Florentine literature, especially Petrarca. Sometimes they are minor authors, but it should be remembered that Petrarchism is a phenomenon that also affected famous authors. So, it is not surprising that, among the great authors, the translator makes thousands of substitutions and this will be demonstrated later.

To show the effectiveness of the translator, example chosen at random from Pulci (*Morgante*, Cantare I) is provided:
L'abate si chiamava Chiaramonte:
era del sangue disceso d'Angrante.
Di sopra alla badia v'era un gran monte
dove abitava alcun fero gigante,
de' quali uno avea nome Passamonte,
l'altro Alabastro, e 'l terzo era Morgante:
con certe frombe gittavan da alto,
ed ogni dì facevan qualche assalto.
The translator makes nine replacements:
L'abbate si chiamava Chiaramonte:
era del sangue disceso d'Angrante.
Di sopra alla abbazia vi era un gran monte
dove abitava un qualche crudele gigante,
dei quali uno aveva nome Passamonte,
l'altro Alabastro, e il terzo era Morgante:
con certe frombe gittavan da alto,
ed ogni dì facevano qualche assalto.

The translation might be arguable but is quite accurate. The truncation of *gittare* ("gittavan") is not identified and has not been changed. The noun *fromba* was also not changed but it appears in some modern dictionaries, although declared of "ancient" use. The example was chosen to test the translator on a work which was not used for extrapolation of definitions. However, by adding new definitions from other works in Florentine vernacular language, the quality of translation will certainly improve.

6. Conclusions

The article demonstrates that the translator, although imperfect like all automatic translators, works with an acceptable degree of precision. It is believed that it may be useful to scholars of Italian literature, philologists or simply to readers of works in the vernacular. It can also be useful to Italian language students to boost their interest in a literature written in the vernacular, which in any case needs to be "assisted" by footnotes to understand the text. Often a text in the vernacular needs footnotes that explain the meaning of some terms or phrases. However, the vernacular literature uses a language with a high symbolic and cultural content. In this case the translator cannot assist, and it is necessary to rely on philological studies. For this reason, the automatic translator should be used only as a means of support for understanding, without expecting a complete and exhaustive understanding of Florentine vernacular texts.

---

4 The rule seems to be there: "when the word ends with *R*, if the word that follows must have male article, one puts *LO*" [14]. Therefore, since in the novella is *messer lo proposto* (and *messer* ends with *r*), *lo* is used as an article.

References

[1] M. W. Madsen. *The Limits of Machine Translation*. Copenhagen: Faculty of Humanities, University of Copenhagen, 2009.

[2] D. Słapek. *Lessicografia computazionale e traduzione automatica*. Firenze: Cesati, 2016.

[3] T. Poibeau. *Machine Translation*. Cambridge, Massachusetts: MIT Press, 2017.

[4] W. J. Hutchins, "Machine translation over fifty years", Histoire, Epistemologie, Langage, vol. XXII, 1, pp. 7-31, 2001.

[5] P. Koehn. *Neural Machine Translation*. Baltimore, Maryland: Johns Hopkins University, 2015.

[6] F. Tamburini, "La linguistica computazionale: un crogiolo di esperienze multidisciplinari. URL http://www.griseldaonline.it/informatica/la-linguistica-computazionale-tamburini.html (Accessed: 3/12/2018).

[7] E. Cresti and A. Panunzi. *Introduzione ai corpora dell'italiano*. Bologna: Il Mulino, 2013.

[8] SDT - *Sentimento del tempo. Petrarchismo e antipetrarchismo nella lirica del Novecento italiano*. Firenze: Leo S. Olschki, 2005.

[9] PVG – F. Petrarca. *Canzoniere*. P. Vecchi Galli (ed.), Milano: Rizzoli, 2012.

[10] VB – G. Boccaccio. *Decameron*, V. Branca (ed.). Torino: Einaudi, 1980.

[11] PMP - *Il Petrarchismo. Un modello di poesia per l'Europa*. vol. 1. Roma: Bulzoni, 2006.

[12] VDC – *Vocabolario degli Accademici della Crusca*, Venezia: Per Combi, e La Noù, 1686.

[13] GB – F. Petrarca. *Trionfi*. G. Bezzola (ed.), Milano: Rizzoli, 1984.

[14] T. G. Da Pofi. *La grammatica volgare trovata ne le opere di Dante, di Francesco Petrarca, di Giovanni Boccaccio, di Cino da Pistoia, di Guittone d'Arezzo*. Napoli: Giovanni Sultzbach, 1589.

[15] RB – F. Petrarca. *Canzoniere*, R. Bettarini (ed.). Torino: Einaudi, 2005.

[16] M. C. Cabani. *Fra omaggio e parodia. Petrarca e petrarchismo nel «Furioso»*. Pisa: Nistri – Lischi, 1990.

[17] TDP – *I territori del petrarchismo. Frontiere e sconfinamenti*. Roma: Bulzoni, 2005.