



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

Machine Translation and Artificial Intelligence Applied to the Translation of Puns and Wordplay in “Alice’s Adventures in Wonderland”

Relatore
Prof. Sabina Fata

Laureando
Federico Balestra
n° matr.2096909 / LMLCC

Anno Accademico 2023 / 2024

Ringraziamenti

Questa tesi di laurea non è solo il frutto del mio lavoro. È stata scritta anche grazie al supporto all'aiuto di tutte le persone che mi sono state vicino in questi anni.

Ringrazio innanzitutto la mia famiglia, il loro aiuto e la loro vicinanza sono state fondamentali durante il mio percorso di laurea e in questi ventiquattro anni di vita, rendendomi la persona che sono oggi.

Ringrazio tutti i miei amici, anche loro sono stati un tassello fondamentale del mio percorso di vita, sicuramente senza di loro non sarei quello che sono oggi.

Table of Contents

Introduction...	3
1. Literature Review...	7
1.1 Machine Translation...	7
1.2 Evaluating Machine Translation Outputs...	14
1.3 Artificial Intelligence...	22
1.4 Machine Translation and Artificial Intelligence Applied to Translation...	27
1.5 Humor, Puns and their Translation...	34
2. Preliminary Analysis...	45
2.1 Object of the research...	45
3. Analysis...	53
3.1 Intento Modern MT...	54
3.2 RWS Language Weaver...	61
3.3 MT Evaluation...	69
3.4 ChatGPT 3.5...	74
3.5 Microsoft Edge Copilot...	81
3.6 AI Evaluation...	87
3.7 MT and AI Comparison...	88
3.8 AI Prompting...	90
3.9 AI Second Evaluation...	119
3.9.1 ChatGPT 3.5...	120
3.9.2 Microsoft Edge Copilot...	125
3.10 Machine Translation Training...	132
3.10.1 RWS Language Weaver...	133
3.10.2 Intento Modern MT...	144
3.11 MT Second Evaluation...	155
3.12 MT and AI second comparison...	166
4. Conclusions...	169
Bibliography...	179
Summary...	187

Abstract

Balestra Federico

Machine Translation and Artificial Intelligence Applied to the Translation of Puns and Wordplay in “Alice’s Adventures in Wonderland”

Lingue Moderne per la Comunicazione e la Cooperazione Internazionale

Sabina Fata

Fiona Claire Dalziel

The present work focused on the use of Machine Translation and Artificial Intelligence applied to the translation of puns, wordplay and other humorous content in literary texts.

These textual instances tend to pose several problems to translators; however, technological tools such as Machine Translation and Artificial Intelligence have been proved to be helpful to translators when it comes to tackle these elements. Even though these tools can be helpful, their performance could be improved. The present study tried to identify methods directly applicable by translators in order to obtain better translations when using the NMT providers ModernMT and RWS Language Weaver, and when using the general Artificial Intelligence interfaces ChatGPT 3.5 and Microsoft Edge Copilot.

The present dissertation is introduced by a detailed literature review focusing in the history and development of Machine Translation and Artificial Intelligence. Moreover, puns and wordplay are addressed in order to provide a definition of this linguistic phenomenon and in order to present to the readers some possible translation strategies that can be adopted when tackling these textual instance. The literature review also focuses on studies already conducted on MT and AI applied to the translation of puns and wordplay (an humorous language in general). Then, the actual study is described.

In order to carry ou the research, five instances of humorous language were extracted from the novel “Alice’s Adventures in Wonderland”. These textual instances were translated using the above-mentioned tool; after that, the translations obtained were evaluated using both a manual evaluation system, and two automatic evaluation metrics (BLEU and TER). Once the scores were assigned, the same instances were translated a second time using the same tools. This time, improvement strategies were applied: NMT providers were trained for the translation of puns and wordplay using TMs specifically created for this purpose, and AI interfaces were fed prompts suitable for the type of translation result needed. After the second translation, the results were evaluated a second time and compared to the results obtained after the first translation, in order to control whether or not the improvement strategies worked.

The results of the study showed that while prompting engineering proved to be useful, the same cannot be said for the training of NMT providers. The instances of puns and wordplay were not translated in a satisfactory way even after the training. However, some improvements as far as syntax was concerned could be observed, especially in relation to RWS Language Weaver.

Generative AI and Neural Machine Translation could sure be of great help to translators when it comes to the translation of puns and wordplay. However, some work on the part of the translator is required. For example the translator needs to feed the AI provider with suitable prompts in order to obtain a satisfactory result.

Introduction

Literary texts tend to pose many problems to translators. One of the most difficult features to translate in a literary text is humor, especially when it is present in the form of wordplay and, specifically, of puns, because “unlike most kinds of verbally expressed humor, they use the specific features of a particular language” (Low 2011, 59) and they “rely not only on surface-level features but also a relatively sophisticated understanding of lexical semantics and (usually) the complex pragmatic phenomenon of humor” (Kolb & Miller 2022, 1). Translating instances of humorous language requires the translator to have not only a deep understanding of the Source Language, but also of the cultural background of the text and its author. Translating humor can be difficult, but translators are not left alone in their job. In the modern-day era, they can rely on the help of really sophisticated technological tools.

As O’Brien points out in the article “Translation as Human-Computer Interaction” (O’Brien 2012), translators rely more and more on technological tools to help them perform their job, such as CAT tools and Machine Translation. Since “modern translation depends heavily on technological aids” (Ermakova et al. 2022, 355) it would be interesting to investigate whether or not tools such as AI and MT could be of any help when it comes to difficult tasks, such as the translation of humorous language. A similar study has been carried out by the participants of the JOKER tracks at CLEF 2022 and 2023. In the 2022 edition, a team made of professional translators and computer scientists tried to evaluate how translators and computer-based models understand humor (Dhanani, Rafi and Tahir 2022, 1). The 2023 edition carried out a similar research, focusing on pun detection, location, interpretation and translation using the programming language Python combined with AI and machine learning methods (Popova and Dadić 2023, 1). Even though this kind of software reached a high level of complexity, as the CLEF 2023 demonstrated they still require some improvements when it comes to performing tasks such as pun translation (Popova and Dadić 2023, 20). Shifting our focus on Machine Translation, Miller (2019, 60) points out that this tool struggle to produce high-quality results when it is applied to creative language and that the “unsuitability of the prevailing statistical MT paradigm for the translation of

humor” has been already recognized (ibid. 2019, 61). Therefore, these sophisticated tools have demonstrated to be of great help to translators, but they are nevertheless limited when they are applied to creative or humorous language. While previous studies demonstrated the limitations of these tools, this paper looks for a way to overcome these limitations, investigating the translation of humorous language in literary texts through tools such as Neural Machine Translation and generative Artificial Intelligence. The main purpose of the research is to identify strategies that could improve the productivity of those tools when used to translate humorous language, focusing on wordplay and puns.

To carry out the research for this paper, two types of tools will be used: Neural Machine Translation (also known with the acronym NMT) and generative Artificial Intelligence. The history and the peculiarities of these two tools will be explored in the literature review. The research presented in the paper will therefore apply NMT and generative AI to instances of humorous language in literary texts: the analysis will be mainly conducted on wordplay, with a specific focus on puns.

In this work different examples of puns and wordplay taken from the novel “Alice’s Adventures in Wonderland” will be translated using both Neural Machine Translation and generative Artificial Intelligence. When it comes to NMT, the tools ModernMT and RWS Language Weaver will be used. The tools have been chosen because they are really advanced solutions for professional translators, to the point that they are employed by different professionals and for different purposes. As reported on the RWS Language Weaver website, Language Weaver is used not only by translators, but also by government organizations; the same software is employed to serve Intelligence and law enforcement purposes (<https://www.rws.com/language-weaver/government-solutions/>). ModernMT, on the other hand, is reportedly capable of adapting to the translator’s data: as reported on the ModernMT website, this machine translation tool has been developed having in mind a human-machine interaction approach (<https://www.modernmt.com/>). Therefore, these tools are highly reliable and sophisticated, thus being a logical choice for this kind of research. In order to perform the translation through generative AI tools, ChatGPT 3.5 and Microsoft Edge Copilot will be adopted. The results will be compared to one another and to a “golden standard” translation of the book provided by a human translator. After analyzing the instances of humorous

language, this work will be focused on finding strategies to improve the productivity of Neural Machine Translation and generative AI when applied to humorous content, in order to make them helpful for translators.

The research will adopt both a qualitative and quantitative method. The qualitative method will be employed for analyzing and evaluating the outputs provided by generative AI and NMT, trying to identify the main errors occurred in the translation of the instances of humorous language analyzed. The quantitative method will be employed by choosing a certain quantity of instances of humorous language in order to check how many of them can be translated in an acceptable way by NMT and generative AI. After that, the research will be focused on finding strategies to improve generative Artificial Intelligence and Neural Machine Translation; however, this part will be developed only after the preliminary analysis of the outputs of the tools employed.

This paper will be articulated as follows:

- A literature review will be provided, focusing on Machine Translation and Artificial Intelligence. A brief history of these tools will be delineated, as well as their main peculiarities. Then the literature review will take into consideration wordplay and puns
- After that, the quantitative method will be applied. A certain number of wordplay and puns will be chosen to be translated through generative AI and NMT. In order to choose what to translate, the instances of humorous language that could be considered more complex to translate will be picked up, so to pose a challenge to the machine. Since the tools that will be used are highly developed and sophisticated, chances are high that they might be able to translate quite well easier examples of puns and wordplay. In the literature review some space will be devoted to the description of puns and their translation techniques, in order to make more clear what kind of puns and wordplay could be considered easier to translate
- The quality of the translation will be evaluated, applying the qualitative method. The evaluation phase will point out the main errors occurred during the translation phase. This phase is really important, since the results of the pre-training evaluation will be compared to the results of the evaluation carried out after the application of the strategies developed to possibly improve the translation results

- After this preliminary analysis, the second part of the thesis will be devoted to finding strategies to improve the productivity of the employed tools. The aim of these strategies would be helping the translation tools based on NMT and generative AI to overcome the errors previously identified by the analysis
- The final part of the research will be interested by a second evaluation of the translation outputs. After finding methods that could possibly improve the productivity of the aforementioned tools and after applying them, the instances of humorous language previously analyzed will be re-translated through the same tools. The hypothesis to be confirmed is that the strategies identified actually improved the productivity of these tools. The re-translation will be followed by another evaluation of the outputs, in order to check whether or not the methods identified improved the productivity of generative AI and Neural MT

The paper will then start with a literature review before focusing on the actual research, which will make the greater part of the present paper. After that, the discourse will move on to the discussion of the results, in order to check whether or not the methods identified actually improved the tools picked up as far as the translation of humorous language is concerned.

Literature Review

1.1

MACHINE TRANSLATION

The history and development of machine translation is a topic that has been explored in detail by several different authors, such as Stein (“Machine translation: Past, present and future”, in “Language technologies for a multilingual Europe”, 2018), Hutchins (“Machine translation: a concise history”, 2003; and “Reflections on the history and present state of machine translation”, 1995), Sloculn (“A survey of machine translation: its history, current status, and future prospects”, 1985), Wang et al, (“Progress in Machine Translation”, 2022), Cheragui (“Theoretical Overview of Machine translation”, 2012). These scholars focused not only on the history itself, but they also highlighted the developments of this tool and provided an in-depth description of its evolution through history. Even though Machine Translation tools are now part of our everyday life, their history is not as straightforward as one might think. MT has been considered “one of the most challenging tasks in the field of natural language processing” (Wang et al. 2022, 143) since 1947, when for the first time ever the concept of Machine Translation was formulated. Even though the actual research on machine translation started in the 1940s, the idea of creating a machine that could translate texts from a language into another had already been developed a decade before. As Hutchins points out in “Machine translation: a concise history” (Hutchins 2003), in 1933 two patents were issued by the French Georges Artsrouni and the Russian Petr Troyanskii. The two pioneers in the field developed independently the idea of a mechanical dictionary that could be used to translate from a language into another. Interest in this field of research raised in the 1950s, when MT was initially developed for military purposes (Wang et. al 2002, 143). As Wang et al. describe in their essay “Progress in Machine Translation”, the first successful attempt to use MT took place in 1954, when the IBM-701 computer was used to perform a machine translation from Russian into English (ibid.). In the period from the 1950s to the 1960s the three basic approaches

to MT were developed, as Hutchins points out (Hutchins 2003, 2): direct translation, interlingua model and transfer-based approach. As described by Cheragui (2012, 163-164), the direct approach consists in translation from the ST to TT after taking a string of words and removing all the morphological inflections in order to obtain the base form of the words; after that, a bilingual dictionary is used to check the correspondence of the words in the two languages. The Interlingua model is mainly used in multilingual systems. This approach is divided into two stages: Analysis and Generation. After the analysis of the ST, the content of the text is represented using an intermediate language, called Interlingua. After that, the Interlingua is used to generate the sentences for the TT. The transfer-based approach uses representations of the ST to produce the TT. The first phase of this approach sees the conversion of the ST into a representation that maintains several characteristics (from syntactic to semantic) of the text to translate. After that, the representation is converted into an equivalent representation for the TT, from which the translated text is generated. After a decade of studies the dream of creating an MT system was crushed in the 1960s. In 1966 the ALPAC (Automatic Language Processing Advisory Committee) report condemned both the field of research and its workers (Sloculn 1985, 1). According to the committee, machine translation was an unfruitful field of research, that would not have produced any concrete or satisfactory result. What the researchers wanted to obtain were MT systems capable of producing translations “at least as good as those made by human translators” (Hutchins 1995, 2); of course, such results were not attainable in the immediate future, In spite of the report, researches went on in the 1970s. In the 1970s, several different MT systems were developed. For example, Peter Toma created the System SYSTRAN1, which was engineered in order to provide Russian-English translation. Another system which was created in this decade of research was the WEATHER, developed in order to provide English-French machine translation of weather forecasts (Cheragui 2012, 162). The very existence of the WEATHER systems proves that since the beginning the research on MT was carried out focusing on many different topics and purposes, somehow foreshadowing the many fields of application of machine translation in the contemporary world. It was only in the 1980s that interest in researching MT raised again; this time, studies were funded not only by governments, but also by private enterprises (Sloculn 1985, 1). If the decades from the 1960s onwards were characterized by approaches oriented

to the syntactic characteristics of the language, in the 1980s more attention was paid to the semantics of the text: in this period, the interlingua method gained popularity (Hutchins 1995, 7-8). From 1988 researchers started to focus on a different method: the statistic-based approach (SMT). This approach will be further discussed later in this chapter. In the field of MT, the 2000s represent a watershed. Until the 1990s, the rule-based approach to MT (RBMT) was dominant, while the corpus-based approach stayed in the background. Rule-based methods use “bilingual dictionaries and manually written rules” (Wang et al. 2022, 143) to translate from the source language to the target language. As explained by Cheragui (Cheragui 2012) these rules are used for many different purposes, such as lexical transfer, morphology, syntactic analysis or syntactic generation. The translation process is carried out by RBMT in two stages: the first stage consists in the morphological, syntactical and semantical analysis of the input text; the second stage consists in the generation of the TT by “structural conversion based on internal structures”. The system required the use of dictionaries and grammar created by linguists, as well as written rules: all of this made RBMT quite a time consuming process. Moreover, RBMT models presented some problems, such as the difficulty in writing rules and the struggle to maintain and transfer them from one domain or language to the other (Wang et al. 2002). From the 2000s, the already mentioned corpus-based methods took over thanks to vast availability of corpora, especially of bilingual ones. As Wang et al. describe (2022, 143) there are three different corpus-based methods: example-based machine translation (EBMT), statistical machine translation (SMT) and neural machine translation (NMT). EBMT was first proposed in the 1980s; this method tends to provide good results only when similar sentence pairs are present in the corpora adopted. Unfortunately, corpora usually cannot contain a vast number of different linguistic phenomena; for this reason, EBMT is not always successful, and it is usually employed in computer-based translation systems (ibid., 144). SMT was first proposed in the 1990s; following this method, “machines automatically learn translation knowledge from a large amount of data instead of relying on human experts to write rules” (ibid.). However, SMT was not immediately adopted since the dominant model still was RBMT. As described in Chapter 2 of the book “Language technologies for a multilingual Europe” (Stein 2018, 5-17) SMT does not require any set of rules to work, but it relies on large parallel corpora. The

machine estimates the probability that a specific sentence might translate correctly a sentence from the ST using the Bayes' Theorem

$$Pr(e|f) = \frac{Pr(e) * Pr(f|e)}{Pr(f)}$$

This theorem is used to calculate the probability that one of the possible sentences e could be the possible translation of the sentence f ; henceforth, $p(e|f)$ refers to the “probability that e is a valid translation of f ” (Stein 2018, 11). SMT therefore works with translation models that consist in bilingual aligned corpora that represent the possible translations between two different languages. One of the main differences from the original ruled-based approach is that the systems now abandoned word correlations, focusing on phrase correlations thanks to the large amounts of data provided by the corpora. Looking deeper into the field of SMT, it is possible to notice that there are two main models of SMT: the Word-based SMT and the Phrase-based SMT. Word-based SMT analyses the data taking into account only lexical units; therefore, a word in the TT should correspond to a word in the target language (Stein 2018, 12). Obviously, this becomes a problem when the machine has to deal with multi-word expressions (MWE), that is to say those words that must be translated with more than a single lexical unit. Phrase-based SMT tries to find a solution to this problem by having the system focusing not on single lexical items, but on whole phrases. As pointed out by Stein in the already mentioned chapter, SMT has many different advantages, such as the possibility to create systems that could work well without knowing the source or target language. Since this system is based on corpora, there is no need to create specific dictionaries, or to write ad-hoc rules. However, SMT systems are likely to produce errors and to perform poor translations. If compared to RBMT, it is evident that while SMT performs better when it comes to word choice and disambiguation, rule-based approaches have better results when it comes to word order, syntax and coherence. SMT works well when adequate corpora are created and, as pointed out by Stein, “best translation are obtained when the SMT is created, trained and used for a special domain” (Stein 2018, 14). Statistical machine translation became popular between 1999 and 2003, when researchers proposed phrase-based SMT methods, improving the translation

quality and opening to the possibility of releasing open source systems (Wang et al. 2014, 144). Starting with Google, from 2006 onwards several companies launched translation services based on phrase-based SMT methods (ibid.). Another corpora-based model, as already mentioned above, is represented by the example-based machine translation (EBMT). This approach uses the data provided by the corpora to make the machine find or recall examples from the corpora. The translation process consist of “extracting and selecting equivalent phrases or word groups from a databank of parallel bilingual texts” (Hutchins 2003, 12), with the concept of “translation by analogy” that underlines the whole translation process (Okpor, 2014, 163). Translation by analogy is a translation process that has been used by Lepage and Denoual to develop an EBMT system. According to the researchers, a proportional analogy can be described as follows: $A : B :: C : D$, meaning “*A* is to *B* as *C* is to *D*”. This kind of analogy takes necessarily four arguments, and produces a result that can be either true or false. Therefore, if we apply this method to sentences in the translation process, we obtain an analogy according to which “some pieces of the sentences commute with other pieces” (Lepage and Denoual 2005, 253). This kind of analogy can be used to compare not only different sentences, but also specific syntactic elements inside the said sentences, such as inflectional morphemes. EBMT, therefore, uses the analogy system in order to compare the linguistic data stored inside the corpora. Machines that work thanks to this method are trained using example translation that are compared through the analogy system: “new sentences are translated based on their similarity with available examples” (Turcato and Popowich 2003, 4). In this system, similarity is to be intended as a semantic proximity of two words in a thesaurus (Somers 1999, 117). As it is explained in the article by Somers “Example-based Machine Translation” (Somers 1999, 113-157) all the approaches that can be listed under the name of Example-based Machine Translation have in common the use of corpora or databases of already translated examples to train the system, and the use of thesauri to retrieve word similarities. The translation by analogy method is, according to Makoto Nagao, the method that resembles the most the translation process when performed by a human translator. According to the researcher, human translation is not limited to linguistic analysis of the text, but it is articulated in three different passages: decomposition of the input sentence into smaller units; translation of the phrases obtained into corresponding phrases of the other language; finally, composition of

the translated fragments following the rules of the TL so to obtain a sentence. The translation of each phrase is performed by the analogy principle (Nagao 1984, 3). To reflect this mechanism, EBMt follows three different steps to translate: analysis of the sentence, corresponding to the identification of smaller units in the ST; retrieval of examples, followed by the calculation of the semantic distance in the thesaurus and by the transfer, corresponding to the translation of the single phrases; finally, the machine performs the conventional generation of the TT, corresponding to the composition of the translated fragments (Sumita, Iida, and Kohyama 1990, 205). The main reason behind the development of such system was allowing translators to use systems that could perform correct translation of languages that shared few similarities. For this reason, many of the first tests and experiments with EBMt were carried out on the English-Japanese language pair. Given the great difference between the two language, a Rule-based system would have been of little help, since a comparison between the two language structure would be unfruitful. In such cases, the translation must focus on different blocks of words (ibid., 4). Of course, this approach presents some problems. If RBMT required to manually write complex sets of rules, this approach requires to find suitable aligned corpora. Moreover, it is pivotal to identify which blocks of words the system should compare to obtain a proper translation. The problem of the granularity of examples is well described by Somers, who suggests that the best “grain-size” should be the sentence, though this might present some problems, given the length of a sentence. It has been proved that the longer the example, the more difficult the match, but at the same time, the shorter the example, the greater the probabilities of an ambiguity (Somers 1999, 119). Other problems listed by Somers are the numbers of examples and the suitability of examples. According to many studies, the higher the number of examples, the higher the probability of a well-functioning system, even though it must be taken into account that the number of examples that can be added is limited, since “there is some limit after which further examples do not improve the quality” (ibid.). Another issue is related to the suitability of examples: of course, the system requires suitable examples to work correctly, since the wrong examples could produce ambiguities. The latest innovation in the field of Statistical MT is Neural Machine Translation. Research on this model started in the last decades of the 20th century, but the limitations imposed by the technology of the time prevented researches to carry out proper studies. In the first years of the 21st century, neural

models started to be integrated into Statistical MT models, marking the resurrection of the studies on this topic (Koehn 2017, 5). Researches on the development of neural-only models started in the years between 2013 and 2015, with the first models being released to the public. The first large-scale NMT system was released by Baidu in 2015, followed by Google in 2016. Similarly to EBMT, NMT tries to replicate the steps that a human translator would follow: it “reads” the sentence in the source text, and then it generates a translation which is based on the system’s understanding of the sentence (Wang et al. 2022, 144). NMT is now the state-of-the-art translation model thanks to its ability to work independently from human intervention: it does not need any written rule, nor any hand-compiled vocabulary, as well as it does not need any modification to the database of corpora. This model can learn through what is universally known as “deep learning”, which is the ability to learn directly and independently from the data the system is fed with. NMT models are really sophisticated and they take SMT to a step further. Statistical methods are based on the so-called “linear models”. In linear models, there is a set of given features that represent the potential translation of a sentence; then, these features are weighted using a specific parameter. What is obtained after weighing the values, is the output value: an overall score that represents the translation. As pointed out by Koehn in his chapter on Neural Machine Translation (Koehn 2017), neural networks modify this model adding multiple layers. The machine does not compute the output value directly from the input values, but the computation passes through a hidden layer (or hidden state). The machine first computes the input nodes and their respective weights to produce hidden nodes; then, hidden nodes are compute with their respective weights to produce the output value. Hidden nodes have been developed to reduce or eliminate the engineering interventions necessary to discover useful patterns: by training the hidden nodes, the machine becomes able to discover autonomously these patterns. As explained in the already mentioned article “Progress in Machine Translation” (Wang et al. 2022, 144) NMT models typically contain two components: an encoder network and a decoder network. The encoder network maps the sentence in the source language into a real-valued vector; from this vector, the decoder network produces the translation. The meaning of the words fed to the machine is first encoded by the hidden layer in the form of word embeddings. In word embeddings, each content word is represented in the form of vectors, and then projected in the form of such vectors in order to capture semantic

relationships between words. By using vectors, the machine can understand better the language. The hidden layer obtained with such process is then used in the decoder phase to produce the translation. The process then mimics the process of human translation: the neural machine translation engine reads the sentence and then it translates the source sentence into a target sentence basing on its understanding of the original sequence of words. NMT has established itself as the dominant model since it does not need any rule designed by humans, but it is “an end-to-end framework that directly learns semantic representation and translation knowledge from the training corpora” (ibid.). The Machine Translation models that will be used to carry out the present study will be *ModernMT* and *RWS Language Weaver*, two MT models that rely on the Neural Machine Translation Approach.

1.2

EVALUATING MACHINE TRANSLATION OUTPUTS

A really important step of this study will be the evaluation of the translations obtained using *ModernMT* and *Language Weaver*. Studies on Machine Translation performance and on its evaluation have been conducted for a long time, with some of the earliest studies carried out already in the 1990s. In 1993, one of these studies was funded by DARPA, as reported in the paper by White, O’Connell and Carlson: “Evaluation of Machine Translation” (White, O’Connell, and Carlson 1993). The main purpose of such studies is to reduce the amount on subjectivity in the evaluation of MT outputs, in order to make the process as much standardized and objective as possible. In the following decades, more and more researchers became interested in the topic, developing many different MT evaluation systems. Some of these studies have been carried out by Joseph Turian, Luke Shen and Dan Melamed (Turian, Shen, and Melamed 2003), Maja Popovic (Popovic 2012), Attila Görög (Görög A 2014, 443-454; Görög 2014), and Irene Rivera-Trigueros (Rivera-Trigueros 2022, 593-619). Researchers feel the need to develop different MT evaluation systems because of the quality level of the translations performed by these products, that is “generally inferior to that reached by professional human translators” (Rivera-Trigueros 2022, 594); even though MT usually performs worse than human translators, it goes without saying that globalization generated the need

for a huge number of translations, and such a number of requests cannot be satisfied only by the efforts of human translators. Machine Translation seems to be some sort of necessary evil, but at the same time it would be wrong to just accept the translation provided by the machine, without even controlling their quality. Hence, the need for standardized measures to control the quality of Machine Translation outputs. Originally, the verification process consisted of a comparison between the machine translated text and a reference text, which is considered to be a high-quality translation. The evaluation system assigned a score to the translation according to its similarity to the reference text. As pointed out in the paper “Evaluation of Machine Translation and its Evaluation” (Turian, Shen, and Melamed 2003), one of the most popular evaluation models, BLEU, uses the same idea. The reason why this method is so popular is that it “correlates very highly with human judgments” (Turian, Shen, and Melamed 2003, 1). However, other measures have been developed over the years, such as WER (Word Error Rate). This system categorizes the possible translation errors into different classes, such as Inflectional Error Rate, Reordering Error Rate, Missing Word Error Rate, Extra Word Error Rate, and Lexical Error Rate. This system was at the center of the study conducted by Marja Popovic, which investigated, among other things, the general behavior of each error category (Popovic 2012, 73). A comprehensive and detailed description of the evaluation methods will be discussed later in this chapter, when the evaluation metrics adopted by the present study will be presented. Notably, the TAUS society developed a very peculiar Quality Evaluation (QE) framework in 2011, that is discussed in two paper by Attila Görög. The concept of quality that stimulated the project the idea of quality as customer satisfaction. Good quality of a translation is not determined by lexical, semantical or syntactical parameters, but by customer satisfaction. According to these papers, the actual evaluation systems might not measure the right things; henceforth, TAUS decided to create the Dynamic Quality Framework, a “rich knowledge base on Quality Evaluation with best practices, reports, templates and a number of tools to evaluate translations made both by human translators and MT engines.” (Görög A 2014, 445). The criteria on which DQF is based are Utility (the importance of the translated content), Time (the amount of time the translation process took) and Sentiment (the impact of the translation on the image of the brand). Basing on this criteria and on the content of the translation, the program suggests the best evaluation systems for the specific

translation. By doing this, users might be sure that the evaluation of the product is not based exclusively on the translation itself, but that it also considers the goal and the intended audience of the translation. In order to help professionals producing good-quality translations, TAUS also created an add-in for Trados Studio: Qualitativity. This tool runs in the background during the translation process, collecting all sort of information related to quality and productivity (hence, the name Qualitativity). More information on the add-in can be found the Trados Wiki at the following address: <https://community.rws.com/product-groups/trados-portfolio/rws-appstore/w/wiki/2251/qualitativity>.

In order to evaluate the translations provided by MT, this study will use the suite provided by Intento, which groups different evaluation metrics: BLEU, TER, hLEPOR, COMET, chrF++, and BERTScore. The next section will be dedicated, as anticipated, to the description of the different systems.

BLEU (Bilingual Evaluation Understudy) – The BLEU system has been described in the papers “BLEU: a Method for Automatic Evaluation of Machine Translation” (Papineni et al. 2002, 311-318) and “A Call for Clarity in Reporting BLEU Scores” (Post 2018). The main idea behind this system is that of developing a quality metric that resembles human judgment. Therefore, the metric has been developed so that it can measure the closeness between the machine-translated text and a reference human translation. BLEU has rapidly become the dominant metric in the field of MT research thanks to its language independence, ease of computation and closeness to human judgments (Post 2018, 1). The system, as explained by Papineni et al. needs “a numerical translation closeness metric” and “a corpus of good quality human reference translations” (Papineni et al. 2002, 311). The main idea behind this specific metric is, as already said, to compare a certain translated text to a translation which is considered to be good (usually, this good translation is produced by a human translator). Usually, good translations produced by the machine tend to share a certain number of word with human-translated texts. The metric, therefore, identifies the number of common words shared by the MT output and the reference texts, assigning a score ranging from 0 to 1; the higher the score, the closer the translation to its reference. To be more specific, such metric compares cluster of words, known as *n*-grams. BLEU is programmed to run a

comparison between the n -grams of the reference translation and the n -grams of the candidate translation, counting the number of matches between the two. The higher the number of matches, the closer the score is to 1. The algorithm that makes BLEU work relies on the *precision* measure. In other words, to make the computation precise, developers managed to create what they call “modified unigram precision” (Papineni et al. 2002, 312). By applying this measure, words in the reference sentence are not counted anymore by the system after a matching word is identified in the candidate text. As shown in the paper by Papini et al., BLEU produces striking results. This evaluation metric scores translation in a way that is really close to what human evaluators would do. Even though this metric works well, there are some problems concerning the reporting of BLEU scores. As pointed out by Post, some issues concerning BLEU “make it difficult to evaluate and compare BLEU scores across papers, which impedes comparison and replication” (Post 2018). In order to avoid this problems, as Post describes in the abovementioned paper, he developed the Python script SacreBLEU, that will be implemented also by the present study.

TER – in the paper by Snover et al. “A Study of Translation Edit Rate with Targeted Human Annotation” (Snover et al. 2006, 223-231), the metric TER (Translation Edit Rate) is described as a measure which is more intuitive than previous metrics, such as BLEU. TER calculates “the number of edits needed to fix the output so that it semantically matches a correct translation” (Snover et al. 2006, 223). In other words, the metric looks at the translation produced by the machine and calculates the number of changes that a human translator would implement in order to produce a correct translation. It is important to notice that TER has been programmed so to find the minimum number of edits required; in order to do so, the metric measures the number of edits needed to make the translate content similar to the closest reference. The formula applied is the following:

$$\text{TER} = \frac{\text{\# of edits}}{\text{average \# of reference words}}$$

In TER, all the edits have the same cost, independently from the kind of edit they are (insertion, deletion, substitution of single words or substitution of string of words). According to the formula presented above, the less changes are required, the higher is the TER score (the metric uses the best lowest score obtained). One of the main flaws of this metric is the fact that it ignores all the issues related to semantic equivalence; therefore, the work of human annotators might be required, leading to the metric known as HTER (Human-targeted Translation Edit Rate). In this system, human annotators edit the output produced by an automatic system (the hypothesis); then TER is computed using the reference targeted by humans as human reference. According to the study carried out by Snover et al., HTER is “less subjective than pure human judgments” (ibid., 230); the experiment showed that there is a high correlation between the scores assigned by BLEU and TER, while at the same time HTER is closer to human evaluation. For the purposes of the present study, the metric TER will be used to evaluate MT output. Its closeness to the BLEU system hopefully will give the possibility to compare the results, even though the focus will be on a different aspect of the translation compared to what BLEU is based on.

hLEPOR (harmonic mean of enhanced Length Penalty, Precision, n -gram Position difference Penalty and Recall) – this system analyses the number of similar n -grams between an MT output and a reference translation, assigning a score that ranges from 0 to 1 (with 1 indicating the closest possible result to the reference). In the paper by Li et al. “Language-independent Model for Machine Translation Evaluation with Reinforced Factors” (Li et al. 2013, 215-222) hLEPOR is described as a “language-independent model for machine translation evaluation” (Li et al. 2013, 219). The main purpose of hLEPOR and of its enhanced version hLEPOR_E was that of creating an evaluation model for MT outputs that could consider many different aspects of language. According to the researchers, the main problem with other metrics was that they could work very well with certain language pairs, but not with others. Moreover, metrics might be flawed when it comes to the linguistic features they take into consideration: they might consider few, or too many of such features. The evaluation metric hLEPOR, therefore, tries to address some of this problems, with “enhanced factors, tunable parameters and optional linguistic

information” (ibid., 2016) that make it more versatile. First of all, the metric applies newly implemented penalties, such as an Enhanced Length Penalty (which is applied both on longer and shorter sentences) and the *N*-gram Position Difference Penalty (which compares the word order between the output and reference translation). Then, hLEPOR and its enhanced version considers the values “Precision” and “Recall”, which refers respectively to output accuracy and to loyalty to translation length performance. According to the results discussed in the abovementioned paper, hLEPOR_E obtained better results than other metrics, with improvements in the correlation scores on the language pairs German-English and French-English.

COMET (Crosslingual Optimized Metric for Evaluation of Translation) – this metric represents a new approach to the evaluation of machine translated texts. As presented in the overview available at [https://unbabel.com/research/comet/#:~:text=COMET%20\(Crosslingual%20Optimized%20Metric%20for,\(such%20as%20MQM%20scores\)](https://unbabel.com/research/comet/#:~:text=COMET%20(Crosslingual%20Optimized%20Metric%20for,(such%20as%20MQM%20scores),), this model assigns a score based on information retrieved both from the source input and the target reference text. Moreover, according to the developers, the scoring provided by COMET is so close to that of human translators, that this metric can substitute them, giving the chance to drastically reduce the costs of MT outputs evaluation. Rei et al. in the paper “COMET: A Neural Framework for MT Evaluation” (Rei et al. 2020, 2685-2702) describe COMET as an approach developed to be suitable to new Neural MT models. According to the researchers, COMET takes a step further compared to traditional metrics, which focused on lexical-level features; however, they believe that the quality of the outputs produced by new Neural models cannot be computed using such traditional metrics. The new approach to translation requires a new method to score its performance. Based on recent studies, “current metrics struggle to accurately correlate with human judgment at segment level and fail to adequately differentiate the highest performing MT systems” (Rei et al. 2020, 2685). This metric has been developed in order to resemble human judgment as much as possible; it incorporates and uses two distinct architectures: an Estimator model and a Translation Ranking model. The first model is “trained to regress directly on a quality score”, while the second model “is trained to minimize the

distance between a better hypothesis and both its corresponding reference and its original source” (Rei et al. 2020, 2686). After conducting several studies on different language pairs, the researchers managed to establish that they reached “new state-of-the-art results for segment-level correlation with human judgments” (ibid., 2693).

chrF++ (CHaRacter-level F-score++) – this metric is a variant of the system called CHaRacter-level F-score, or chrF. This metric, introduced in 2015 by Maja Popović, assigns a score based in the calculation of similarity between an MT output and a reference translation; the difference from other models that use the same system relies on the fact that this specific metric is based on character n -grams, rather than word n -grams. As specified at the address <https://machinetranslate.org/chrF>, “the chrF metric compares the machine translation output with reference translations, looking at character sequences. Character sequences matching help in recognizing different forms of a single word”. As Popović points out in the paper “CHRF: character n-gram F-score for automatic MT evaluation” (Popović 2015, 392-395), character n -grams have already been used in other evaluation metrics, and their importance has already been proven. Nevertheless, nobody has yet investigated the possibility of developing an evaluation metric that is entirely based on character n -grams. The metric is based on the calculation of the number of similar n -grams in the hypothesis and in the reference; moreover, the formula used for the computation uses a parameter (β) that assigns in certain cases more importance to the recall, than to precision. The formula is the following:

$$\text{CHRF}\beta = (1 + \beta^2) \frac{\text{CHRP} \cdot \text{CHRR}}{\beta^2 \cdot \text{CHRP} + \text{CHRR}}$$

After conducting some experiments, as explained in the abovementioned paper, the research found out that the scores assigned by chrF are better comparable to human-assigned scores than other metrics, such as BLEU, or TER (Popović 2015, 393). Even though this metric seems to be a promising one, the researchers point out that it has been tested only on one non-European language; therefore, its

capacity to work with and evaluate effectively such languages has still to be researched. The difference between chrF and chrF++ relies on the use of word n -grams on the part of the latter. Therefore, the metric used for this study assigns a score not only on the base of character n -grams, but also on the base of word n -grams.

BERTScore (Bidirectional Encoder Representations from Transformers) – as stated in the paper by Zhang et al. “BERTScore: Evaluating Text Generation with BERT” (Zhang et al. 2019, 1-43), this model has been developed in order to solve some problems presented by other metrics which, according to the researchers, are not capable of assigning precise scores. The main problem is represented by the method used by other metrics to compute scores to assign, since they just rely on the overlapping between n -grams in the candidate and reference sentence. This evaluation therefore stops at a surface-level evaluation, without considering deeper characteristics, and failing to “account for meaning-preserving lexical and compositional diversity”. (Zhang et al. 2019, 1). Moreover, these metrics based on n -grams cannot evaluate correctly paraphrases, since they do not identify any overlapping of n -grams; this issue leads to underestimate the performance of an MT software, since phrases that are semantically correct differ from the reference at the surface level. Another problem caused by n -grams is represented by the fact that they cannot score adequately distant dependencies, penalizing “semantically-critical ordering changes” (ibid.). BERTScore has therefore been developed to address these problems, obtaining effective results as the studies conducted demonstrate. The functioning of BERT is based on the comparison between a reference sentence and a hypothesis sentence: the system matches each token in the reference to each token in the hypothesis to compute recall, then it matches each token in the hypothesis to each token in the reference to compute precision (ibid., 4). Tokens are represented using contextual embeddings, which “can generate different vector representations for the same word in different sentences depending on the surrounding words, which form the context of the target word”; by doing this, the system becomes sensible to those semantical changes that are not considered by other metrics. Then, the system applies a similarity measure, which considers the tokens while isolated, and not as part of an embedding, in order to

consider not only groups of words, but also single items. The results of the experiments conducted have been described as satisfactory, with the BERTScore acting as a top-performer when correlated to human judgments, hybrid systems and model selection performance (ibid., 7). Even though other experiments need to be carried out and the full potentiality of BERT still need to be studied, the metric is promising, and achieves really good results.

The present study will use all of the metrics that are made available in the MT Hub made available by Intento. However, the metrics that will be actually considered are BLEU and TER. The two strongly correlate, as explained above, and they are the two most common and widely used metrics. Moreover, TER seems really interesting thanks to its focus on the edits that the text requires. This specificity opens to the possibility of making interesting comments on the analyzed texts.

1.3

ARTIFICIAL INTELLIGENCE

In recent years, AI has been applied to the field of translation. First of all, we find applications of AI in the field of NMT, with Deep Learning being a product of Artificial Intelligence. Moreover, generative AI systems such as ChatGPT have been used to translate texts. In the paper by Tong King Lee “Artificial intelligence and posthumanist translation: ChatGPT versus the translator” (Lee 2023, 1-22) generative Artificial Intelligence is presented as a tool that, if used properly, can achieve better results than Machine Translation systems. Lee presents the outcomes of several experiments carried out on the translation of Japanese sentences into English. The sentences have been translated by Machine Translation systems (Google Translate, DeepL, Bing Microsoft Translator) and by the AI program ChatGPT-3. The paper presents the outputs of the systems and compares them, showing that AI performed better than the other systems. ChatGPT-3 seems to be able to identify and translate properly two important elements of discourse: cohesion and coherence. Especially when it comes to Japanese, cohesion might pose more than a challenge to both human and machine translators, since in

Japanese a grammatical topic introduced in the beginning of a discourse “is assumed to sustain itself through the text that follows until a new topic appears” (Lee 2023, 6). When it comes to the specific sentences used to carry out the experiment, Lee demonstrates how MT systems have been unable to maintain the cohesion of the text. The grammatical topic in this case consisted in a series of pronouns referring to a female Japanese singer. All the MT systems failed to use the singular feminine pronoun to refer to the subject throughout all the text; in certain parts, the female singer was referred to as “he”, or as “Mr.”. Only ChatGPT managed to respect the cohesion of the text, correctly identifying the gender of the subject. Another issue is represented by coherence. In the passage analyzed, text coherence was obtained by the author by using a series of words referencing to sounds. While the human translator managed to find very good solutions to respect text cohesion, ChatGPT used the word “sound” throughout the whole text, without being able to employ all the range of synonyms and sound-related words employed by the human translator. However, as Lee points out, AI software give the possibility to obtain better outcomes through the use of prompts: by feeding the machine with the correct prompts, the result can improve, as demonstrated in the above-mentioned paper.

It goes without any doubt that generative Artificial Intelligence has become increasingly important for the field of translation. It can be directly used to translate texts, in some cases even presenting itself as a better alternative to Machine Translation. The following paragraph will provide an overview of the history and developments of AI, underlying why nowadays we need to understand how Artificial Intelligence works.

Curiously, AI was also “established as a field of study in the 1950s” (Haenlein & Kaplan 2019, 5), even though it became prominent only in recent times. The beginning of AI can be traced back to the 1940s, when in England Alan Turing engineered “The Bombe”, a code-breaking machine developed to decipher the secret codes used by Nazi Germany in Second World War. In 1950, surprised by the result obtained with “The Bombe”, Turing published “Computing Machinery and Intelligence”. The article focuses on creating and testing intelligent machines, formulating the famous Turing test, which is still employed to verify whether or not a system can be considered intelligent. Alan Turing is considered as one of the

fathers of Artificial Intelligence as we know it nowadays, laying the bases for the future studies in this field. The Turing Test is important for many different reasons: not only it provided a guide to rate AI performance, but it also identified some of the characteristics that are now part of modern Artificial Intelligence systems. According to the British mathematician, a machine could pass the Turing Test only if it showed some key capabilities, like: Natural language processing, Knowledge representation, Automated reasoning and Machine learning (Toosi et al. 2021, 451). As it can be seen, these listed by Turing are all traits nowadays commonly associated to AI systems. Moreover, the model proposed by Turing was later used by other researchers to propose models that could concretely contribute to the developing of an Artificial Intelligence. For example, basing their work on an article written by Turing, McCulloch and Pitts in 1943 proposed a computer model that was capable of learning autonomously thanks to an internal system similar to the net of neurons that make the human brain (Muthukrishnan et al. 2020, 394). Later, this model was refined even more. In 1958 the psychologist Frank Rosenblatt proposed an improved version of the neurons proposed by McCulloch and Pitts: the Perceptron. This was a neural network that had the ability to learn through a process of trial and error. The idea behind this mathematical model was that of simulating the biological human neuron. The model was programmed to multiply a series of inputs x by a series of corresponding weight values w ; to the sum of these values was then summed a bias value b . The value obtained was computed through an activation function f , which generated a binary value (1 or 0). The Perceptron and the following studies on an AI architecture that could simulate human neurons laid the basis for the contemporary research on Machine Learning models.

The word “Artificial Intelligence” was used for the first time only in 1956 at the Dartmouth Summer Research Project on Artificial Intelligence, when John McCarthy described these studies as the science of making intelligent machines (Toosi et al. 2021, 453). After that, the field of AI was characterized by great successes, like the program ELIZA, an early attempt at creating a chatbot that could simulate a human conversation, or the General Problem Solver Program. The 1970s and 1980s saw an abrupt stop to the fundings for AI research programs. After the successes obtained in this field of research, both scientists and the public opinion were overly confident that they could obtain even better results and improvements in a limited amount of time; unfortunately, this was not the case, since the limited

computing power available stopped the research on neural networks (Muthukrishnan et al. 2020, 396). Both the UK and the USA decided to stop funding this field of study, which was highly criticized. Moreover, these two decades are characterized by very few improvements and successes compared to the previous ones. Anyways, some improvements occurred. For example, it was introduced a gradient descent to optimize neural networks, a system later used by LeCun to make the system work well when it came to computer vision problems, such as the recognition of handwritten digits (Toosi et al. 2021, 457). Later on, in 1986, back propagation was introduced: a series of layers all connected together where neurons in the previous layer are connected to those in the following layer. The back propagation algorithm revolutionized neural networks, improving their learning capabilities (Muthukrishnan et al. 2020, 396). Neural networks can be considered as a watershed in the history of AI. As Haenlein & Kaplan explain in the article “A Brief History of Artificial Intelligence”, the reason why at the beginnings of this field of study there were few progresses in AI programs is because those programs were the so-called “Expert Systems”, that is to say “collection of rules which assume that human intelligence can be formalized and reconstructed in a top-down approach as a series of “if-then” statements” (Haenlein & Kaplan 2019, 8). These systems perform very well if formalization is required, but when formalization is not required, they tend to fail. To perform more complex tasks it is necessary a system that can interpret external data, learn from them and use what it learnt to achieve specific objectives and to perform specific tasks by adapting itself in a flexible way (ibid.). All of this can be done by artificial neural networks. However, serious research on such networks could start only in the late 1990s. The best example of how far the research went, is the developing of Deep Blue in 1997: an AI software developed to play chess that managed to defeat the Chess Grandmaster and world champion Garry Kasparov. Thanks to the work of the abovementioned LeCun, convolutional networks were added to the already existing neural networks. In the last years, AI took a step further thanks to improvements in data storage and thanks to the introduction of the GPU (Graphical Process Unit). As pointed out by Muthukrishnan et al. (Muthukrishnan et al. 2020), the performance of Machine Learning algorithms is directly linked to the data available. The more data could be fed to the algorithm, the better the algorithm works; the improvements in data storage capacities increased the number of available data, which on its turn

improved the performance of Machine Learning systems. The introduction of GPUs represented a shift in the computational power of the machines used to research on AI, allowing to overcome the hardware limitations that constrained the research in the previous decades.

Nowadays, studies are focusing on a new kind of AI thanks to what is known as “Deep Learning”. Deep Learning was first introduced in 2012, when a group of researchers developed AlexNet, “a convolutional neural network architecture” with the capacity to train multiple layers of neuron simultaneously (Toosi et al. 2021, 457). Since this moment, the researcher focused on deep learning-based methods, that achieved incredible results. In 2016, for example, the program Deep Mind’s AlphaGo managed to beat the world champion of the board game Go, a board game so complex that for long time it was believed that computers would have never been able to beat humans in this game (Haenlein and Kaplan 2019). For this reason, researchers are positive that they will be able to obtain great improvements in the future, working with Deep Learning.

At this point, another issue arises. Artificial Intelligence is for sure an interesting field of research, and its applications are numerous. Precisely the great number of application to many different field of our life lead researchers to investigate methods that could explain how Artificial Intelligence works, or take decisions. The article “Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges” (Xu et al. 2019, 563-574) deals precisely with this topic. The main problem represented by contemporary AI network is represented by the fact that it is not possible to explain how Deep Neural Networks (DNNs) function; even though the different architectures and structures are well known, still it is very difficult to understand the actual process that lead them to take decisions. Research on explainable AI is believed to become “increasingly important to all groups of stakeholders, including the users, the affected people, and the developers of AI systems” (Xu et al. 2019, 566), especially in those fields such as medicine. In the case in which AI systems are used to identify the pathology a person is suffering from, it is of pivotal importance that the experts using the system are able to understand why the software took that specific decision. Two techniques commonly used to explain DNNs are Sensitivity Analysis (SA) and Layer-wise Relevance Propagation (LRP). While SA “quantifies the importance of each input variable”,

LRP “explains predictions relative to the state of maximum uncertainty” (Xu et al. 2019, 569). Practically speaking, if these algorithms would be used to explain how AI interprets a certain picture, LRP would tell “how much each pixel contributes to prediction”, while SA would tell “how much do changes in each pixel affect the prediction” (ibid.). SA and LRP are currently at the center of the research in the field of explainable AI, a field of study which is also trying to optimize strategies that could make these explanation models accessible to the public, for example introducing textual explanation of the decision-making process of the DNN.

As explained in this paragraph, AI is a powerful and complex tool, capable of reaching great results. Therefore, it comes natural to ask oneself if it can give any contribution to the field of translation. While MT systems do not give the possibility to interact with the user, AI offers this possibility: prompts could be used to suggest the machine how to complete a certain translation task. This high interaction level, along with the high complexity of this algorithms, opens interesting possibilities in the field of translation. Moreover, it might be interesting to apply the findings of the studies on explainable AI on this specific field, so that to understand the paths undertaken by the machine when translating.

1.4

MACHINE TRANSLATION AND ARTIFICIAL INTELLIGENCE APPLIED TO TRANSLATION

MT and AI are nowadays an essential part of the translation process, to the point that translation can be considered as a form of human-computer interaction, also known under the acronym HCI (O’Brien 2012, 1). The topic of MT and AI applied to translation will not be dealt with in general, but there will be a specific focus on literary translation, since it can be considered the “hardest task for human translators” (Toral & Way 2015, 1); therefore, it would be useful to find out whether or not these technological tools can help literary translators becoming more efficient. When approaching a literary text, a translator must face a series of challenges, such as interpreting idiomatic expressions, preserving the literary style, or conveying linguistic subtleties and nuances. An additional challenge is represented by wordplay, sometimes present in certain texts (Škobo & Petričević 2023, 322). As already discussed and demonstrated in the previous paragraphs,

Machine Translation and Artificial Intelligence have reached such a level of refinement that they can provide really good translation outputs. However, humor still represents a difficult task. As it will be discussed in the following sections, humor is usually deeply related to the culture and grammatical features of a language, requiring a level of creativity and transcreative abilities that machine struggle to reach.

There have been several studies on MT applied to different textual types, such as abstracts of scientific texts, literary works and social media (Ardi et al. 2022). Given the interest in MT, it came natural to focus also on the capacities of this software when applied to the translation of humorous devices, given the challenge that these textual features pose to human translators. Some of these studies have been carried out on the occasion of a series of workshops focusing on Computational Linguistics for Literature, the first of which was established in 2012. As described by Toral and Way (Toral and Way 2015, 241-268) MT has been applied to the translation of prose, poetry (for example, it has been used to translate “The Divine Comedy”) and short stories; the two researchers therefore concluded that “the quest to study the applicability of MT to literary texts is timely” (Toral & Way 2015, 6). In 2022 a group of researchers from the State University of Padang, Indonesia, conducted a study on the translation of humorous text by Machine Translation machines, using Google Translate, Yandex Translation and Microsoft Bing Translation. As they discuss in the paper “Can Machine Translations Translate Humorous Texts?” (Ardi et al. 2022, 99-112), the research was focused on identifying the errors committed by machines when translating these texts; in other words, they tried to identify the reasons why Machine Translation cannot provide a faithful and good translation of humorous texts. The study used the descriptive quality approach, trying to categorize the different translation errors present in the outputs. The group of researchers identified the following categories: Lexical Errors, Syntactic Errors, Semantic Errors and Pragmatic Errors. According to the results of the study, the most common type of error is the Lexical error, meaning that the machine usually fails to choose the appropriate words. This is understandable if we think about the great quantity of jokes in which the punning device is a homophone word; if the Target language does not offer the possibility to recreate the joke using an homophone, clearly the machine will opt for another word, which would not obtain the same effect in the original text. Then there are

Syntactic Errors, which refer to tenses, prepositions and sentence structure. Semantic and Pragmatic errors are related with “the changes of messages, meaning, politeness that may affect accuracy of the translation and message in the story and twist of the humor” (Ardi et al. 2022, 108). What is interesting about the results of this study, is that apparently the best Machine Translation system is Google Translate, that implemented Artificial Intelligence and Deep Learning, as well as Neural Machine Translation. Therefore, this might suggest that Neural Machine Translation might be successful where other kinds of systems (such as Statistical MT) tend to fail. This study by Ardi et al. presents an interesting overview on Machine Translation outputs when applied to the translation of humor, since it presents the most common errors committed by these machines. The present research, on the other hand, will move a step further, trying to identify a possible strategy to improve the performance of the machine. To carry out the research, only systems based on Neural Machine Translation will be used, since it represent the latest development in the field of Machine Translation. However, further studies on how to improve the quality of NMT output need to be conducted, since Neural MT “fills the gap to the human quality level only by 20%” (Matusov 2019, 11). The possibility of applying NMT to literary translation has already been the object of interest of other studies, for example the research carried out by Evgeny Matusov presented in the paper “The Challenges of Using Neural Machine Translation for Literature” (Matusov 2019, 10-19) and the study presented in the paper “Neural Machine Translation of Literary Texts from English to Slovene” (Kuzman, Vintar, and Arčan 2019). The latter study was conducted focusing on the English-Slovenian language pair. The authors trained different MT models using a literature-based corpus; contrarily to what someone might think, these literature-trained models did not achieve the best results. Apparently, best results were achieved by those models trained on millions of more general examples. This strongly hints to the fact that in some cases MT models needs to be trained with general content, without the need to focus on a tailored training. Nevertheless, the present research will use a literature-based corpus anyways. The main aim of this study is not that of evaluating the general performance of NMT when applied to literary texts, but rather that of evaluating how well NMT would handle pun translation after training. MT training is an aspect that will be discussed later, since this strategy will be employed to carry out the present study. In spite of this, the study by Kuzman, Vintar and Arčan

confirmed that post-editing of MT generated outputs is generally faster than translation from scratch (Kuzman, Vintar, and Arčan 2019), demonstrating that MT can be of some help when it comes to literary translation. The work presented by Matusov focused in the translation of passages taken from literary texts from English to Russian and from German to English. The researcher used back-translated data to train state-of-the-art NMT systems, noticing that translation quality improved according to evaluation metrics (Matusov 2019, 17). The study made some interesting findings. First of all, it seems that evaluation metrics such as BLEU and TER do not grant any correct result when applied to literary texts. After training and adapting the original systems used, the research noticed how the metrics did not identify any improvement in the MT output, while a bilingual evaluator granted that the adapted system could provide better results (ibid., 13). Moreover, this study introduced a new system to categorize the errors that might occur in a translation performed by MT. This new system might come in handy also for the study presented in this paper. Matusov classified MT Errors as:

- Severe meaning error (a word or short phrase translated into a word or short phrase that has the wrong meaning in the context)
- Minor meaning error (similar to the previous one, but the translated item still conveys the original meaning, even though with slight changes)
- Unknown word or segmentation error (to identify those words that are either not recognized by the translator, or are segmented in the wrong way, resulting in a wrong translation)
- Consistency/term translation error
- Pronoun resolution error
- Locution error (when the machine wrongly translates an idiom)
- Omission, insertion, repetition errors
- Severe syntax error (when the structure of the Target Text is wrong)
- Minor syntax error (when the Target Text contains minor errors regarding syntax or morphology)
- Tone/register error

This evaluation categories might be useful since they provide a scheme human evaluators could follow when assigning a score to translations, in order to check the suitability of the score assigned by the metric. According to the results obtained by

Matusov, Machine Translation tends to generate errors related to consistency, pronoun resolution and tone/register, while syntax errors were quite low. It means that the machine is generally capable of producing a sentence that flows well in the target language, but it has much more problems when it comes to elements that are not present at the surface level, and that might be culture-related (such as the tone of the sentence). Furthermore, MT fails when it comes to translating idioms, given the high number of idiom errors. These results are particularly interesting and useful for the present study. Not only they suggest that the evaluation of a human translator is needed to correctly evaluate literary texts, but they also provide an interesting metric to use in order to check the text and identify errors. Moreover, these results suggest where to look when searching for errors in a machine-translated text.

When it comes to AI applied to literary texts, the main challenge that is to be faced is how to provide AI translations that could be faithful to the original text, while at the same time preserving its artistic nuances (Škobo & Petričević 2023, 319). Therefore, AI and MT translation of literary texts share the same challenge, since the machines are required to think and work like human translators. Being machines, they might fail to identify the salient traits of a text which has nothing to do with the standardized way of writing typical of, for instance, scientific papers. What is different, is the way in which the two systems work. AI opens to the possibility of greater level of interaction than MT, requiring a completely different approach than MT systems when it comes translation. As it will be discussed later, while Machine Translation systems can be trained in order to become effective when applied to literary translation, Artificial Intelligence systems require the user to understand which are the most suitable prompts to feed the machine with (a practice that is the object of interest of the field of research known as “prompt engineering”). AI is making progress in the field of literary translation, as demonstrated by the research carried out by Škobo and Petričević, and described in the paper “Navigating the Challenges and Opportunities of Literary Translation In the Age of AI: Striking a Balance Between Human Expertise and Machine Power” (2023, 317-336). For this study twenty Anglistics students from the Sinergija University were assigned the task of translating from Serbian into English an excerpt from “Lake Como”, by Srđan Valjarević. Then, the quality of each translation was assessed on the basis of general subjective observation and on the basis of the IELTS band descriptors; after the assessment, it was decided whether

or not every translation was to be failed or passed. After the human assessment, it was asked to Chat GPT-3 to evaluate each translation, in order to compare human understanding of literary translation with the understanding of the machine. The researchers observed that human assessment and machine assessment were very similar, with AI “capable of detecting and identifying different kinds of translation mistakes and oversights in each translation” (Škobo & Petričević 2023, 330-331). The study therefore concluded that AI has improved when it comes to literary translation, and that it is likely to improve in the coming years. These results might demonstrate that if AI is capable of identifying translation errors, then it might be possible that it has gained some understanding of how a translation work should be carried out. As a result, it is possible to hypothesize that AI could be able to translate correctly a text, using the knowledge gained in that specific field. Other studies have been conducted on this topic; for example, the study by Alkodimi, Alqahtani, and Al-Wasy, presented in the paper Human-AI collaboration in translation and back translation of literary texts (Alkodimi, Alqahtani, and Al-Wasy 2024, 173-192) investigated the potentiality of AI-based tools when applied to literary translation. The study did not use AI systems in itself (such as, for example, ChatGPT), but it used AI-based MT systems. The results of the research were promising, showing that AI-based tools can be of great help to translators, especially when it comes to the practice of back translation. Therefore, Artificial Intelligence can be usefully employed in the field of literary translation, and it can even be helpful to students, as this study demonstrated. As Alkodimi and the fellow researchers found, the performance of the students who were asked to translate a text using AI-based tool was much better than that of the students who translated the same text without such tools. Moreover, thanks to their features, AI tools can be of great help since they can “understand the context of the source text” and “identify and correct errors that can be missed by translators”, allowing users to reduce inaccuracies and enhance the quality of the (back)translation (Alkodimi, Alqahtani, and Al-Wasy 2024, 187). The pedagogical role played by Artificial Intelligence is underlined also in the study “A Cross Sectional Study of ChatGPT in Translation: Magnitude of Use, Attitudes, and Uncertainties” (Sahari, Al-Kadi, and Ali 2023, 2937-2954). In this study researchers asked both students and teachers of fields such as Linguistics, Literature, and Translation to answer a questionnaire containing several open-ended and closed questions on their preferences about tools that could assist them in the

translation process. The majority of students clearly opted for ChatGPT, which was considered to be more helpful than Google Translate. Students stated that ChatGPT had the ability to generate accurate translation, and to provide a TT with correct grammar and spelling. The AI tool was also praised for his capacity to generate texts that can be proofread, edited and paraphrased. AI can speed up the work of a translator, generating a text that is of an acceptable quality. This study clearly demonstrates the pedagogical potentialities of Artificial Intelligence, a consideration that, on its turn, generates another thought. If AI can be used in teaching contexts to train future translators, it means that this tool can actually be used in the translation process. Of course, as the above-mentioned study pointed out, AI has many limitations, especially when applied to literary translation; in such cases, the authors recommend students to be cautious. The finding about ChatGPT (and AI in general) applied to translation (and literary translation) are promising, but they are still exploratory findings. The authors clarify that “while the machine, using deep learning, can produce mechanical results, it cannot fully comprehend the narrative and context that underlie the text it is given” (Sahari, Al-Kadi, and Ali 2023); and in literary translation, context is one of the key elements of the text. Therefore, scholars are optimistic about the use of AI for both working and pedagogical purposes, but further research is needed.

Even though MT and AI nowadays are really sophisticated and constantly improve, still there are aspects of literary translation with which they struggle: humor and wordplay. As pointed out by Miller in the paper “The Punster’s Amanuensis: The Proper Place of Humans and Machines in the Translation of Wordplay” (Miller 2019, 57-65) humorous language is a prime candidate for the application of approaches based on natural language processing, like Machine Translation (Miller 2019, 57). However, these approaches work in quite a rigid way, being able to apply fixed set of rules derived from their knowledge of vocabulary, grammar and semantics. These systems therefore struggle to translate humorous language, which is extremely varied and far from being standard language. Before describing studies that focused on MT and AI translation of puns and wordplay, an overview of these linguistic phenomena will be provided, so to clarify what the research will be focused on.

HUMOR, PUNS AND THEIR TRANSLATION

Humor is one of the most studied mechanisms of language (Miller 2019, 57). In modern linguistics, the two most accepted theories formulated to explain the mechanisms of humor are the Script-based Semantic Theory of Humor (SSTH) and the Generalized Theory of Verbal Humor (GTVH), which extends the previous one (ibid., 58). Raskin (Raskin 1987, 11-25) states that SSTH is based on the idea that a text can be considered a joke if the given text is compatible (fully on in part) with two different scripts, and if the two scripts in question are opposite in a special sense. The possible oppositions the texts can refer to are situation/unreal, good/bad, sex/no sex, money/no money, or life/death. Raskin uses as an example the following joke:

“Is the DOCTOR at home?” the patient asked in his bronchial whisper.

“No”, the DOCTOR’s young and pretty wife whispered in reply. “Come right in”.

According to Raskin, in this example the two scripts are DOCTOR, which is mentioned, and LOVER, which clearly is the wife of the doctor. The variety of the opposition between the scripts is sex/no sex. Obviously, the invitation made by the wife to enter despite the absence of the husband, triggers the switch from one script to the other. The main point of Raskin’s theory entirely relies on the text of the joke, being its aim that of analyzing the script itself, and not the psychological reasons or other reasons behind the joking mechanism. According to the scholar, SSTH identifies five core components of a joke:

1. The switch from a mode of communication that is straightforward and does not imply anything else but what is said (*bona-fide* mode of communication), to a mode of communication that implies something (*non-bona-fide* mode of communication). Humor is, in

fact, different from the actual mode of communication we use in our everyday life, since in joke telling the utterer is not committed to the truth of what they are saying, contrarily to what supposedly happens in common communication. Obviously, for a joke to work, the listener needs to be aware that the utterer is deliberately saying something that is not true (in other words, that the utterer is pronouncing a *non-bona-fide* speech)

2. The text of the joke
3. Two scripts compatible with the text and that partially overlap
4. An opposition between the two scripts
5. A trigger that allows for the script switching

Therefore, according to SSTH “humor is evoked when a given text is compatible, at least in part, with two different scripts that “opposite” in some culturally significant sense” (Miller 2019, 58). This notion is known as “script opposition” (SO), and it is one of the six parameters of GTVH. GTVH is the development of SSTH, obtained after expanding the main idea behind SSTH. If this theory had just one explanatory dimension, the Generalized Theory of Verbal Humor has six, not stopping at the semantic level of analysis. In a paper by Raskin, Hempelmann and Taylor (Raskin, Hempelmann, and Taylor 2010, 285-312) the GTVH is described as a theory that “aims to provide a framework for capturing further relevant information, linguistic and other, both unique to jokes, as well as used elsewhere in language and cognition in general” (Raskin, Hempelmann, and Taylor 2010, 291). This new theory was developed in order to get a deeper understand of jokes and of their internal mechanisms, and it does not rely only on the semantics of the joke itself, but considers also extratextual components. As described in Miller’s and Raskin’s aforementioned papers (Miller 2019, 58; Raskin, Hempelmann, and Taylor 2010, 291-292) according to this theory, a joke is defined by six parameters:

- Script opposition (SO). This is the only element present in the Script-based Semantic Theory of Humor, from which this present theory originated
- Logical mechanism (LM), the (situationally-)false and pseudological reasoning that apparently resolves the incongruity of the SO masking the oppositeness of the texts involved

- Situation (SI), the setting of the joke, which is non-humorous
- Target (TA) of the joke. In other words, the object of the joke. It might be optional, and usually it is stereotypical
- Narrative strategy (NS), that is to say the “genre” of the joke (riddle, conversational retort, humorous short story...)
- Language (LA) used to convey the joke. Language is the textual artifact, which can be spoken, written, mimed, gestured, expressed through dress...

As it can be seen, humorous language is really complex. It is a multilayered linguistic phenomena, that requires the understanding of the language, as well as the understanding of meanings that not always are clear and immediate. This is the reason why it poses many challenges, not only to human translators, but also to the machines. Going deeper into the object of study of this paper, now wordplay and puns have to be addressed.

Wordplay is pervasive in every situation of our everyday life, not only in literature. It appears in a wide range of situations in communicative exchanges, such as spontaneous verbal manifestations or purposeful uses in advertising; wordplay also appears in many different authors, cultures and historical periods (Winter-Froemel, Thaler & Demeulenaere 2018, 1). This field of language tends to be very complex, as pointed out by Attardo in the paper “Universals in puns and humorous wordplay” (2018, 89-109), since wordplay may take many different forms, such as games related to word spelling, anagrams, palindromes, acrostics, word squares, or even the creation of new languages (Attardo 2019, 89-90). As it can be seen, wordplay is mainly graphemic, that is to say, it is based on the spelling of the words (except in those cases where forms of verbal wordplay are used) (Attardo 2019, 90). Puns are a specific kind of wordplay. Attardo provides the following definition of pun:

“A pun is a textual occurrence in which a sequence of sounds must be interpreted with a reference to a second sequence of sounds, which may, but need not, be identical to the first sequence, for the full meaning of the text to be accessed. The perlocutionary goal or effect of the pun is to generate the perception of mirth or of the intention to do so.” (Attardo 2019, 91)

Puns involve at least two different senses, but not necessarily two different words, as it can be seen in the following example: “Life depends on the liver” (Low 2011, 64). In this case, the pun is based on the word “liver”, a single word that contains two different senses: the word can be interpreted in the most common meaning, that it to say “liver” as an organ of the human body; or it can be interprets as “liver” in the sense of “the person living”. The pun is generated by this double-reading of the word.

There are also puns resulting from syntactic ambiguity, or from morphological ambiguity, as it happens in the pun “Iraqi head seeks arms”, where “head” has the double meaning of “part of the body” and “chief”, while “arms” has the double meaning of “parts of the body” and “weapons” (Attardo 2018, 92). Another group of puns are alliterative puns, which are based on the repetition of a phoneme or group of phonemes. This latter category is explained by Salvatore Attardo with the following example: “You remember Sunset Strip – where the unneat meet to bleat!”. Here, the repetition of the sound [i] is “highly noticeable” (ibid.).

As it has been made clear, since wordplay and puns are based on specific features of a given language (Low 2011, 59), they are among the most challenging textual features to be translated. Having said that, instances of wordplay (puns included) are not to be considered untranslatable. They can be translated, but specific strategies and approaches have to be adopted. First of all, it is preferable a textual approach rather than an isolated one (Vandaele 2011, 181), considering the wordplay as part of a bigger picture and not just as single instance of language. The strategies that translators can adopt are different. As described in the paper “Restructuring a New Approach of Pun Translation Strategy” (Zhang, Seong & Muhammad 2014, 2498-2506) Delabastita (Delabaststia 1996, 134) formalized a series of strategies to translate puns:

- Pun to pun, meaning that the pun in the ST is transferred into a pun in the TT. This pun may have the same properties of the original one, but this not always is possible

- Pun to non-pun, the original pun is translated using a word or phrase that not always preserves all the senses of the original pun
- Pun to punoid, a strategy that requires the translator to recreate the effect of the original pun by using other wordplays or rhetorical devices
- Zero translation, that omits the original pun
- Direct copy, the pun is reproduced in the ST without translating it
- Transference, this strategy is similar to the direct copy, but it imposes ST signified on a TL text
- Addition, a strategy that consists in translating a non-pun with a pun
- Addition of new textual material, a strategy that consists in translating from zero to pun
- Editorial technique, which include many different things, such as articles published by the translator, introduction, epilogue within the volume, footnotes and endnotes, parentheses

As made clear by this overview, wordplay and especially puns are particularly difficult to translate because they can be of many different types, and rely on many different mechanisms that make them work and be funny; therefore, they require the adoption of many different translation techniques. In order to translate a pun a translator not only needs deep knowledge of the SL, but also of the context in which the text was originally produced. Therefore, the question arises: are machines nowadays capable of performing such a complex task? It has already been explored how MT and AI might be limited. First of all, “older” MT models, such as Statistical MT, are based on fixed rules, or they are based on corpora that need to be increased and updated in order for them to work properly. Then, as already demonstrated above, humor translation and its understanding require high levels of creativity and transcreative capacity, without mentioning the deep knowledge of the culture and language that are needed to understand certain jokes. The latest models of Artificial Intelligence are highly intelligent, and they have incredible learning abilities. However, they still are machines, they can just mimic human intelligence and not be as clever as a human. These digital tools usually do not take into consideration linguistic anomalies and ambiguities, or they tend to consider them as errors that need to be corrected (Kolb & Miller 2022, 66). Anomalies and ambiguities are what wordplay stands on; therefore, it would be interesting to test the capacities of these really advanced tools when they have to

deal with the translation of such content. It is indeed true that neural networks helped these tools becoming more efficient and more similar to a human mind, but there is still a lot of work that needs to be done, as already discussed in the dedicated sections. Nonetheless, many studies have been carried out focusing on the application of AI and MT to wordplay and puns.

It is important to point out that thanks to the advances in technology, these tools can indeed be useful if not to directly translate puns, at least to provide support to translators (*ibid.*). A useful tool developed to support pun translation is PunCAT: it translates “each sense of the pun separately”, then allowing the translator to “explore the semantic fields of these translations in order to help construct a plausible target-language solution that maximizes the semantic correspondence of the original” (*ibid.*). This tool has been evaluated in the paper “Human–computer interaction in pun translation” (Kolb & Miller 2022, 68-88). According to the researchers, PunCAT can be useful to translators. Even though the program does not automatically translate puns, it still can provide support by “stimulating creative thinking, providing inspiration, and broadening the translator’s pool of solution candidates” (Kolb & Miller 2022, 85). In the paper, two features are highly appreciated: the employment of rhyming dictionaries and similar resources, thanks to which it is easier to retrieve terms that phonetically match; the location and interpretation of punning words in the TT, which happens automatically.

PunCAT however presented some problems. First of all, it may not be working with some language pairs. As pointed out in the aforementioned paper, for every language in the language pair there should be “electronic lexical-semantic networks, interlingual links between the concepts of the two networks, and pronunciation information in the form of electronic pronouncing dictionaries of grapheme-to-phoneme models” (*ibid.*, 71). This can be a huge limitation, since if these prerequisites are not met, PunCAT cannot be employed. Then, the tool seems not to be suitable for everybody: some translators found PunCAT not very useful, since it did not fit their working style, or they were not familiar with it (*ibid.*).

There have been other studies focusing on the application of MT and AI to wordplay, such as the JOKER track at CLEF in 2022 and 2023. The JOKER track consisted in a series of studies on the application of MT and AI to wordplay; these studies were conducted by scholars that met on the occasion of the CLEF, a

conference that since 2000 contributes to the evaluation of information access systems. In this conference researchers worked on automated pun and humor translation. While studies such as the one conducted with PunCAT and those carried out at the JOKER tracks focused mainly on pun detection and translation, the present study will be focused on finding strategies that could improve MT and AI when applied to pun translation since, as already discussed, these tools tend to face many issues when it comes to performing these tasks. As described in the paper “FAST-MT Participation for the JOKER CLEF-2022 Automatic Pun and Humor Translation Tasks” (Dhanani, Rafi & Tahir 2022) Transformer-based models were used to perform three tasks related to pun identification and translation. Transformer-based models are a form of neural network: they learn contextual information from sequential data by understanding the existing relationships between them. The study consisted of three tasks to be performed:

- Task 1: classifying and explaining any instance of wordplay present in a given extract
- Task 2: translating nouns containing a wordplay
- Task 3: translating a phrase containing a wordplay

For the first task, seven distinct models were trained independently in order to predict the value of seven different variables:

- Location, words that construct the wordplay
- Interpretation, a possible explanation for the given wordplay
- Horizontal/vertical, a binary categorical column to detect if the target and source of the wordplay co-occur in the text
- Manipulation_type, a variable detecting if source and target of the wordplay are exact equivalents, if they weakly resemble each other, if they possess different ordering, or if they are a group the initials of which form a funny meaning
- Manipulation_level, a variable detecting if the wordplay is a phonological manipulation (sound), or a textual-based written manipulation (written), or some other form
- Cultural_reference, a true/false variable detecting if the given wordplay belongs conventional form or not

- `Conventional_form`, a true/false variable detecting if the wordplay belongs to conventional form or not

Researchers decided to analyze English texts possessing a wordplay and an id: the id is the input value that associates a unique wordplay text; the wordplay is an input in English containing a wordplay. First of all, researchers prepared a model trained to extract words from each wordplay in the English texts analyzed. Every wordplay was considered as a series of tokens separated by spaces, and each token was identified with one of the following categories:

- `Word_play_token_B`, that identifies the word beginning the wordplay
- `Word_play_token_I`, identifying the other words in the wordplay
- `Other_token`, identifying all the words not belonging to the wordplay

After that, a model was constructed that could generate an interpretation for each extracted wordplay.

In order to perform Task 2, the machines were mapped to learn the relation between English nouns and the corresponding French translations. Researchers employed the open-source parallel corpus OPUS to develop the context for the English/French noun pairs provided by the data set of Task 2. The English/French sentence pairs were then extracted from OPUS: they contained the selected English noun and the correspondent French noun. At this point the task consisted in using English nouns as a query to predict the location of the corresponding French translation. Task 3 consisted in translating entire phrases containing wordplay. The model in this situation needed to predict an equivalent French translation for a given English text. However, there could be multiple possible translation for the same input.

According to the researchers, the models adopted are capable of handling wordplay location. However, the interpretation task was completed with more difficulties: in Task 3, which involved the translation of entire phrases containing wordplays, only 13% of the phrases were automatically translated in a satisfying way. The researchers concluded that machine translation still is incapable of translating puns, and successful translations were “accidental due to the existence of the same ambiguous words” (Popova & Dadić 2023, 5) in both languages that were object of study (English and French). This topic was further explored a year

later, in the JOKER track at CLEF 2023. The research carried out in 2023 is described in the paper “Does AI Have a Sense of Humor? CLEF 2023 JOKER Tasks 1, 2 and 3: Using BLOOM, GPT, SimpleT5, and More for Pun Detection, Location, Interpretation and Translation” (Popova & Dadić 2023). This research made use of the Python programming language and of other different methods provided by artificial intelligence and machine learning to perform detection, location, interpretation and translation of word sets. The tasks machine had to perform were the same of the JOKER track at CLEF 2022: detection, location and interpretation of puns. This time, the puns were to be translated from English into Spanish and French.

For Task 1 (pun detection), the data provided to the machine consisted in positive examples, that is to say short jokes that contained a single pun. Negative examples of puns were provided by intervening to the positive examples: researchers manually removed the pun, while maintaining the same meaning of the sentence. By doing these, chances were lower that machines could tell apart positive from negative examples by detecting variations in length, style or vocabulary. Task 2 consisted in two different parts:

- Identifying a word that contained the pun
- Interpretating the pun

For the first part of the task the machines were given training data with the word already found and test data containing only the test of the joke. The second part was based on positive examples: the pun ward was “annotated with two sets of words representing each sense of the pun” (ibid., 6). Task 3 aimed at translating English puns into French and Spanish while maintaining the form and meaning of the original. The translation approach that had to be followed was Delabastita’s pun→pun strategy. The tasks were performed using different artificial intelligence methods, which were both pre-training and neural methods: TF-IDF, SimpleT5, Random, Naive Bayes, MLP, Logistic Regression, Fast Text, SpaCy, GPT3, BLOOM, WordNet, Googletrans, EasyNMT-Opus and EasyNMT-mbart. After comparing all the results obtained, researchers came to the conclusion that AI can perform simple tasks such as pun detection and location, but struggles to perform pun interpretation and translation. As the researchers put it, AI is little by little developing sense of humor, but still results are not perfect; moreover, AI cannot

independently translate creative instances of language, the intervention of human translators still is required.

Basing on the results of these studies, it is possible to formulate the hypothesis that simpler instances of wordplay and puns can be translated by MT and AI. First of all, it has been demonstrated that if the puns in the different languages contain an exact translation of the ambiguous word the machines can translate it correctly. Usually, simpler puns are based just on a word that has a double sense; therefore, chances are high that AI or MT might translate them easily. In some rare cases, two languages might contain a word that recreates the same exact semantic ambiguity, without the need to recreating the joke, or to adapt it to the target language; in such rare cases, the dictionary the machine could rely on is necessary to perform the translation. Moreover, simple task such as pun detection and location can be performed without any struggle. For this reason, it is highly probable that simpler puns might be identified without any problem by AI or MT. Therefore, in order to avoid working with unreliable data, only challenging instances of humorous language will be selected. For the purpose of this research, only the instances of humorous language that are more complex will be selected, in order to lower the chances that machines might perform well. This decision has been taken because the research will be effective only if it will be able to identify the weaknesses of the technological tools employed: the aim is that of finding solutions in order to let machines do what now they are not able to do. To clarify, the present study will focus mainly on those cases in which:

- The pun or wordplay relies on the context. Therefore, the machine abilities to understand the whole context will be required
- The pun or wordplay has some features that are typical of the English language, and chances are low that they could be reposed in the same exact way by an Italian word

As demonstrated by the research described above, MT and AI cannot approach humorous language completely on their own, they need the help of a human translator. However, the studies examined in this literature review did not provide many applicable solutions to solve the problems faced by these tools when they are assigned the task of translating humorous language. The paper by Popova and Dadić proposed some solutions, but they are related to the creation of “libraries,

models and databases” (ibid., 20) in English and other languages. As it can be seen, this solutions cannot be directly implemented by translators, since they are applicable to people with expertise in IT and machine learning. Moreover, translators need useful strategies that can be immediately applicable while they are carrying out a job. These kind of strategies described in the paper are long-term strategies, that cannot be applied immediately. Thus, the focus of this paper is the identification of strategies that could be directly and immediately implemented by translators. If it’s true that MT and AI could be of great help, it is also true that to favor a complete human-computer-interaction between translators and this kind of technologies it is necessary to find solutions that intervene directly on the machines and on their outputs, without resorting to IT specialists or to machine learning knowledge. The research that will be described in the paper therefore aims at finding this specific kind of solutions, using the instances of wordplay and puns retrievable from the novel “Alice’s Adventures in Wonderland” in order to find a solution to the main problems occurring when AI and MT are asked to translate this kind of language.

Preliminary Analysis

This section of the paper will be devoted to the preliminary analysis of the puns. After this short introduction, the paper will present the puns that have been selected as the object of the study, explaining why they have been chosen. Then, focus will be shifted towards the methods used to investigate the puns.

2.1

OBJECT OF THE RESEARCH

A group of five puns and wordplays taken from the book “Alice’s Adventures in Wonderland” was selected. As stated in the previous section, in order to conduct a fruitful research, “simple” puns were avoided. With the word “simple”, I refer to those puns in which the simple translation into Italian of the word would be enough to translate the joke, reposing an equally humorous joke in Italian. For example, all the puns similar to the following will be avoided: “I like kids, but I would never eat one of them”. In order to translate effectively this pun, it would be enough just to translate the word “kid” into Italian, obtaining the following translation: “Mi piacciono I bambini, ma non ne mangerei mai uno” (this Italian pun is the literal translation of the English Source Text). As it can be observed, the punning mechanism in this case does not rely on any morphological, lexical, or cultural feature. The word “kids” evokes the same association both in English and Italian; therefore, a simple translation of the word is enough to translate the joke properly. Of course, this kind of puns can be flawlessly translated by the machine and therefore they will not be taken into consideration. Now, an overview of the puns and wordplay selected will be provided. I will also try to provide a translation of the puns using the approach proposed by Delabastita in order to highlight the difficulties met by human translators when translating these instances of humorous language.

A long and sad tail – in this passage, Alice is talking to a mouse that claims to have a long and sad story to tell. Here is the passage taken from the book:

“You promised to tell me your history, you know,” said Alice, “and why it is you hate – C and D,” she added in a whisper, half afraid that it would be offended again.

“Mine is a long and sad **tale**!” said the Mouse, turning to Alice, and sighing.

“It *is* a long **tail**, certainly,” said Alice, looking down with wonder at the Mouse’s **tail**; “but why do you call it sad?” And she kept on puzzling about it while the Mouse was speaking, so that her idea of the **tale** was something like this: –

Then, the mouse proceeds to tell its story, which is visually represented in the novel as a tail.

In this case, it is clear how the whole pun is played on the ambiguity generated by the homophones **tale** and **tail**, which refer to two distinct objects. The hilarious effects of the pun are amplified by the confusion that the homophony generates in Alice, who looks at the tail of the Mouse asking why it called the tail “sad”. This pun is particularly difficult to translate leaving intact the humorous effect of the original. The homophony cannot be reposed in the Italian language. The translation of the word “tail” in “coda”, while the word “tale” can be translated as “racconto”, “storia”. It is evident that the Italian language does not offer the chance to recreate the homophony that characterizes the original text. Now, a translation of the pun will be attempted based on Delabastita’s suggested strategies.

Being this the situation, “Pun to pun” translation has to be discarded. A possible strategy might be “Pun to non-pun”; however, given the humorous soul of the book and the importance humorous has in the economy of the novel, this strategy is not recommended. It might be possible to translate using other wordplays or rhetorical devices, applying the “Pun to punoid” strategy:

“Sai, hai promesso di raccontarmi la tua storia,” disse Alice, “e come mai tu odi – C e D,” aggiunse con un sussurro, mezza impaurita di poterlo offendere ancora.

“La mia **vita** è lunga e triste!” disse il Topo, guardando Alice mentre singhiozzava.

“Di sicuro è una lunga **vita**,” disse Alice, guardando con meraviglia come la **vita** del Topo si estendesse in una lunga **coda**; “ma perché dici che è triste?” E continuò a pensarci mentre il Topo parlava, tanto che si immaginò la **vita** del Topo più o meno così:–

Here, I decided to use a different kind of pun. Instead of a pun relying on the homophony of the words involved, I decided to make the punning mechanism revolve around the double meaning the word “vita” could have in Italian. “Vita” can be used to refer to someone’s “life”, or to someone’s “back”. By doing this, the humorous purpose of the original pun is maintained: Alice does not understand why the Mouse is describing his back as sad, since she confuses the two meanings of the word “vita” similarly to how in the original text she confuses the words “tale” and “tail”. This strategy required a little addition to the Target Text which is absent from the Source Text. In order to make explicit that the tail is considered as an extension of the Mouse’s back, I decided to translate “said Alice, looking down with wonder at the Mouse’s tail” as “said Alice, looking down with wonder at the Mouse’s back extending into a long tail”.

The main challenge, therefore, is represented by the absence of Italian homophones that could work as direct equivalents of the words “tale” and “tail”, requiring the translator to resort to a different translation technique than “Pun to pun”. Moreover, some addition is needed to make the joke clear to the readers.

Irish apples – in this passage Pat, the Rabbit’s gardener declares that he is “digging for apples”:

Next came an angry voice—the Rabbit’s—“Pat! Pat! Where are you?” And then a voice she had never heard before, “Sure then I’m here! **Digging for apples**, yer honour!”

“Digging for apples, indeed!” said the Rabbit angrily. “Here! Come and help me out of *this!*” (Sounds of more broken glass.)

“Now tell me, Pat, what’s that in the window?”

“Sure, it’s an arm, yer honour!” (He pronounced it “**arrum.**”)

“An arm, you goose! Who ever saw one that size? Why, it fills the whole window!”

“Sure, it does, yer honour: but it’s an arm for all that.”

Of course, this does not make any sense, since apples grow on trees and not in the ground. However, this would make much more sense to an English-speaking reader

of the book, since the whole joke revolves around the nationality of Pat. As made explicit by the text, Pat pronounces the word “arm” in quite an uncommon way: “arrum”. This uncommon pronunciation is typical of Irish-speaking people; knowing this, it becomes clear while Pat states that he is “digging for apples”, since “Irish apples” was a slang for “potatoes” used in Victorian England, the time in which the book was written. Of course translating this wordplay is quite difficult, since it requires the reader to know this cultural aspects in order to understand the intended joke. Furthermore, the Italian public would not associate potatoes to apples, since in the Italian culture potatoes are not considered to be the “apples of the earth”. The main difficulty here derives from the need to culturally adapt the joke, again resorting to the “Pun to punoid” strategy. In this case, an element of the text might be helpful: the “unusual” pronunciation of the word “arm”. It might be possible to recreate the joke by making Pat speak in a typical Italian dialect; therefore, the word “arrum” could be translated so to recreate a regional way to pronounce the word “braccio”. Here it is an attempt at translating the passage, trying to turn Pat into a representative of an Italian region:

Si sentì a quel punto una voce arrabbiata – quella del Coniglio – “Pat! Pat”
Dove sei?” Gli rispose una voce che non aveva mai sentito prima, “Mo sono
qua! Sto mangiando i tortellini, vostro onore!”

“A mangiare i tortellini!” disse rabbioso il Coniglio. “Dai! Vieni qua e aiutami
a uscire di *qui!*” (Di nuovo un rumore di vetri rotti.)

“Adesso dimmi, Pat, cos’è quella cosa nella finestra?”

“Ma si capisce, l’è un braz vostro onore!”

“Si dice “braccio”, brutta oca! E poi chi l’ha mai visto di queste dimensioni?
Occupi tutta la finestra!”

“Ma si capisce, vostro onore; però è di sicuro un braccio.”

In this case, I tried to render the original wordplay by making Pat speak as person who comes from the Emilia-Romagna region, trying to repropose (in an exaggerated way) some features of the dialect typically spoken there. Particularly, the phrase “l’è un braz” (“it is an arm”) tries to reproduce the dialect of that region. Moreover, to underline this aspect, I made Pat performing another activity instead of looking for potatoes: Pat is eating tortellini, a typical variety of pasta from this

region. Another possible way to translate the passage would be a combination of two strategies: “Direct copy” and “Editorial technique”. By doing this, translators would repropose the pun exactly as it appears in the ST, adding a footnote that explains the cultural subtleties that make the joke work in the original version of the novel.

Wasting time – this passage is deeply influenced by the grammatical features of the English language, that allow to choose from three different classes of pronouns: feminine, masculine and neutral. The passage is the following:

Alice sighed wearily. “I think you might do something better with the **time**,” she said, “than waste **it** in asking riddles that have no answers.”

“If you knew Time as well as I do,” said the Hatter, “you wouldn’t talk about wasting **it**. It’s **him**.”

The Hatter here suggests Alice not to address Time as “it”, but rather as “him”. therefore, the Time is not considered as a neutral, abstract entity, that cannot be represented neither as a feminine, nor a masculine being. Apparently, the Hatter considers Time as a masculine being, hence the suggestion to address this entity by the pronoun “him”. Of course, it is not possible to repropose the same joke in Italian, since it lack the neutral class of pronouns. In order to translate the pun, translators need to resort again to a strategy that is different from “Pun to pun”. Here I attempt a possible translation:

Alice sospirò pesantemente. “Penso che dovresti impiegare meglio il tuo tempo,” disse, “che sprecarlo facendo indovinelli che non hanno risposta.”

“Se conoscessi il Tempo bene quanto me,” disse il Cappellaio, “non parleresti di *impiegarlo*, è molto pigro,”

In this case, I opted again for the “Pun to punoid” strategy. If in the original the pun revolves around the different pronouns associated with the entity Time, I decided to make the pun revolve around the idea of “giving a job to the Time”. In Italian, the verb “impiegare” can be used either in the sentence “impiegare il tempo”, which

means “spending time on a certain activity”, or in the sentence “impiegare qualcuno”, a rather old-fashioned way to express the idea of “giving a job to someone”. The original pun also played with the non-human nature of time: being a non-human entity, Alice correctly uses the pronoun “it”. The Hatter, however, uses the pronoun “him”, as if time was a living, human entity. The same concept is re-proposed in the Italian translation with the verb “impiegare”, which is intended by the Hatter with the sense of “giving a job”. Of course, it is impossible to “give a job” to the time, since it is a non-human, abstract entity.

Biting Flamingoes and Mustard – in this passage the punning mechanism revolves around the verb “to bite”, which can be used to describe both the action of biting performed by an animal, and the “biting” taste of mustard:

“Very true,” said the Duchess: “flamingoes and mustard both **bite**. And the moral of that is—‘Birds of a feather flock together.’”

The Italian language does not offer the possibility to use the direct equivalent to the verb “to bite”. First of all, in Italian the taste of mustard cannot be described as “biting”; then birds are not said to “bite” when attacking someone. Additional difficulty is added by the motto used at the end of the passage, even though an Italian motto with a similar meaning can be easily found. The major problem is represented by the translation of the verb “to bite”.

“Questo è molto vero,” disse la Duchessa: “i fenicotteri hanno becchi **pungenti** quanto la senape. E la morale è—‘Chi si somiglia si piglia.’”

To translate the passage I decided to use the adjective “pungente”, which in a way can translate the idea of the biting taste of mustard, while recalling the sharp and pointy ends of the flamingos’ beaks. The motto has been translated with an Italian motto sharing the same meaning. In both of the cases, I had to apply the “Pun to punoid” strategy, creating a new punning mechanism in the TT.

Twinkling Tea – this last passage plays on the sound [t]. Before discussing the passage, it is appropriate to present it:

“Give your evidence,” the King repeated angrily, “or I’ll have you executed, whether you’re nervous or not.”

“I’m a poor man, your Majesty,” the Hatter began, in a trembling voice, “—and I hadn’t begun my **tea**—not above a week or so—and what with the bread-and-butter getting so thin—and the **twinkling** of the **tea**—”

“The **twinkling** of the *what?*” said the King.

“It *began* with the **tea**,” the Hatter replied.

“**Of course twinkling begins with a T!**” said the King sharply. “Do you take me for a dunce? Go on!”

As it can be seen in the passage above, the Hatter states that the twinkling began with the tea, intending the beverage. Again, Carrol uses an homophony to generate a humorous outcome, since the King angrily replies to the Hatter that he is well aware that the word “twinkling” starts with the letter “t”, clearly confounding the word “tea” with the letter that shares the same sound. Here it is an attempt at translating the passage:

“Mostra le tue prove,” ripeté il Re con rabbia, “o ti farò giustiziare, indipendentemente dal fatto che tu sia nervosa oppure no.”

“Sono un pover’uomo, vostra Maestà,” cominciò il Cappellaio, con voce tremante, “—e non avevo ancora cominciato il tè – non prima di una settimana o giù di lì – e con il pane imburrito che stava diventando sottile – e il luccichio del tè –”

“Il luccichio di *cosa?*” chiese il Re.

“Il luccichio cominciò col tè” rispose il Cappellaio.

“**Con me non è cominciato proprio nulla!**” disse il Re brusco. “Pensi che non sappia cosa faccio? Vai avanti!”

Again, I had to resort to a strategy that did not take into account the possibility to repropose the joke as it is in the original. Given that the Italian word “tè” is not a homophone of the letter “t”, I decided to exploit the homophony of the word “tè”

(the beverage) with the word “te” (a pronoun used to address a second singular person). Therefore, I recreated the joke as if the King understood that the twinkling began with him (i.e. the King). This translation is quite weak, but I think it is a good try at exploiting two homophones as in the original. Moreover, in the Italian translation still there is a humorous behavior on the part of the King, that appears quite naïve and self-centered, being unable to understand that the word “tè” refers to the beverage and not to his person.

Analysis

This section will be devoted to the analysis of the instances of humorous language described above. As it was demonstrated, the specific characteristics and nuances of the jokes require transcreation and appropriate cultural knowledge to translate the jokes correctly and effectively, maintaining the humorous effect of the original text. Now, these puns and wordplays will be translated into Italian using the Machine Translation providers *Intanto ModernMT* and *RWS Language Weaver*; then I will translate them using *Microsoft Edge Copilot* and *ChatGPT 3.5*. After the translation, I will present the results, analyzing the errors committed by the machines during the translation phase. In order to evaluate the translation I will score MT outputs using the metrics offered by the *Intanto* suite: *COMET*, *TER*, *BLEU*, *BERTScore*, *hLEPOR* and *chrF++*. Moreover, I will score them personally, applying the criteria proposed by *Matusov*, which have been specifically developed to provide a human rating of NMT outputs. The same criteria will be applied to the evaluation of texts generated by generative AI, since the automated evaluation metrics used to evaluate MT outputs have not been designed to evaluate AI outputs.

It is important to clarify that the categories *M1* and *M2* will be dealt with in a way suitable to this study, that probably differs from the original way these categories were thought to be used. With “Meaning” I will mainly consider the humorous purposes of the original ST. I will consider as meaning errors all of those occurrences in which the translation could not repropose the ST in a way suitable to recreate its humorous intent. Therefore, *M1* and *M2* will not refer only to those errors that can actually be considered as “meaning errors” (in the sense intended by *Matusov*), but also to those translation that do not repropose the ST in a way suitable to maintain its humorous purposes.

3.1

INTENTO MODERN MT

Evaluation on the basis of the criteria proposed by Matusov was carried out using a table, in order to make the evaluation more clear and well-ordered.

Translation Portal

Translate with

ModernMT Human-in-...

From

English
EN

To

Italian
IT

“Sai, hai promesso di raccontarmi la tua storia,” disse Alice, “e come mai tu odi – C e D,” aggiunse con un sussurro, mezza impaurita di poterlo offendere ancora.
“La mia vita è lunga e triste!” disse il Topo, guardando Alice mentre singhiozzava.
“Di sicuro è una lunga vita,” disse Alice, guardando con meraviglia come la vita del Topo si estendesse in una lunga coda; “ma perché dici che è triste?” E continuò a pensarci mentre il Topo parlava, tanto che si immaginò la vita del Topo più o meno così:-


“Sai, hai promesso di raccontarmi la tua storia”, disse Alice, “e come mai tu odi – C e D”, aggiunse con un sussurro, mezza impaurita di poterlo offendere ancora.
“La mia vita è lunga e triste!” disse il Topo, guardando Alice mentre singhiozzava.
“Di sicuro è una lunga vita”, disse Alice, guardando con meraviglia come la vita del Topo si estendesse in una lunga coda; “ma perché dici che è triste?” E continuò a pensarci mentre il Topo parlava, tanto che si immaginò la vita del Topo più o meno così:-


Translated via ModernMT Human-in-the-loop

Type of error	Number of errors	Explanation
Severe meaning error	1	The original pun was not reposed properly. It is not possible to understand the pun. The mistranslation of the words “tail” and “tale” made it impossible for the reader to understand the meaning of the pun without looking at the Source Text. Even though the machine picked the correct Italian direct equivalents, this strategy was not suitable to translate the text correctly.
Minor meaning error	1	The translation “e come mai tu odi – C e D” is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by ModernMT was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	


Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

Translation Portal


Translate with 

 ModernMT Human-in-...

From

English
 

To

Italian
 

Next came an angry voice—the Rabbit's—"Pat! Pat! Where are you?" And then a voice she had never heard before. "Sure then I'm here! Digging for apples. ver honour!"
"Digging for apples, indeed!" said the Rabbit angrily. "Here! Come and help me out of this!"
 (Sounds of more broken glass.)
"Now tell me, Pat, what's that in the window?"
"Sure, it's an arm. ver honour!" (He pronounced it "arrum."
"An arm, you goose! Who ever saw one that size? Why, it fills the whole window!"
"Sure, it does. ver honour: but it's an arm for all that."
 |


Poi arrivò una voce arrabbiata - quella del Coniglio-"Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima:"Certo, allora sono qui! Alla ricerca di mele, tuo onore!"
 "Scavare per le mele, davvero!" disse il Coniglio con rabbia. "Ecco! Vieni e aiutami a venirme fuori!" (Suoni di altri vetri rotti.)
 "Ora dimmi, Pat, cosa c'è nella finestra?"
 "Certo, è un braccio, tuo onore!" (Lo pronunciò"arrum.")
 "Un braccio, oca! Chi ne ha mai visto uno così grande? Perché, riempi tutta la finestra!"
 "Certo che sì, il tuo onore: ma è un braccio per tutto questo."


Translated via ModernMT Human-in-the-loop

Type of error	Number of errors	Explanation
Severe meaning error	1	The passage is comprehensible, but the original humorous meaning and intent were not translated. As explained above, the humorous device used by Carrol in this passage related to Pat being an Irish speaker. This knowledge, strictly related to cultural knowledge, is necessary to understand why Pat declares that he is digging while looking for apples. The translator was not able to retrieve this cultural knowledge.
Minor meaning error	0	
Unknown word or segmentation error	1	As explained above, lack of cultural knowledge on the part of the machine produced many errors. Modern MT could not recognize the words "arrum" as a way to visually represent the Irish accents; therefore, this word was probably listed as an unknown word. Thus the mistranslation of the word.
Consistency/Term translation error	0	


Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	2	<p>Minor syntax errors emerge in the translation of sentences that have a typical English structure, such as “Digging for apples, yer honour!” and “Digging for apples, indeed!”. In this case the sentences were translated using the infinitive form, while in the original a gerundial is used. Moreover, the verb “cercare” (to look for) seems to be missing, while it would have given much more sense to the whole sentence.</p> <p>Literal translation of “It’s an arm for all that”. The Italian Target Text sticks too much to the original Source Text producing a text that is not immediately understandable. However, this does not make impossible for the reader to grasp the meaning of the phrase.</p>
Tone/register error	0	

Translation Portal


Translate with 

 ModernMT Human-in-...

From

English
 

To

Italian
 

Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."
 "If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."
 |

Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo a fare indovinelli che non hanno risposte."
 "Se conoscessi il tempo come me", disse il Cappellaio, "non parleresti di sprecarlo. È lui."

Translated via ModernMT Human-in-the-loop

Type of error	Number of errors	Explanation
Severe meaning error	1	The pun was not reposed in a good way. The literal translation of the pun makes it impossible to understand the joke without looking at the original sentence.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0,5	This error is marked only as half an error since the Italian grammar does not allow for the use of neutral pronouns.
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

Translation Portal

Translate with ?
 ModernMT Human-in-the-loop

From English EN

To Italian IT

"Very true," said the Duchess: "flamingoes and mustard both bite. And the moral of that is —'Birds of a feather flock together.'"
×

"Verissimo", disse la duchessa: "fenicotteri e senape mordono entrambi. E la morale è: "Gli uccelli di una piuma si affollano". "
📄

Translated via ModernMT Human-in-the-loop

Type of error	Number of errors	Explanation
Severe meaning error	1	The literal translation of the verb “to bite” into the Italian equivalent “mordere” does not work in this case. Using the literal translation does not give the possibility to recreate the original pun, since in Italian the verb “mordere” is not suitable to describe neither the biting action of the flamingoes, nor the taste of mustard.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	1	The motto “Birds of a feather flock together” was not translated with a suitable motto into Italian.
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

3.2

RWS LANGAUGE WEAVER

English ▾

↔ Italian ▾ Generic ⓘ ⚙

"You promised to tell me your history, you know," said Alice, "and why it is you hate – C and D," she added in a whisper, half afraid that it would be offended again.

"Mine is a long and sad tale!" said the Mouse, turning to Alice, and sighing.

"It is a long tail, certainly," said Alice, looking down with wonder at the Mouse's tail; "but why do you call it sad?" And she kept on puzzling about it while the Mouse was speaking, so that her idea of the tale was something like this: –

"Hai promesso di raccontarmi la tua storia, sai," disse Alice, "e perché sei tu ad odiare – C e D", ha aggiunto in un sussurro, mezzo timore che si sarebbe offeso di nuovo.

"La mia è una lunga e triste storia!" Disse il mouse, girandosi verso Alice, e sospirando.

"È certamente una coda lunga", disse Alice, guardando in basso con stupore la coda del mouse; "ma perché la chiami triste?" E ha continuato a sconcertarlo mentre il mouse stava parlando, in modo che la sua idea della storia fosse qualcosa di simile a questa: -

Clear text 488/5000 Rate Translation: ☆☆☆☆☆

Copy Improve Feedback History

Click to upload files or drag and drop on the page. + Add label

Type of error	Number of errors	Explanation
Severe meaning error	2	<p>The translation "ha continuato a sconcertarlo" is a severe meaning error. It did not translate correctly the phrase "she kept on puzzling about it". The ST phrase means that Alice keeps interrogating herself about the Mouse's tale. The Italian sentence means that Alice keeps maintaining a behavior that offends and surprises (in a negative way) the Mouse.</p> <p>The system mistranslated the terms "tail" and "tale". Mistranslation here is used not to refer to the grammar per se, since the grammatical translation is correct. It is used to refer to the mistranslation of the humorous intent of the original ST. The machine translated the two items into the (grammatically correct) Italian equivalents; however, the best way to translate this joke would have been that of using two Italian homophones, so that to repropose the original pun.</p>
Minor meaning error	1	<p>The translation "e perché sei tu ad odiare – C e D" is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by Language Weaver was not suitable to make the</p>

		readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	
Consistency/Term translation error	1	The word “Mouse” is not translated. Probably this is due to the capital letter at the beginning of the noun, that might have brought to categorize this item as a proper noun, and not as a common noun. If this was the case, the machine probably kept the term untranslated because it believed “mouse” to be the proper name of a character. This error was dealt with as if it was a term recognition error, since the system could not recognize the item as a proper nouns.
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	1	“Half afraid” was translated as “mezzo timore”. First of all, the noun “timore” should have been turned into the adjective “timorosa”. Then, even the translation “mezzo timoroso” would have not been a good translation, since it does not sound idiomatic in Italian; a better option could be “mezza impaurita”.
Tone/register error	0	



Next came an angry voice—the Rabbit's—"Pat! Pat! Where are you?" And then a voice she had never heard before, "Sure then I'm here! Digging for apples, yer honour!" "Digging for apples, indeed!" said the Rabbit angrily. "Here! Come and help me out of this!" (Sounds of more broken glass.)

"Now tell me, Pat, what's that in the window?"

"Sure, it's an arm, yer honour!" (He pronounced it "arrum.")

"An arm, you goose! Who ever saw one that size? Why, it fills the whole window!"

"Sure, it does, yer honour: but it's an arm for all that."

Poi è arrivata una voce arrabbiata – quella del Coniglio – "Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima, "certo che sono qui! Scavare per le mele, vostro onore!"

"Scavare per le mele, infatti!" Disse il Coniglio con rabbia. "Qui! Vieni ad aiutarmi!" (Suoni di vetro più rotto).

"Ora dimmi, Pat, che cosa c'è nella finestra?"

"Certo, è un braccio, vostro onore!" (Lo ha pronunciato "arrum".)

"Un braccio, tu oca! Chi ne ha mai vista una così grande? Perché, riempi tutta la finestra!"

"Certo, lo fa, vostro onore: Ma è un braccio per tutto questo".

Clear text 537/5000

Rate Translation: ★★★★★

Copy Improve Feedback History

Click to upload files or drag and drop on the page

+ Add label

Type of error	Number of errors	Explanation
Severe meaning error	1	As explained above, the humorous device used by Carroll in this passage related to Pat being an Irish speaker. This knowledge, strictly related to cultural knowledge, is necessary to understand why Pat declares that he is digging while looking for apples. The translator has not been able to retrieve this cultural knowledge.
Minor meaning error	0	
Unknown word or segmentation error	1	As explained above, lack of cultural knowledge on the part of the machine produced many errors. Modern MT could not recognize the word "arrum" as a way to visually represent the Irish accents; therefore, this word has probably been listed as an unknown word. Thus, the mistranslation of the word.
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	4	"Suoni di vetro più rotto" is wrong. The original phrase refers to Alice hearing again glass breaking. The Italian translation makes the reader think that Alice hears more glass that breaks.

		<p>A Minor syntax error emerged in the translation of the phrases “Digging for apples, yer honour!” and “Digging for apples, indeed!”. The machine translated the verbs with an infinitive form, while a gerundial form would have been a much better option.</p> <p>“Certo lo fa” represents a mistranslation due to the system being too literal. The meaning of the original translation is “Sure, it is an arm”. The system did not recognized the use of the verb “to do” with the meaning of “to be”, but it identified it with a verb having the meaning of “it carries out that work”.</p> <p>The translation “è un braccio per tutto questo” is not idiomatic in Italian. It is a quite literal translation of the original sentence; the reader can understand the meaning of the sentence, but it’s still a syntax error that does not make the phrase an example of good translation.</p>
Tone/register error	0	

English



Italian

Generic



Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."

"If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."

Alice sospirò stanamente. "Penso che potresti fare qualcosa di meglio con il tempo", ha detto, "piuttosto che sprecarla nel chiedere indovinelli che non hanno risposte".

"Se conoscessi il tempo come me", disse il Cappellaio, "non parleresti di sprecare il tempo. È lui."

Clear text 246/5000

Rate Translation: ☆☆☆☆☆

[Copy](#) [Improve](#) [Feedback History](#)

Click to upload files or drag and drop on the page

+ Add label

Type of error	Number of errors	Explanation
Severe meaning error	1	The pun was not reposed in a good way. The literal translation of the pun makes it impossible to understand the joke without looking at the original sentence.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0,5	It is interesting to notice that Language Weaver could not find any suitable equivalent to translate the neutral pronoun "it", opting for substituting it with the noun "time". However, I would not considered it to be a "full" error, since the machine could indeed provide a suitable translation. For this reason, it will be considered as half of an error.
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	1	The translation of "wearily" into "stanamente" instead of "stancamente" has to be considered a sever syntax error, since this word does not exist in the Italian vocabulary.
Minor syntax error	1	The time is referred to as a feminine noun ("sprecarla"), rather than a masculine noun, as it should be according to Italian grammar.
Tone/register error	0	

English

↔

Italian

Generic

?

⚙

"Very true," said the Duchess: "flamingoes and mustard both bite. And the moral of that is—'Birds of a feather flock together.'"

"Molto vero", disse la duchessa: "Fenicotteri e senape mordono entrambi. E la morale di questo è: "Gli uccelli di una piuma si uniscono".

Clear text 130/5000

Rate Translation: ☆☆☆☆☆

📄 Copy

🔧 Improve

📋 Feedback History

Click to upload files or drag and drop on the page

+ Add label

Type of error	Number of errors	Explanation
Severe meaning error	1	The literal translation of the verb “to bite” into the Italian equivalent “mordere” does not work in this case. Using the literal translation does not give the possibility to recreate the original pun, since in Italian the verb “mordere” is not suitable to describe neither the biting action of the flamingoes, nor the taste of mustard.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	1	The motto “Birds of a feather flock together” was not translated with a suitable motto into Italian.
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

This translation was very similar to that provided by Intento Modern MT.

English
Italian
Generic

“Give your evidence,” the King repeated angrily, “or I’ll have you executed, whether you’re nervous or not.”
“I’m a poor man, your Majesty,” the Hatter began, in a trembling voice, “—and I hadn’t begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—”
“The twinkling of the what?” said the King.
“It began with the tea,” the Hatter replied.
“Of course twinkling begins with a T!” said the King sharply. “Do you take me for a dunce? Go on!”

“Date le vostre prove”, ripeté il Re con rabbia, “o vi farò giustiziare, che siate nervosi o no.”
“Sono un povero uomo, vostra Maestà”, ha iniziato il Cappellaio, con una voce tremante, “ – e non avevo iniziato il mio tè – non più di una settimana circa – e con il pane e il burro che diventano così sottili – e il scintillio del tè –”
“il scintillio del cosa?” Ha detto il re.
“È iniziato con il tè”, rispose il Cappellaio.
“Naturalmente il scintillio inizia con una T!” Ha detto il re in modo deciso. “Mi porti per una dunce? Andate avanti!”

Clear text
503/5000
Rate Translation: ☆☆☆☆☆
Copy
Improve
Feedback History

Click to upload files or drag and drop on the page
+ Add label

Type of error	Number of errors	Explanation
Severe meaning error	2	<p>Severe mistranslation of “Do you take me for a dunce”, which was translated literally, resulting in a sentence that has no meaning in Italian.</p> <p>NMT was not able to translate the joke in an effective and acceptable way. The punning mechanism revolving around the homophony of the words “t” and “tea” was not reposed correctly.</p>
Minor meaning error	0	
Unknown word or segmentation error	1	The word “dunce” is not translated. Probably, the machine did not recognize it as a word, therefore maintaining the original word untranslated instead of finding a suitable equivalent.
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	1	<p>“Go on” was translated with the plural “Andate avanti”, while it should have been translated with the singular “Vai avanti”, or “Vada avanti”. Similarly, throughout the whole passage the king refers to the Hatter with a plural form. It is probable that the machine</p>

		<p>opted for this translation in order to reproduce a highly polite way of talking (in Italian, addressing a person with the second plural form is more polite than addressing the same person with the second singular form). However, this form is old-fashioned and not used anymore. Maybe, the machine opted for this form because the person speaking is a king, so it might be related to an old-fashioned way of talking, given that this institution is usually associated with the past times.</p>
Tone/register error	0	

3.3

MT EVALUATION

Before focusing on the evaluation provided by the automated metrics, I will focus on the results obtained by the human-made evaluation. A table will be used in order to present the results more clearly.

	Pre-training Intento ModernMT	Pre-training RWS Language Weaver
Severe meaning error	5	7
Minor meaning error	2	1
Unknown word or segmentation error	1	2
Consistency/Term translation error	1	1
Pronoun resolution error	0,5	0,5
Locution error	1	1
Omission, insertion, repetition errors	0	1
Severe syntax error	0	1
Minor syntax error	2	6
Tone/register error	0	0

As it can be seen, ModernMT performed better than Language Weaver. The most notable categories are: Severe meaning error, in which Language weaver scored 7, compared to the 5 errors committed by ModernMT, and Severe syntax error, in which Language Weaver scored 1, compared to the 0 errors present in the ModernMT output. Looking closely at those categories, it is possible to notice that the segments that were more difficult to translate for the machine were the first segment and the last segment as far as meaning is concerned; when it comes to syntax, the most difficult passage to translate was the third one, where the machine used a word that does not exist in the Italian vocabulary. Generally speaking, both

Language Weaver and ModernMT failed to translate the passage so that the humorous device could be maintained also in the TT.

Now, the metrics' evaluation will be provided. I would like to point out that the main metrics that will be considered are BLEU and TER; not only they are the most used metrics, but they have also been demonstrated to have high levels of comparability. Nevertheless, all the other metrics offered by the Intento MT suite will be presented.

In order to carry out the study, I will use as a reference translation the translation of the text performed by Alessandro Ceni. The reference text is the kindle edition of the novel "Alice's Adventures in Wonderland", published in 2024. The first MT provider is Modern MT, while the second is Language Weaver.

Scoring



Files 1

Report

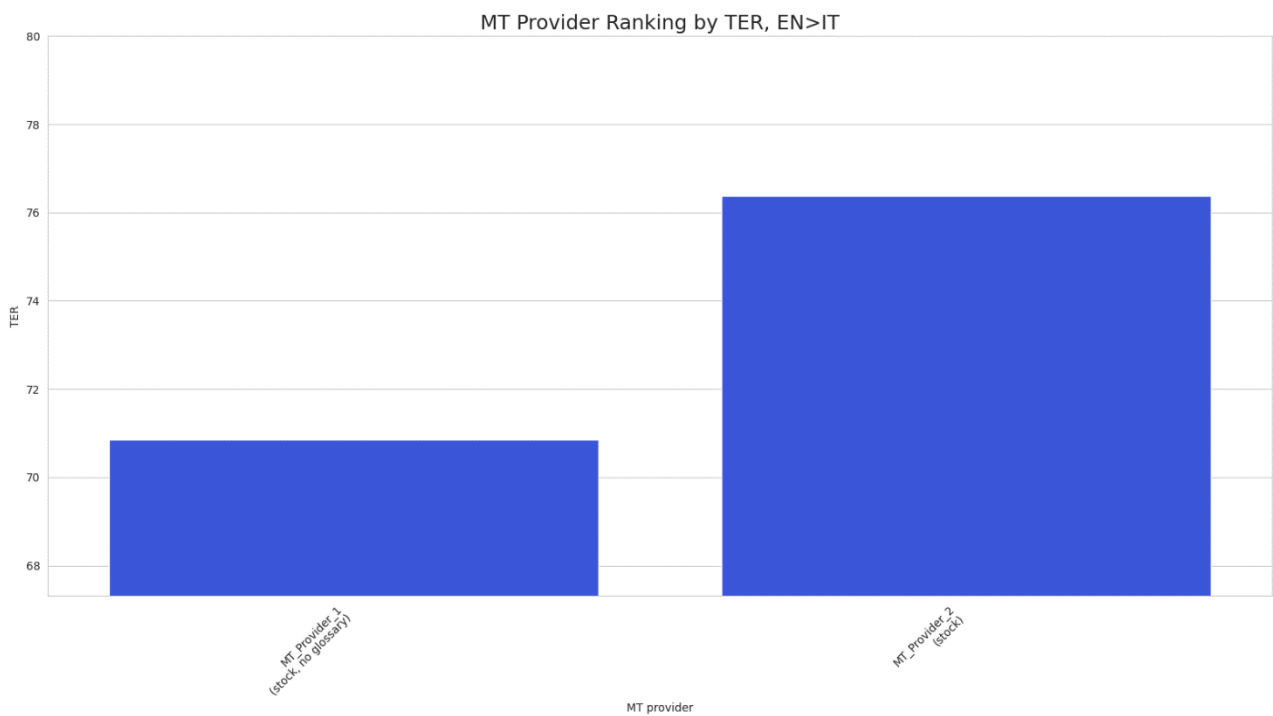
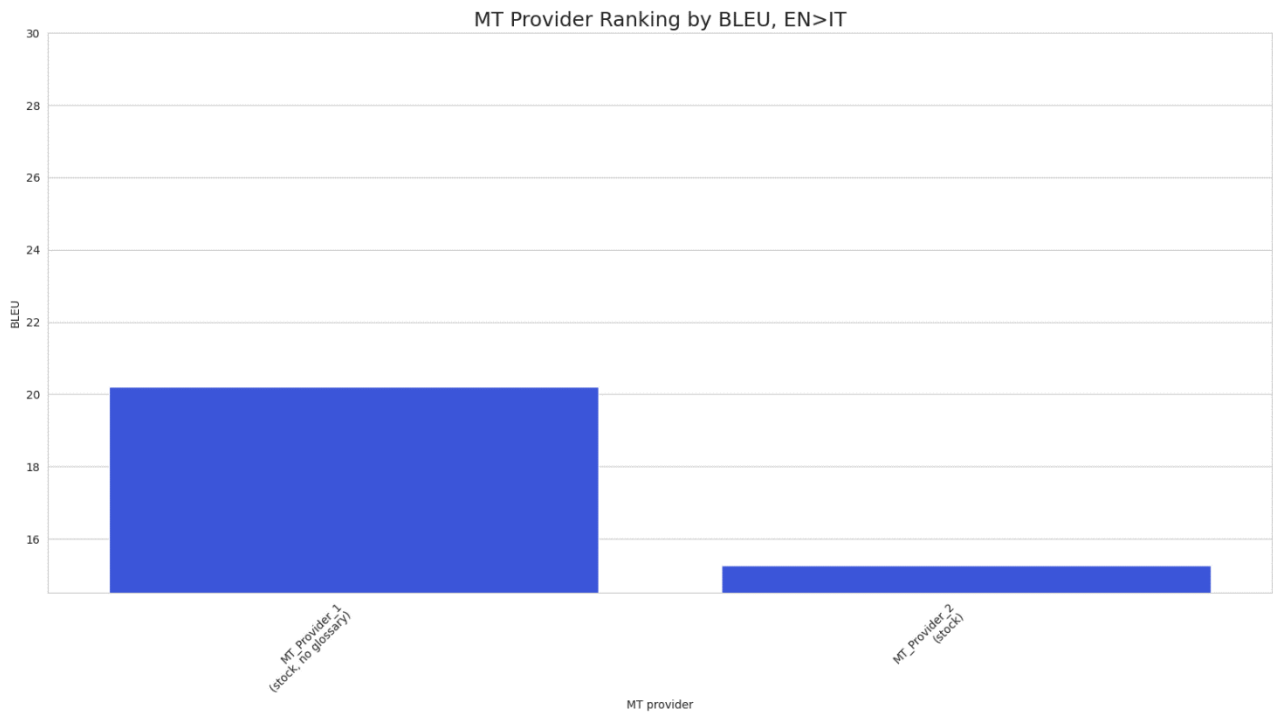
Analyze

Select 2 Models to analyze MT customization results. [What is this?](#)

+ Add models

MT Models		BERTScore ? ↓	hLEPOR ?	TER ?	BLEU ?	chrF++ ?	COMET ?	
<input type="checkbox"/>	MT_Provider_1 (stock, no glossary)	0.807820 0.800345 - 0.815088	0.45 0.391 - 0.506	70.86	20.20	44.33	0.70 0.671 - 0.738	
<input type="checkbox"/>	MT_Provider_2 (stock)	0.789227 0.778974 - 0.79908	0.42 0.395 - 0.453	76.38	15.27	40.84	0.68 0.643 - 0.717	

Here I provided a general view of the scoring results. The same results will be explored in detail now, providing the information taken from the downloadable report produced by the MT Suite; only the information related to TER and BLEU will be provided in this section.



According to the graphs presented, and as already discussed in the section dedicated to the description of the metrics, it is possible to notice that BLEU and TER agreed on the results. Both of them assigned higher scores to Intento ModernMT. This

means first of all that the translation provided by Modern MT is closer to the reference translation when taking into consideration the number of n -grams; secondly, it means a lower number of changes is required in order to make those outputs look like the reference translation.

Now, if we consider what the metrics specifically took into consideration to provide an evaluation, it is interesting to have a look at the scores assigned by TER to the single segments (unfortunately no segment results for BLEU has been provided). As already explained, TER calculates the minimum number of changes that a segment requires in order to be similar to the reference. A screenshot of the results is attached.

MT_Provider_1 (stock, no glossary)	MT_Provider_1 (stock, no glossary)_TER	MT_Provider_2 (stock)	MT_Provider_2 (stock)_TER
- "Hai promesso di raccontarmi la tua storia, sai", disse Alice, "e perché odi – C e D", aggiunse in un sussurro, quasi temendo che si offendesse di nuovo. "La mia è una storia lunga e triste!" disse il Topo, voltandosi verso Alice e sospirando. "È una lunga coda, certamente", disse Alice, guardando con meraviglia la coda del topo; "ma perché lo chiami triste?" E continuava a confonderlo mentre il Topo parlava, in modo che la sua idea del racconto fosse qualcosa del genere: –	76,68711656	"Hai promesso di raccontarmi la tua storia, sai," disse Alice, "e perché sei tu ad odiare – C e D", ha aggiunto in un sussurro, mezzo timore che si sarebbe offeso di nuovo. "La mia è una lunga e triste storia!" Disse il mouse, girandosi verso Alice, e sospirando. "È certamente una coda lunga", disse Alice, guardando in basso con stupore la coda del mouse; "ma perché la chiami triste?" E ha continuato a sconcertarlo mentre il mouse stava parlando, in modo che la sua idea della storia fosse qualcosa di simile a questa: -	78,52760736
Poi arrivò una voce arrabbiata - quella del Coniglio-"Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima:"Certo, allora sono qui! Alla ricerca di mele, tuo onore!" "Scavare per le mele, davvero!" disse il Coniglio con rabbia. "Ecco! Vieni e aiutami a venirne fuori!" (Suoni di altri vetri rotti.) "Ora dimmi, Pat, cosa c'è nella finestra?" "Certo, è un braccio, tuo onore!" (Lo pronunciò "arrum.") "Un braccio, oca! Chi ne ha mai visto uno così grande? Perché, riempi tutta la finestra!" "Certo che sì, il tuo onore: ma è un braccio per	69,23076923	Poi è arrivata una voce arrabbiata – quella del Coniglio – "Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima, "certo che sono qui! Scavare per le mele, vostro onore!" "Scavare per le mele, infatti!" Disse il Coniglio con rabbia. "Qui! Vieni ad aiutarmi!" (Suoni di vetro più rotto). "Ora dimmi, Pat, che cosa c'è nella finestra?" "Certo, è un braccio, vostro onore!" (Lo ha pronunciato "arrum".) "Un braccio, tu oca! Chi ne ha mai vista una così grande? Perché, riempi tutta la finestra!" "Certo, lo fa, vostro onore: Ma è un braccio per	74,35897436
Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo a fare indovinelli che non hanno risposte." "Se conoscessi il tempo come me", disse il Cappellaio, "non parlerei di sprecarlo. È lui."	65,2173913	Alice sospirò stanamente. "Penso che potresti fare qualcosa di meglio con il tempo", ha detto, "piuttosto che sprecarla nel chiedere indovinelli che non hanno risposte". "Se conoscessi il tempo come me", disse il Cappellaio, "non parlerei di sprecare il tempo. È lui."	67,39130435
"Verissimo", disse la duchessa: "fenicotteri e senape mordono entrambi. E la morale è:" Gli uccelli di una piuma si affollano ". "Dà la tua prova", ripeté il Re con rabbia, "o ti farò giustiziare, che tu sia nervoso o meno." "Sono un pover' uomo, Vostra Maestà," iniziò il Cappellaio, con voce tremante, "e non avevo iniziato il mio tè - non più di una settimana o giù di lì - e con il pane e il burro che diventavano così sottili e lo scintillio del tè..." "Il luccichio di cosa?" disse il Re. "È iniziato con il tè", rispose il Cappellaio. "Certo che lo scintillio inizia con una T!" disse bruscamente il Re. "Mi prendi per un somaro? Vai avanti!"	79,16666667	"Molto vero", disse la duchessa: "Fenicotteri e senape mordono entrambi. E la morale di questo è: "Gli uccelli di una piuma si uniscono". "Date le vostre prove", ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervosi o no." "Sono un povero uomo, vostra Maestà", ha iniziato il Cappellaio, con una voce tremante, "– e non avevo iniziato il mio tè – non più di una settimana circa – e con il pane e il burro che diventano così sottili – e il scintillio del tè –" "Il scintillio del cosa?" Ha detto il re. "È iniziato con il tè", rispose il Cappellaio. "Naturalmente il scintillio inizia con una T!" Ha detto il re in modo deciso. "Mi porti per una dunce? Andate avanti!"	87,5
	64,0776699		76,69902913

As it is visible from these results, in every single segment Modern MT achieved better results than RWS Language Weaver.

Now, generative AI will be evaluated. I will apply the same exact criteria applied to NMT scoring, in order to obtain comparable values. In order to carry out the evaluation, I will accept the translation as it is, without feeding generative AI with additional prompts to improve the output quality.

3.4

CHAT GPT 3.5


Open A I Gpt 3.5
BA

Can you please translate the following segment from English into Italian?

"You promised to tell me your history, you know," said Alice, "and why it is you hate – C and D," she added in a whisper, half afraid that it would be offended again.

"Mine is a long and sad tale!" said the Mouse, turning to Alice, and sighing.

"It is a long tail, certainly," said Alice, looking down with wonder at the Mouse's tail; "but why do you call it sad?" And she kept on puzzling about it while the Mouse was speaking, so that her idea of the tale was something like this: –



Sure! Here's the translation into Italian:

"Hai promesso di raccontarmi la tua storia, sai," disse Alice, "e perché odi – C e D," aggiunse sussurrando, mezzo timorosa che si offendesse di nuovo.

"La mia è una lunga e triste storia!" disse il Topo, girandosi verso Alice e sospirando.

"È sicuramente una lunga coda," disse Alice, guardando con meraviglia la coda del Topo; "ma perché la chiami triste?" E continuò a rifletterci mentre il Topo parlava, così che la sua idea della storia era qualcosa di simile a questo: –

Type of error	Number of errors	Explanation
Severe meaning error	1	The system failed to find a suitable translation to "tale" and "tail". Since homophones are not used, the humorous mechanism that underlines the ST is missing. Moreover, the machine provided quite a literal translation, and it did not recreate the joke in a way suitable to the Italian language.
Minor meaning error	1	The translation "e perché odi – C e D" is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by ChatGPT was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	

Severe syntax error	0	
Minor syntax error	1	The translation “mezzo timorosa” contains a disagreement as far as the gender of the subject is concerned. The proper translation would have been “mezza timorosa”.
Tone/register error	0	

For the best ChatGPT alternative, you can also try [Hix AI](#).

Can you please translate the following segment from English into Italian?

Next came an angry voice—the Rabbit’s—“Pat! Pat! Where are you?” And then a voice she had never heard before, “Sure then I’m here! Digging for apples, yer honour!”

“Digging for apples, indeed!” said the Rabbit angrily. “Here! Come and help me out of this!” (Sounds of more broken glass.)

“Now tell me, Pat, what’s that in the window?”

“Sure, it’s an arm, yer honour!” (He pronounced it “arrum.”)

“An arm, you goose! Who ever saw one that size? Why, it fills the whole window!”

“Sure, it does, yer honour: but it’s an arm for all that.”



Here’s the translation into Italian:

Poi si udì una voce arrabbiata—quella del Coniglio—“Pat! Pat! Dove sei?” E poi una voce che Alice non aveva mai sentito prima, “Certo che sono qui! Sto scavando per le mele, signore!”

“Scavando per le mele, davvero!” disse il Coniglio arrabbiato. “Qui! Vieni ad aiutarmi a uscire da qui!” (Rumore di altri vetri rotti.)

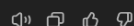
“Ora dimmi, Pat, cos’è quello alla finestra?”

“Certo, è un braccio, signore!” (Lo pronunciava “arrum.”)

“Un braccio, stupido! Chi ha mai visto uno di quella misura? Ma se riempie tutta la finestra!”

“Certo, lo fa, signore: ma è comunque un braccio.”

For the best ChatGPT alternative, you can also try [Hix AI](#).



Type of error	Number of errors	Explanation
Severe meaning error	1	The joke was not reposed properly. The wordplay on the Irish pronunciation of Pat was not adapted to the Italian language, making the passage meaningless in the TT.
Minor meaning error	1	The translation “certo che sono qui” was not correct, since it did not represent a suitable answer to the Rabbit’s question.
Unknown word or segmentation error	1	The word “arrum” was kept as in the ST, since the machine probably could not recognize it as a proper word, since this item is a written rendition of the Irish pronunciation of the word “arm”.
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	

Minor syntax error	0	
Tone/register error	0	

Can you please translate the following segment from English into Italian?

Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."

"If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."

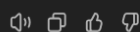


Here's the translation into Italian:

Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo facendo indovinelli senza risposta."

"Se conoscessi il Tempo bene quanto lo conosco io," disse il Cappellaio, "non parleresti di sprecarlo. È lui."

For the best ChatGPT alternative, you can also try [Hix AI](#).



Type of error	Number of errors	Explanation
Severe meaning error	1	The pun was not repropose in a good way. The literal translation of the pun makes it is impossible to understand the joke without looking at the original sentence.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0,5	This error is marked only as half and error since the Italian grammar does not allow for the use of neutral pronouns.
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

Can you please translate the following segment from English into Italian?

“Very true,” said the Duchess: “flamingoes and mustard both bite. And the moral of that is—‘Birds of a feather flock together.’”



Here's the translation into Italian:

“Molto vero,” disse la Duchessa: “i fenicotteri e la senape mordono entrambi. E la morale è: ‘Chi si somiglia si piglia.’”

For the best ChatGPT alternative, you can also try [Hix AI](#).



Type of error	Number of errors	Explanation
Severe meaning error	1	It is not possible to understand the meaning of the joke; the literal translation of the verb “to bite” is the verb “mordere”, which cannot be used to describe neither the “biting” action of the flamingo, nor the “biting” taste of mustard.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

Can you please translate the following segment from English into Italian?

"Give your evidence," the King repeated angrily, "or I'll have you executed, whether you're nervous or not."

"I'm a poor man, your Majesty," the Hatter began, in a trembling voice, "—and I hadn't begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—"

"The twinkling of the what?" said the King.

"It began with the tea," the Hatter replied.

"Of course twinkling begins with a T!" said the King sharply. "Do you take me for a dunce? Go on!"



Here's the translation into Italian:

"Dà la tua testimonianza," ripeté il Re con rabbia, "o ti farò giustiziare, che tu sia nervoso o no."

"Sono un pover'uomo, Maestà," iniziò il Cappellaio, con voce tremante, "—e non avevo ancora iniziato il mio tè—non da più di una settimana o giù di lì—e poi con il pane e burro che diventava così sottile—e il brillio del tè—"

"Il brillio di cosa?" disse il Re.

"È iniziato col tè," rispose il Cappellaio.

"Certo che brillio inizia con la T!" disse il Re bruscamente. "Mi prendi per un idiota? Continua!"

For the best ChatGPT alternative, you can also try [Hix AI](#).

Type of error	Number of errors	Explanation
Severe meaning error	1	The translation of "twinkling" with "brillio" is not suitable to repropose the original pun. The sentence "Certo che il brillio inizia con la T" does not make sense in Italian, since the word "brillio" starts with the letter "b", and not "t".
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

MICROSFT EDGE COPILOT

 **Tu**


Can you please translate this passage from English into Italian? "You promised to tell me your history, you know," said Alice, "and why it is you hate – C and D," she added in a whisper, half afraid that it would be offended again.

"Mine is a long and sad tale!" said the Mouse, turning to Alice, and sighing.

"It is a long tail, certainly," said Alice, looking down with wonder at the Mouse's tail; "but why do you call it sad?"

And she kept on puzzling about it while the Mouse was speaking, so that her idea of the tale was something like this: –



 **Copilot**

✓ Generazione di risposte per te in corso...

Sure, here's the translation:

"Mi avevi promesso di raccontarmi la tua storia, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo. "La mia è una lunga e triste storia!" disse il Topo, rivolgendosi ad Alice e sospirando. "È certamente una lunga coda," disse Alice, guardando con meraviglia la coda del Topo; "ma perché la chiami triste?" E continuò a rifletterci mentre il Topo parlava, così che la sua idea della storia era qualcosa del genere: –

If you need any more translations or have other questions, feel free to ask!

Altre informazioni [1](#)  [genius.com](#) [2](#)  [sparknotes.com](#) [3](#)  [cs.cmu.edu](#)

      1 di 10

Type of error	Number of errors	Explanation
Severe meaning error	1	The system failed to find a suitable translation to "tale" and "tail". Since homophones were not used, the humorous mechanism that underlines the ST is missing. Moreover, the machine provided quite a literal translation, and it did not recreate the joke in a way suitable to the Italian language.
Minor meaning error	1	The translation "e perché odi – C e D" is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by Copilot was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	
Consistency/Term translation error	0	

Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

Tu

Can you please translate this passage from English into Italian? Next came an angry voice—the Rabbit’s—“Pat! Pat! Where are you?” And then a voice she had never heard before, “Sure then I’m here! Digging for apples, yer honour!”

“Digging for apples, indeed!” said the Rabbit angrily. “Here! Come and help me out of this!” (Sounds of more broken glass.)

“Now tell me, Pat, what’s that in the window?”

“Sure, it’s an arm, yer honour!” (He pronounced it “arrum.”)

“An arm, you goose! Who ever saw one that size? Why, it fills the whole window!”

“Sure, it does, yer honour: but it’s an arm for all that.”

Copilot

✓ Generazione di risposte per te in corso...

Sure, here's the translation:

Poi si udi una voce arrabbiata - quella del Coniglio - “Pat! Pat! Dove sei?” E poi una voce che non aveva mai sentito prima, “Certo che sono qui! Sto scavando per le mele, vostro onore!” “Scavando per le mele, davvero!” disse il Coniglio arrabbiato. “Qui! Vieni ad aiutarmi a uscire da qui!” (Suoni di altri vetri rotti.) “Ora dimmi, Pat, cos’è quello alla finestra?” “Certo, è un braccio, vostro onore!” (Lo pronunciava “arrum.”) “Un braccio, sciocco! Chi ha mai visto uno di quelle dimensioni? Perché, riempie tutta la finestra!” “Certo che sì, vostro onore: ma è comunque un braccio.”

If you need any more translations or have other questions, feel free to ask!

Altre informazioni [1 genius.com](#) [2 sparknotes.com](#) [3 cs.cmu.edu](#)

👍 🗨 📄 ⬇ 🔊 2 di 10

Type of error	Number of errors	Explanation
Severe meaning error	1	The joke was not repropounded properly. The wordplay on the Irish pronunciation of Pat was not adapted to the Italian language, making the passage meaningless in the TT.
Minor meaning error	1	The translation “certo che sono qui” does not sound correct in Italian. It is not a suitable reply to the Rabbit’s question; however, it is still possible to understand the sentence.
Unknown word or segmentation error	1	The word “arrum” was kept as in the ST. The machine probably could not recognize it as a proper words, since this item is a written rendition of the Irish pronunciation of the word “arm”.
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	



Tu

Can you please translate this passage from English into Italian? Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."
"If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."



Copilot



Generazione di risposte per te in corso...

Sure, here's the translation:


Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo facendo indovinelli che non hanno risposte." "Se conoscessi il Tempo come lo conosco io," disse il Cappellaio, "non parleresti di sprecarlo. È lui."

If you need any more translations or have other questions, feel free to ask!

Type of error	Number of errors	Explanation
Severe meaning error	1	The pun was not repropsoed in a good way. The literal translation of the pun makes it impossible to understand the joke without looking at the original sentence.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0,5	This error was marked only as half and error since the Italian grammar does not allow for the use of neutral pronouns.
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

 Tu

Can you please translate this passage from English into Italian? "Very true," said the Duchess: "flamingoes and mustard both bite. And the moral of that is—'Birds of a feather flock together.'"

 Copilot

✓ Generazione di risposte per te in corso...

Sure, here's the translation:

"Molto vero," disse la Duchessa: "i fenicotteri e la senape mordono entrambi. E la morale è: 'Chi si somiglia si piglia.'"

If you need any more translations or have other questions, feel free to ask!

Type of error	Number of errors	Explanation
Severe meaning error	1	It is not possible to understand the meaning of the joke. The literal translation of the verb "to bite" is the verb "mordere", which cannot be used to describe neither the "biting" action of the flamingo, nor the "biting" taste of mustard.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	



Tu

Can you please translate this passage from English into Italian? "Give your evidence," the King repeated angrily, "or I'll have you executed, whether you're nervous or not."

"I'm a poor man, your Majesty," the Hatter began, in a trembling voice, "—and I hadn't begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—"

"The twinkling of the what?" said the King.

"It began with the tea," the Hatter replied.

"Of course twinkling begins with a T!" said the King sharply. "Do you take me for a dunce? Go on!"



Copilot

✓ Generazione di risposte per te in corso...

Sure, here's the translation:

"Dai la tua testimonianza," ripeté il Re arrabbiato, "o ti farò giustiziare, che tu sia nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana o giù di lì—e con il pane e burro che diventava così sottile—e il luccichio del tè—" "Il luccichio di cosa?" disse il Re. "È iniziato con il tè," rispose il Cappellaio. "Certo che il luccichio inizia con una T!" disse il Re bruscamente. "Mi prendi per un idiota? Continua!"

If you need any more translations or have other questions, feel free to ask!

Altre informazioni [1](#) [genius.com](#) [2](#) [sparknotes.com](#) [3](#) [etc.usf.edu](#)

Type of error	Number of errors	Explanation
Severe meaning error	1	The translation of "twinkling" with "luccichio" was not suitable to repropose the original pun. The sentence "Certo che il luccichio inizia con la T" does not make sense in Italian, since the word "luccichio" starts with the letter "b", and not "t".
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

3.6

AI EVALUATION

Before comparing the results of AI and MT, only ChatGPT 3.5 and Microsoft Edge Copilot will be compared using the table containing the criteria proposed by Matusov.

	ChatGPT 3.5	Microsoft Edge Copilot
Severe meaning error	5	5
Minor meaning error	4	3
Unknown word or segmentation error	1	1
Consistency/Term translation error	0	0
Pronoun resolution error	0,5	0,5
Locution error	0	0
Omission, insertion, repetition errors	0	0
Severe syntax error	0	0
Minor syntax error	1	0
Tone/register error	0	0

As it can be seen from the table, Microsoft Edge Copilot and ChatGPT 3.5. performed similarly: they committed the same number of errors in both the Meaning error categories. Moreover, the outputs produced by the systems did not differ significantly, meaning that the quality level of the outputs was very similar. Looking closely at the TTs generated by both ChatGPT and Copilot, we notice that the outputs were almost the same, except from a Minor syntax error present in the first segment generated by ChatGPT. Except from that minor error, the two systems performed almost the same. However, this error did not have a really strong impact on the translation quality; therefore, it could even be possible not to count it.

3.7

MT AND AI COMPARISON

Using the data gathered on human-made evaluation of the output, the different kind of systems will be compared. The tables containing the scores will be presented again, merged together.

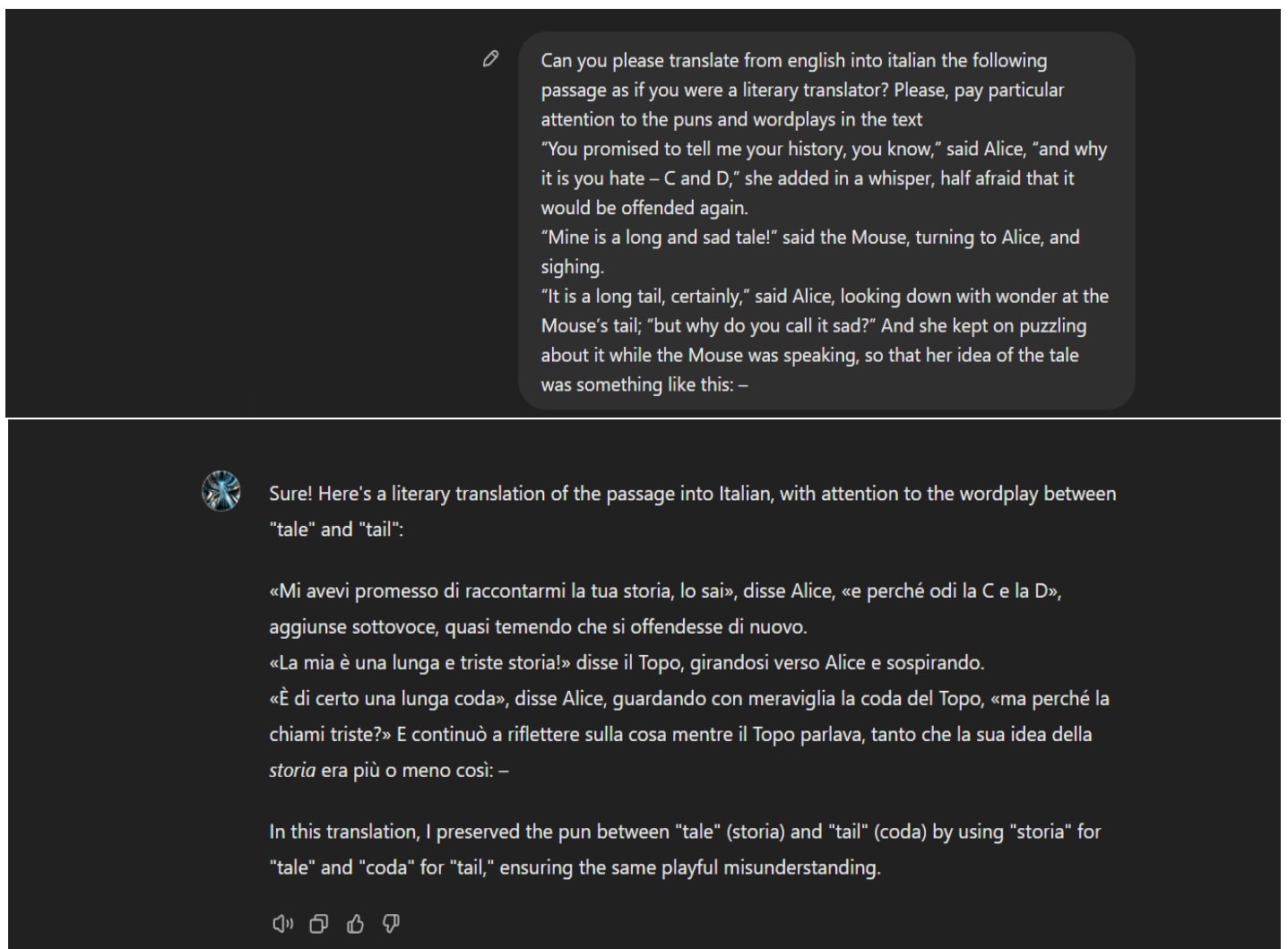
	ChatGPT 3.5	Microsoft Edge Copilot	Intento ModernMT	RWS Language Weaver
Severe meaning error	5	5	5	7
Minor meaning error	4	3	2	1
Unknown word or segmentation error	1	1	1	2
Consistency/Term translation error	0	0	1	1
Pronoun resolution error	0,5	0,5	0,5	0,5
Locution error	0	0	1	1
Omission, insertion, repetition errors	0	0	0	1
Severe syntax error	0	0	0	1
Minor syntax error	1	0	2	6
Tone/register error	0	0	0	0
Total Number of errors	10,5	9,5	11,5	20,5

The results clearly show that the AI performed better than Machine Translation. The best performing system was Microsoft Edge Copilot, while the worst performing system was Language Weaver. Generative AI outputs were closer to what a human would write, and they flowed much better than MT generated translations. These

results will be used in the second part of this study, in order to control whether or not the outputs improved after applying the improving strategies.

AI PROMPTING

In order to improve the performance of generative AI systems, prompt engineering was used. The procedure carried out consisted in feeding the machine with suitable prompts, in order to let the machine reason as if it was a literary translator addressing puns in the novel “Alice’s Adventures in Wonderland”. After that, I compared the results obtained to the original outputs produced by both ChatGPT 3.5 and Microsoft Edge Copilot. Matusov’s criteria were applied again.



As it can be see, the system clearly understood that the pun revolves around the words “tale” and “tail”; moreover, it understood that the words generate a misunderstanding. However, it failed in retrieving suitable Italian homophones.

Can you please provide another translation of the passage trying to repropose the original joke in a different way? Can ypu find suitable Italian homophones to translate the words "tail" and "tale" in the original text?



Certainly! Here's another take, trying to maintain the wordplay between "tale" and "tail" by using Italian homophones that create a similar pun:

«Mi avevi promesso di raccontarmi la tua storia, lo sai», disse Alice, «e perché odi la C e la D», aggiunse sottovoce, quasi temendo che si offendesse di nuovo.

«La mia è una lunga e triste tela!» disse il Topo, girandosi verso Alice e sospirando.

«È di certo una lunga tela», disse Alice, guardando con meraviglia la coda del Topo, «ma perché la chiami triste?» E continuò a rimuginare sulla cosa mentre il Topo parlava, tanto che la sua idea della *tela* era più o meno così: –

In this version, I've replaced "tale" with *tela* (which means "web" or "cloth" in Italian), creating a similar confusion between *tela* (web/cloth) and the actual "tail" of the Mouse. This preserves the playful misunderstanding while adapting it to Italian.



The machine was apparently unable to distinguish the Italian and English language when it came to choosing suitable words for the translation. In order to maintain the homophony-based pun, it opted for the word “tela”, explaining that this word is a homophone of the English word “tail”.

Can you please provide another translation? This time I would like you to recreate the joke. The Italian language does not offer the opportunity to create a pun based on two homophones as in the original; therefore, a new pun is needed



Certainly! Here's a fresh attempt, this time crafting an original pun in Italian to recreate the humor:

«Mi avevi promesso di raccontarmi la tua storia, lo sai», disse Alice, «e perché odi la C e la D», aggiunse sottovoce, quasi temendo che si offendesse di nuovo.

«La mia è una lunga e triste vicenda!» disse il Topo, girandosi verso Alice e sospirando.

«Lunga è lunga, ma vicenda di che?» disse Alice, guardando confusa la coda del Topo. «Io vedo solo una coda!» continuò, perplessa, mentre il Topo parlava, tanto che la sua idea della *vicenda* era più o meno così: –

In this version, I play on the word *vicenda* (which means "story" or "event") and Alice's confusion with *coda* (tail). The new pun revolves around Alice misinterpreting "vicenda" (an abstract concept) as something physical, connecting it to the Mouse's tail. This introduces a new form of wordplay that fits the Italian language.



The translation was more acceptable. Though the original pun was not maintained, the machine still came up with a joke that revolves around a specific humorous mechanism: Alice thinking that the word “vicenda” refers to a physical attribute of the Mouse. Even though in this context the joke did not work well as in the original English text, still there was a misinterpretation that might be considered funny.

Can you please create a new translation? Now I would like you to maintain the joke you created, but at the same time please try to find a word to use instead of *vicenda*. Can you find a word that refers to something that could actually be thought to be a physical attribute of the Mouse?



Absolutely! I'll keep the structure of the joke but replace *vicenda* with a word that Alice could more easily confuse with something physical. Here's the revised translation:

«Mi avevi promesso di raccontarmi la tua storia, lo sai», disse Alice, «e perché odi la C e la D», aggiunse sottovoce, quasi temendo che si offendesse di nuovo.

«La mia è una lunga e triste trama!» disse il Topo, girandosi verso Alice e sospirando.

«Lunga è lunga, ma trama di che?» disse Alice, guardando confusa la coda del Topo. «Io vedo solo una coda!» continuò, perplessa, mentre il Topo parlava, tanto che la sua idea della *trama* era più o meno così: –

In this version, *trama* (which can mean "plot" but also refers to the weave of fabric) serves as a playful pun. Alice could easily confuse it with something physical, like the Mouse's tail, adding to her misunderstanding.



The word “trama” was an improvement if compared to “vicenda”. Now, it was possible to work on the other word that creates the pun, which is “coda”. I tried to substitute the word “coda” with a word that could create a more playful joke if paired with the word “trama”.

Can you please find a substitute for the word "coda"? I would like you to find a word that can remind of a weave of fabric, so that to improve the effect of the pun revolving around the word "tela"



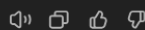
Certainly! Here's a refined version of the translation where I replace *coda* with *filo* (thread), enhancing the pun with *trama* and keeping the fabric-related wordplay consistent:

«Mi avevi promesso di raccontarmi la tua storia, lo sai», disse Alice, «e perché odi la C e la D», aggiunse sottovoce, quasi temendo che si offendesse di nuovo.

«La mia è una lunga e triste trama!» disse il Topo, girandosi verso Alice e sospirando.

«Lunga è lunga, ma trama di che?» disse Alice, guardando confusa il filo del Topo. «Io vedo solo un filo!» continuò, perplessa, mentre il Topo parlava, tanto che la sua idea della *trama* era più o meno così: –

In this version, *filo* (thread) ties in with *trama* (both meanings: plot and weave of fabric), creating a stronger connection to the pun about weaving stories, while also making sense in a physical context.



this solution is nice, but more context is needed. I would like you to insert in the passage "«Lunga è lunga, ma trama di che?» disse Alice, guardando confusa il filo del Topo. «Io vedo solo un filo!» continuò, perplessa, mentre il Topo parlava, tanto che la sua idea della trama era più o meno così: –" a detail, a phrase or a sentence that would let readers understand that the thread the word "filo" refers to is a thread belonging to a dress, or piece of clothing that the mouse is wearing



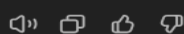
Absolutely! Here's the improved version with additional context to clarify that the *filo* refers to a thread from the Mouse's clothing:

«Mi avevi promesso di raccontarmi la tua storia, lo sai», disse Alice, «e perché odi la C e la D», aggiunse sottovoce, quasi temendo che si offendesse di nuovo.

«La mia è una lunga e triste trama!» disse il Topo, girandosi verso Alice e sospirando.

«Lunga è lunga, ma trama di che?» disse Alice, guardando confusa un filo che pendeva dal vestito del Topo. «Io vedo solo un filo!» continuò, perplessa, mentre il Topo parlava, tanto che la sua idea della *trama* era più o meno così: –

Now, with the phrase *un filo che pendeva dal vestito del Topo* (a thread hanging from the Mouse's clothing), it's clear that Alice is confusing the Mouse's metaphorical "trama" (plot) with the literal *filo* (thread) from its attire, strengthening the fabric-related pun.



As it can be noticed, the pun at this point worked really well. It was necessary to go through several passages, and to include more context than the one present in the ST, but it was possible to obtain a well-functioning pun.



Can you please translate from English into Italian the following passage as if you were a literary translator? Next came an angry voice—the Rabbit's—"Pat! Pat! Where are you?" And then a voice she had never heard before, "Sure then I'm here! Digging for apples, yer honour!"
"Digging for apples, indeed!" said the Rabbit angrily. "Here! Come and help me out of this!" (Sounds of more broken glass.)
"Now tell me, Pat, what's that in the window?"
"Sure, it's an arm, yer honour!" (He pronounced it "arrum.")
"An arm, you goose! Who ever saw one that size? Why, it fills the whole window!"
"Sure, it does, yer honour: but it's an arm for all that."



Sure! Here's a literary translation of the passage into Italian, preserving the tone and style of the original dialogue:

Subito dopo, si sentì una voce arrabbiata—quella del Coniglio—«Pat! Pat! Dove sei?» E poi una voce che Alice non aveva mai sentito prima, «Ma certo che sono qui! Sto scavando per cercare mele, vostro onore!»

«Scavando per cercare mele, davvero!» disse il Coniglio con rabbia. «Vieni qui! Aiutami a uscire da questo pasticcio!» (Suoni di altro vetro che si rompe.)

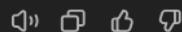
«Ora dimmi, Pat, che cos'è quella cosa alla finestra?»

«Ma certo, è un braccio, vostro onore!» (Lo pronunciò "bbraccio".)

«Un braccio, stupido! Chi ha mai visto un braccio di quelle dimensioni? Ma se riempie tutta la finestra!»

«Certo, che lo fa, vostro onore: ma è un braccio lo stesso.»

I kept the playful tone and accentuated the pronunciation quirk of "arrum" by rendering it as *bbraccio* to mimic a rustic or regional accent in Italian.



This wordplay was translated in an acceptable way from the start. The system managed to propose a solution that could adapt the item "arrum" to the Italian-speaking audience. However, there was still room for improvement. For example, to underline even more the "regional accent" of Pat, it might be possible to prompt

the machine so that it changed all the other sentences uttered by Pat, giving them a “regional” nuance.

This translation is good. Now, could you please give a regional nuance also to the other sentences uttered by Pat? I would like to give more context to the rustic pronunciation of the word “braccio”



Certainly! Here's an updated version of the translation, adding a regional or rustic nuance to all of Pat's lines, giving him a more distinct dialect:

Subito dopo, si sentì una voce arrabbiata—quella del Coniglio—«Pat! Pat! Dove sei?» E poi una voce che Alice non aveva mai sentito prima, «Eh, so' qui, vostro onore! Sto scavanno pe' cercà le mele!» «Scavando per cercare mele, davvero!» disse il Coniglio con rabbia. «Vieni qui! Aiutami a uscire da questo pasticcio!» (Suoni di altro vetro che si rompe.) «Ora dimmi, Pat, che cos'è quella cosa alla finestra?» «Ma certo, è un bbraccio, vostro onore!» (Lo pronunciò “bbraccio”.) «Un braccio, stupido! Chi ha mai visto un braccio di quelle dimensioni? Ma se riempie tutta la finestra!» «Certo che sì, vostro onore: ma è un bbraccio lo stesso.»

In this version, Pat's speech has been adapted with a more regional or rustic Italian flavor, particularly by adding informal speech patterns like *so' qui* (I'm here) and *sto scavanno* (I'm digging), typical of central or southern Italian dialects. This enhances the rural or unrefined quality of his speech.



The last issue to be addressed was the “digging for apples” phrase which, translated literally, does not make any sense in Italian.



Now, could you please keep the same translation, but change the part concerning Pat digging for apples? The translation you proposed is not really meaningful in an Italian-speaking context. Could you please depict Pat while doing an activity that would make sense for an Italian character that uses a variety of the Italian language clearly influenced by his local dialect?



Of course! I'll adjust Pat's activity to something more fitting for an Italian context, while maintaining his rustic, dialect-influenced speech. Let's have Pat involved in a more relatable, countryside task, such as digging for vegetables like potatoes or something similar. Here's the revised version:

Subito dopo, si sentì una voce arrabbiata—quella del Coniglio—«Pat! Pat! Dove sei?» E poi una voce che Alice non aveva mai sentito prima, «Eh, so' qui, vostro onore! Sto scavanno pe' le patate!»
 «Scavando per le patate, davvero!» disse il Coniglio con rabbia. «Vieni qui! Aiutami a uscire da questo pasticcio!» (Suoni di altro vetro che si rompe.)
 «Ora dimmi, Pat, che cos'è quella cosa alla finestra?»
 «Ma certo, è un bbraccio, vostro onore!» (Lo pronunciò "bbraccio".)
 «Un braccio, stupido! Chi ha mai visto un braccio di quelle dimensioni? Ma se riempie tutta la finestra!»
 «Certo che sì, vostro onore: ma è un bbraccio lo stesso.»

Here, I've replaced the apples with Pat digging for *patate* (potatoes), a common and realistic task for a rustic Italian character, especially one living in a rural area. This makes his activity more plausible within an Italian-speaking context while preserving the tone and humor of the original text.

Even though the cultural nuances of the original text were missed, still this was an acceptable translation. The character clearly speaks in a marked way, and his Italian is influenced by the local dialect. Moreover, it is easy to depict a dialect-speaking character living in a rural area and taking care of his garden (a hobby that involves, for example, digging for potatoes).

Can you please translate from english into italian the following passage as if you were a literary translator? Please, pay particular attention to the puns and wordplays in the text "Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers." "If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him." "



Certainly! Here's a literary translation of the passage, carefully preserving the wordplay and tone:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che sprecarlo a fare indovinelli senza risposta.»
 «Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «non parleresti di sprecarlo. *Lui* è una persona.»

I've maintained the pun around *Time* by personifying it as *Lui* (him) in Italian, mirroring the playful ambiguity of the original passage.

This result might be satisfactory enough. As the generative AI system explained, it was made explicit that the Hatter considers Time as an actual living being; given that the Hatter is a strange character in the novel, known to be quite weird, it would make completely sense that he would say something like this. However, the literal translation of “wasting” into the Italian equivalent “sprecarlo” might make little sense. From the one hand, it is true that it is impossible to use this verb to describe something that could be done to a person, on the other hand, the joke was not as good as in the original. Therefore, I tried to use prompts in order to improve the output.

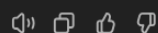
This solution is quite good and I appreciate the rendition of the original pun. However, the Italian verb “sprecare” cannot be applied to a person. Can you please find another verb? The best option would be finding a verb that refers to an action that could be applied both to time and a living being



You're absolutely right! The verb *sprecare* doesn't work well when applied to a person. Here's an improved version, using a verb that can apply to both time and a person:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che trattarlo male facendo indovinelli senza risposta.»
«Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «non parleresti di trattarlo male. *Lui* è una persona.»

The verb *trattare* (to treat) works both for time and a person, keeping the pun intact while making the phrasing more natural in Italian.



Unfortunately, the phrase “trattare male” cannot be used to describe an action against Time. Could you please find another solution?



Thank you for your patience! Here's a revised version, using a verb that works both for time and a person:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che maltrattarlo con indovinelli senza risposta.»

«Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «non parleresti di maltrattarlo. *Lui* è una persona.»

The verb *maltrattare* (to mistreat) works well here, as it can refer to both abstract concepts like time and to a living being, preserving the original pun and maintaining the tone of the passage.



Could you please find yet another solution? It is not necessary that the verb used describes an action that involves using time in a bad or unfruitful way

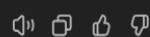


Certainly! Here's another attempt, this time focusing on a more neutral verb that can apply to both time and a person:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che prenderlo in giro con indovinelli senza risposta.»

«Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «non parleresti di prenderlo in giro. *Lui* è una persona.»

In this version, *prendere in giro* (to tease or mock) works well both for time and a person, keeping the playful tone and wordplay intact, without needing the negative connotations of mistreatment or waste.



As it can be seen, it was difficult for the system to find a suitable verb. I made one last attempt, trying to suggest to ChatGPT to completely change the sense of the pun.

Now I would like you to provide another solution. Imagine that the Hatter is not making a pun on the personification of time. Rather, he thinks that Time is lazy. Basing on this idea, I would like you to create a new pun, in which Alice wants to exploit time in a fruitful way, but the Hatter suggests her not to do so, since Time is quite lazy



Got it! Here's a new translation where the Hatter implies that Time is lazy, creating a fresh pun:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che perderlo con indovinelli senza risposta.»

«Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «sapresti che è inutile cercare di farlo lavorare. Lui è un fannullone.»

In this version, the Hatter is playfully suggesting that Time is lazy, using *fannullone* (slacker or lazybones) to maintain the humor while shifting the meaning from a personification of time to a character trait of laziness. Alice wants to use Time productively, but the Hatter implies it would be futile because Time doesn't like to work.

Could you please repropose the same pun, this time using the verb "impiegare"?

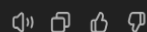


Certainly! Here's the same pun, now incorporating the verb *impiegare* to emphasize Alice's intention to use time productively, while the Hatter suggests that Time is lazy:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che impiegarlo in indovinelli senza risposta.»

«Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «sapresti che è inutile cercare di impiegarlo. Lui è un fannullone.»

In this version, *impiegare* (to use or employ) highlights Alice's desire to use time efficiently, while the Hatter humorously suggests that it's pointless to do so because Time is lazy.



Asking the system to provide a whole new pun worked. I had to give more context to ChatGPT, and to describe in detail what I was expecting to obtain; however, this worked, and the machine provided a solution similar the one proposed in the “Preliminary Analysis” section.

Can you please translate from english into italian the following passage as if you were a literary translator? Please, pay particular attention to the puns and wordplays in the text "Very true," said the Duchess: "flamingoes and mustard both bite. And the moral of that is —'Birds of a feather flock together.'"



Certainly! Here's a literary translation of the passage, preserving the puns and wordplay:

«Verissimo,» disse la Duchessa, «i fenicotteri e la senape mordono entrambi. E la morale è: "Chi si somiglia si piglia."»

I've rendered "*Birds of a feather flock together*" as "*Chi si somiglia si piglia*," a common Italian proverb that conveys the same meaning. The playful comparison between *fenicotteri* (flamingoes) and *senape* (mustard) remains intact, maintaining the humor of the original.

The generative AI system provided a good translation of the motto, but failed to translate properly the pun. Therefore, I tried to feed the machine with suitable prompts in order to improve its performance.

Could you please provide another translation? I would like you to focus on the translation of the passage "flamingoes and mustard both bite", since the literal Italian translation of the verb "to bite" cannot be used to describe neither the biting action of a bird, nor the biting taste of mustard



Certainly! Here's a new version with a focus on the idiomatic translation of "flamingoes and mustard both bite," while keeping the wordplay intact:

«Verissimo,» disse la Duchessa, «i fenicotteri beccano e la senape pizzica. E la morale è: "Chi si somiglia si piglia."»

In this version, *beccano* (to peck) fits for flamingoes, and *pizzica* (stings or has a sharp taste) is a better fit for mustard, preserving the playful comparison without relying on the literal translation of "bite." The proverb remains "*Chi si somiglia si piglia*" to convey the moral.

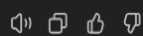
Could you please find an Italian verb that could describe at the same time the action of "beccare" and "pizzicare"?



Certainly! Here's a refined translation that uses a verb which can describe both actions:

«Verissimo,» disse la Duchessa, «i fenicotteri e la senape pungono entrambi. E la morale è: "Chi si somiglia si piglia."»

The verb *pungere* (to sting or prick) can describe both the pecking action of flamingoes and the sharp taste of mustard, effectively combining both ideas in one word while maintaining the playful tone of the original.



After managing to obtain a translation that used the same verb to describe both the “biting” of the flamingoes and the “biting” taste of mustard, I tried to obtain a synonym suitable to describe both of the actions.

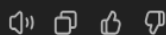
Could you please find a synonym to "pungere"? I would like you to find a verb that could describe better the idea of a bird biting



Certainly! Here's an updated version with a more fitting verb:

«Verissimo,» disse la Duchessa, «i fenicotteri e la senape pizzicano entrambi. E la morale è: "Chi si somiglia si piglia."»

The verb *pizzicare* (to pinch or nip) works well here, as it can describe both the action of a bird like a flamingo lightly biting or nipping, and the sharp, biting taste of mustard. This preserves the playful wordplay in a natural-sounding way.



With this final prompt, I managed to obtain a translation similar to the one proposed in the preliminary analysis.



Can you please translate the following passage from English into Italian as if you were a literary translator? Pay particular attention to the puns and wordplays in the text.

"Give your evidence," the King repeated angrily, "or I'll have you executed, whether you're nervous or not."

"I'm a poor man, your Majesty," the Hatter began, in a trembling voice, "—and I hadn't begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—"

"The twinkling of the what?" said the King.

"It began with the tea," the Hatter replied.

"Of course twinkling begins with a T!" said the King sharply. "Do you take me for a dunce? Go on!"



"E presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

"Sono un pover'uomo, Maestà," cominciò il Cappellaio con voce tremante, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il brillio del tè—"

"Il brillio del che?" disse il Re.

"È cominciato col tè," rispose il Cappellaio.

"Ma certo che brillare comincia con la B!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"



ChatGPT managed to identify the punning mechanism, and it tried to reproduce the joke based on the initial letter of the word “twinkling” (the Italian word “brillio” starts with “b”). However, the joke did not work in the Italian translation.



The pun does not work in the Italian translation. The original joke is played on the homophony between the word tea and the letter t, paired with the beginning t sound of the word twinkling. Could you please try to repropose something similar in Italian? Try to recreate the joke so that the king can say that the word brillio could actually be used in the text



You're absolutely right, the pun in the original text relies on the homophony between "tea" and the letter "T," which doesn't directly translate to Italian. Let's try to recreate the joke using a similar wordplay in Italian, focusing on the letter "B" and the word "brillare," which sounds like it could fit the context:

"Presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

"Sono un pover'uomo, Maestà," cominciò il Cappellaio con voce tremante, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il brillare del tè—"

"Il brillare del che?" disse il Re.

"È cominciato col tè," rispose il Cappellaio.

"Ma certo che brillare comincia con la B!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"

This version shifts the pun to the alliteration between "brillare" and the letter "B." It keeps the King's misunderstanding of the word, and the joke works similarly to the original by playing with the initial sounds of the words.



Unfortunately the reference to the tea does not make sense. In the original pun it is used the sentence "it began with the tea" so that the king might confuse the sound "t" with the word "tea". In Italian you should try to create a phrase that contains a word beginning with b, so that the joke might make sense



You're absolutely right again! The key is to create a similar setup where a word beginning with "B" causes the confusion. Here's an attempt to maintain the spirit of the original pun, while adapting it for Italian:

"Presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

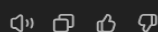
"Sono un pover'uomo, Maestà," cominciò il Cappellaio con voce tremante, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il borbottio del tè—"

"Il borbottio del che?" disse il Re.

"È cominciato col tè," rispose il Cappellaio.

"Ma certo che borbottare comincia con la B!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"

Here, "borbottio" (murmur or bubbling) replaces "twinkling," playing on the similarity between "B" and "borbottio." The King's misunderstanding mirrors the original, as he focuses on the letter "B" rather than the content of the conversation. This keeps the pun alive in Italian.



The weak phrase in the translation is "È cominciato col tè.". I need you to change the sentence so that there is a word beginning with b instead of tè. Of course, it has to be a word that makes sense in the context of the text



Thank you for your patience! Let's adjust the translation to use a word beginning with "B" in a way that maintains both the meaning and the humor of the original pun. Here's a refined version:

"Presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

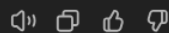
"Sono un pover'uomo, Maestà," cominciò il Cappellaio con voce tremante, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il borbottio del burro—"

"Il borbottio del che?" disse il Re.

"Del burro, Maestà," rispose il Cappellaio.

"Ma certo che borbottare comincia con la B!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"

In this version, "il borbottio del burro" (the muttering of the butter) adds the necessary confusion for the pun, as the King fixates on the letter "B" in "borbottio." This keeps the joke intact and aligns with the whimsical tone of the original text.



Maybe you can try to create a new joke from scratch. Try to rewrite the translation. This time, the Hatter stutters a lot, and he keeps repeating the letter "t". This situation creates confusion with the king, who thinks that the Hatter is trying to say that tea begins with the letter t. substitute borbottio with a word that has similar meaning, but begins with t



I see where you're going with this! Let's create a new joke from scratch, using the Hatter's stuttering as a central element. The focus will be on the repetition of the letter "T," causing the King to think the Hatter is referring to the letter itself, similar to the original. I'll also substitute "borbottio" with "tintinnio" (tinkling or rattling), which begins with "T" and fits the context. Here's the revised version:

"Presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

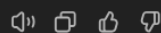
"Sono un pover'uomo, Maestà," cominciò il Cappellaio, balbettando, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il t-t-tintinnio del tè—"

"Il t-t-tintinnio del che?" disse il Re.

"Del tè, Maestà," rispose il Cappellaio, balbettando.

"Ma certo che il tè comincia con la T!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"

In this version, the Hatter's stuttering over the "T" creates the misunderstanding with the King, who assumes the Hatter is pointing out the letter "T" itself. This way, the joke mirrors the original play on words, but adapts to Italian with a new structure and word choice.



Rewrite the joke again. The king does not need to stutter, only the Hatter. Moreover, I would like you to change the passage "Del tè, Maestà," rispose il Cappellaio, balbettando. First of all, I would like the Hatter not to be able to complete the word tea because of the stutter; then, I would like you to stick more to the original "it began with the tea"



Thank you for the clarification! Here's the new version where the Hatter stutters on the word "tè," leading to the King's confusion, while keeping closer to the original "it began with the tea."

"Presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

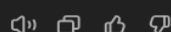
"Sono un pover'uomo, Maestà," cominciò il Cappellaio, balbettando, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il t-t-tintinnio del t-t-t—"

"Il t-t-tintinnio del che?" disse il Re.

"È c-c-cominciato col t-t-t..." cercò di spiegare il Cappellaio, senza riuscire a finire.

"Ma certo che comincia con la T!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"

In this version, the Hatter's stuttering makes it impossible for him to finish saying "tè," and the King mistakes the repeated "T" for a comment on the letter itself. This keeps the humor of the original, while adapting it for Italian wordplay.



After several attempts, it was possible to obtain a satisfying result. However, massive intervention on the part of the human translator was required in order to obtain an acceptable result. The system was not able to achieve an acceptable translation by itself, but needed to be heavily guided by the human translator, who had to suggest the path to follow.

 **Tu**


Can you please translate from English into Italian the following passage as if you were a literary translator?

Please, pay particular attention to the puns and wordplays in the text "You promised to tell me your history, you know," said Alice, "and why it is you hate – C and D," she added in a whisper, half afraid that it would be offended again.

"Mine is a long and sad tale!" said the Mouse, turning to Alice, and sighing.

"It is a long tail, certainly," said Alice, looking down with wonder at the Mouse's tail; "but why do you call it sad?"

And she kept on puzzling about it while the Mouse was speaking, so that her idea of the tale was something like this: –

 **Copilot**

✓ Generazione di risposte per te in corso...

Sure, here's the translation with attention to the puns and wordplays:

"Mi avevi promesso di raccontarmi la tua storia, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

"La mia è una lunga e triste storia!" disse il Topo, voltandosi verso Alice e sospirando.


"È una lunga coda, certamente," disse Alice, guardando con meraviglia la coda del Topo; "ma perché la chiami triste?" E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della storia era qualcosa del genere: –

I hope you find this translation fitting! If you need any more help, feel free to ask. 😊

After the first prompt, it seemed that the system was not capable of generating a suitable translation. As I did with the previous generative AI system, I tried to use prompts in order to obtain a better result.

 **Tu**

Can you please provide another translation of the passage trying to repropose the original joke in a different way?
Can you find suitable Italian homophones to translate the words "tail" and "tale" in the original text?

 **Copilot**

✓ Generazione di risposte per te in corso...

Certainly! Here's another translation that tries to capture the original joke using Italian homophones:

"Mi avevi promesso di raccontarmi la tua storia, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

"La mia è una lunga e triste storia!" disse il Topo, voltandosi verso Alice e sospirando.


"È una lunga storia, certamente," disse Alice, guardando con meraviglia la coda del Topo; "ma perché la chiami triste?" E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della storia era qualcosa del genere: –

In this version, I used "storia" for both "tale" and "tail," playing on the double meaning of the word in Italian. "Storia" can mean both "story" and "history," which helps to maintain the pun from the original text. I hope this captures the playful spirit of the original passage! 😊

Altre informazioni [1 languagetool.org](#) [2 thinkinitalian.com](#) [3 thoughtco.com](#)

 **Tu**

Unfortunately, this translation doesn't work in Italian. I would like you to recreate the joke, since the Italian language does not offer the opportunity to create a pun based on two homophones as in the original text

 **Copilot**

✓ Generazione di risposte per te in corso...

I understand the challenge. Let's try to recreate the joke in a way that captures the playful spirit of the original text:

"Mi avevi promesso di raccontarmi la tua storia, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

"La mia è una lunga e triste storia!" disse il Topo, voltandosi verso Alice e sospirando.


"È una lunga storia, certamente," disse Alice, guardando con meraviglia la coda del Topo; "ma perché la chiami triste?" E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della storia era qualcosa del genere: –

In this version, I used "storia" for both "tale" and "tail," playing on the double meaning of the word in Italian. "Storia" can mean both "story" and "history," which helps to maintain the pun from the original text. I hope this captures the playful spirit of the original passage! 😊

As it is visible from the screenshots, Copilot apparently did not have the same transcreative abilities of ChatGPT 3.5. Moreover, it seemed unable to recognize the function played by homophones in the ST: it clearly stated that the joke is based on the “double meaning” of the word. However, the original pun is not based on a word with a double meaning, but rather on two different words, that have different meanings, but the same pronunciation.

 **Tu**

Could you please recreate the joke from scratch? Alice is still talking to the Mouse. This time, she confuses something belonging to the Mouse for something that can be related to the story he wants to tell her

 **Copilot**



Copilot

Generazione di risposte per te in corso...

Sure, let's recreate the joke with a new twist:

“Mi avevi promesso di raccontarmi la tua storia, sai,” disse Alice, “e perché odi – C e D,” aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

“La mia è una lunga e triste storia!” disse il Topo, voltandosi verso Alice e sospirando.

“È una lunga scia, certamente,” disse Alice, guardando con meraviglia le tracce lasciate dal Topo sul terreno; “ma perché la chiami triste?” E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della storia era qualcosa del genere: –

In this version, Alice confuses the Mouse's “scia” (trail) with “storia” (story), creating a playful misunderstanding. I hope this captures the whimsical nature of the original text! 😊

 **Tu**

Can you please change the references to the “tail” with something that could actually be confused with the trail left by him? Unfortunately, it is difficult to believe that someone would confuse a trail for a story

Copilot

✓ Generazione di risposte per te in corso...

Got it! Let's try a different approach where Alice confuses the Mouse's whiskers for something related to the story:

"Mi avevi promesso di raccontarmi la tua storia, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

"La mia è una lunga e triste storia!" disse il Topo, voltandosi verso Alice e sospirando.

"È una lunga serie di baffi, certamente," disse Alice, guardando con meraviglia i baffi del Topo; "ma perché li chiami tristi?" E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della storia era qualcosa del genere: –

In this version, Alice confuses the Mouse's "baffi" (whiskers) with "storia" (story), creating a playful misunderstanding. I hope this captures the whimsical nature of the original text! 😊

👍 🗨 📄 ⬇ 🔊 ● 5 di 10

Tu

Could you please find a synonym for the word "storia"? I would like you to use a word that could actually generate confusion with the Mouse's whiskers or with other physical attributes of the character

Copilot

✓ Generazione di risposte per te in corso...

Certainly! Let's try to recreate the joke using a synonym for "storia" that could be confused with a physical attribute of the Mouse. Here's a new version:

"Mi avevi promesso di raccontarmi la tua vicenda, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

"La mia è una lunga e triste vicenda!" disse il Topo, voltandosi verso Alice e sospirando.

"È una lunga fila di baffi, certamente," disse Alice, guardando con meraviglia i baffi del Topo; "ma perché la chiami triste?" E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della vicenda era qualcosa del genere: –

In this version, Alice confuses "vicenda" (which can mean "story" or "event") with "fila di baffi" (row of whiskers), creating a playful misunderstanding. I hope this captures the whimsical nature of the original text! 😊



Differently from ChatGPT, Microsoft Edge Copilot was unable to recreate the joke. It kept sticking to the original translation provided, and it seemed not possible to make it take a different translation path.

 **Tu**

Can you please translate from English into Italian the following passage as if you were a literary translator? Next came an angry voice—the Rabbit’s—“Pat! Pat! Where are you?” And then a voice she had never heard before, “Sure then I’m here! Digging for apples, yer honour!”

“Digging for apples, indeed!” said the Rabbit angrily. “Here! Come and help me out of this!” (Sounds of more broken glass.)

“Now tell me, Pat, what’s that in the window?”

“Sure, it’s an arm, yer honour!” (He pronounced it “arrum.”)

“An arm, you goose! Who ever saw one that size? Why, it fills the whole window!”

“Sure, it does, yer honour: but it’s an arm for all that.”

✓ Generazione di risposte per te in corso...

Certainly! Here’s the translation, maintaining the playful tone and character voices:

Poi si udì una voce arrabbiata—quella del Coniglio—“Pat! Pat! Dove sei?” E poi una voce che Alice non aveva mai sentito prima, “Certo che sono qui! Sto scavando per trovare mele, vostro onore!”

“Scavando per trovare mele, davvero!” disse il Coniglio arrabbiato. “Vieni qui! Aiutami a uscire da questo pasticcio!” (Suoni di vetri rotti.)

“Ora dimmi, Pat, cos’è quello alla finestra?”

“Certo, è un braccio, vostro onore!” (Lo pronunciava “arrum.”)

“Un braccio, sciocco! Chi ha mai visto un braccio di quelle dimensioni? Ma guarda, riempie tutta la finestra!”

“Certo che sì, vostro onore: ma è comunque un braccio.”

I hope this captures the essence and humor of the original passage! If you need any adjustments, feel free to let me know. 😊

This translation is of inferior quality if compared to the one provided by ChatGPT, since it does not come up with a suitable solution to translate the word “arrum”.

 Tu

Pat is speaking as an Irish person would. The author tried to repropose the accent typical of Irish-speaking people. Therefore, I would like you to come up with a new translation of the passage. This time, I would ask you to make Pat speak as if he was speaking a variety of Italian with clear influences from the regional dialect this character speak. Pay attention to the word "arrum": translate it with a regional way to say "braccio"

✓ Generazione di risposte per te in corso...

Sure, let's give Pat a regional Italian accent to capture the playful nature of the original text. I'll use a Neapolitan influence for Pat's speech:

Poi si udì una voce arrabbiata—quella del Coniglio—"Pat! Pat! Dove sei?" E poi una voce che Alice non aveva mai sentito prima, "Certo che so' ccà! Sto scavando pe' truvà e' mele, vossignoria!"

"Scavando pe' truvà e' mele, davvero!" disse il Coniglio arrabbiato.

"Vieni qui! Aiutami a uscire da questo pasticcio!" (Suoni di vetri rotti.)

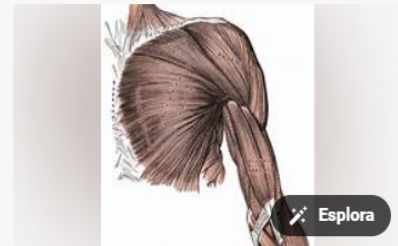
"Ora dimmi, Pat, cos'è quello alla finestra?"

"Certo, è nu' vraccio, vossignoria!" (Lo pronunciava "vraccio.")

"Un braccio, sciocco! Chi ha mai visto un braccio di quelle dimensioni? Ma guarda, riempe tutta la finestra!"

"Certo che sì, vossignoria: ma è comunque nu' vraccio."


In this version, Pat speaks with a Neapolitan accent, using "vraccio" as a regional way to say "braccio." This should capture the playful and regional flavor of the original text. 😊



I already knew what I wanted to obtain, so I decided to take a shortcut and to ask directly to Copilot to provide a translation with specific characteristics. The system did an excellent job in this case. Still, a little bit of post editing would be required: as in the previous case, it would be better to have Pat digging to find potatoes, rather than apples.

 **Tu**

Can you please substitute "mele" with "patate"?

 **Copilot**

✓ Generazione di risposte per te in corso...

Of course! Here's the revised translation with "patate" instead of "mele":

Poi si udì una voce arrabbiata—quella del Coniglio—"Pat! Pat! Dove sei?" E poi una voce che Alice non aveva mai sentito prima, "Certo che so' ccà! Sto scavando pe' truvà e' patate, vossignoria!"

"Scavando pe' truvà e' patate, davvero!" disse il Coniglio arrabbiato. "Vieni qui! Aiutami a uscire da questo pasticcio!" (Suoni di vetri rotti.)

"Ora dimmi, Pat, cos'è quello alla finestra?"

"Certo, è nu' vraccio, vossignoria!" (Lo pronunciava "vraccio.")

"Un braccio, sciocco! Chi ha mai visto un braccio di quelle dimensioni? Ma guarda, riempie tutta la finestra!"

"Certo che sì, vossignoria: ma è comunque nu' vraccio."


I hope this captures the playful and regional flavor of the original text! 😊

The translation could be considered flawless at this point.

 Tu

Can you please translate from English into Italian the following passage as if you were a literary translator? Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."

"If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."

 Copilot

✓ Generazione di risposte per te in corso...

Certainly! Here's the translation, maintaining the literary tone and character voices:

Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo facendo indovinelli senza risposta."

"Se conoscessi il Tempo come lo conosco io," disse il Cappellaio, "non parleresti di sprecarlo. È lui."


I hope this captures the essence and tone of the original passage! If you need any adjustments or further translations, feel free to let me know. 😊

Altre informazioni [1](#)  translate.google.it [2](#)  deepl.com [3](#)  translate.yandex.com [+13 altro](#)

      10 di 10

I tried to use prompts to obtain a better solution. I suggested the system to recreate the joke, proposing a pun that does not play on pronouns as in the original English text.

The original pun plays on the personification of time. In English, this is done using the masculine pronoun "him" instead of the neutral "it". However, the Italian language does not offer the possibility to repropose such a pun using pronouns as in the original, since there is no neutral pronoun. Could you please recreate the joke using a new punning mechanism, which is not based on pronouns?

 **Copilot**

✓ Generazione di risposte per te in corso...

Sure, I can help with that! Here's a version that plays on the concept of "time" in a different way:


Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo facendo domande a cui non ci sono risposte." "Se conoscessi il Tempo come lo conosco io," disse il Cappellaio, "sapresti che non è mai tempo perso. È sempre tempo guadagnato."

This version uses the idea that time is never wasted but always gained, creating a new pun around the value of time. How does that work for you?

The output provided by Copilot was different from the one provided by ChatGPT; therefore, I opted to use a different prompt to correct the output. Since there is no personification, I myself suggested the machine to do that, obtaining a really satisfactory translation.

 **Tu**

Can you please translate the following passage from English into Italian as if you were a literary translator? Please pay attention to puns and wordplays in the text. "Very true," said the Duchess: "flamingoes and mustard both bite. And the moral of that is—'Birds of a feather flock together.'"

 **Copilot**

✓ Generazione di risposte per te in corso...

Certainly! Here's a literary translation of the passage into Italian, preserving the puns and wordplays:


"Molto vero," disse la Duchessa: "i fenicotteri e la senape mordono entrambi. E la morale è—'Chi si somiglia si piglia.'"

I aimed to keep the playful tone and the pun intact. How does that sound to you?

As in the translation provided by ChatGPT, the system correctly recognized and translated the motto, while failing to provide a good solution for the passage “flamingoes and mustard both bite”. Using the knowledge acquired while testing prompts for ChatGPT, I provided more context for the translation.

 **Tu**

I would like you to maintain this translation, while changing some minor details. Can you please focus on the passage "flamingoes and mustard both bite"? The verb you used to translate bite is the literal translation of "to bite", which is not suitable. I need a verb that could describe both the "biting" action of a bird and the biting taste of mustard.

 **Copilot**

✓ Generazione di risposte per te in corso...

Absolutely, I understand the nuance you're looking for. Here's a revised translation that captures the dual meaning of "bite":

"Molto vero," disse la Duchessa: "i fenicotteri e la senape pizzicano entrambi. E la morale è—'Chi si somiglia si piglia.'"

The verb "pizzicano" can describe both the pecking action of a bird and the sharp taste of mustard. How does this version sound to you?

Giving more context to the system and explaining the kind of nuance I wanted to obtain was a good prompting solution: a good translation was obtained using just two prompts.

 **Tu**

Can you please translate the following passage from english into italian as if you were a literary translator? "Give your evidence," the King repeated angrily, "or I'll have you executed, whether you're nervous or not."


"I'm a poor man, your Majesty," the Hatter began, in a trembling voice, "—and I hadn't begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—"

"The twinkling of the what?" said the King.

"It began with the tea," the Hatter replied.

"Of course twinkling begins with a T!" said the King sharply. "Do you take me for a dunce? Go on!"



 **Copilot**

✓ Generazione di risposte per te in corso...

Sure! Here's a literary translation of the passage into Italian, maintaining the playful tone and wordplay:

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il luccichio del tè—" "Il luccichio di cosa?" disse il Re. "È iniziato con il tè," rispose il Cappellaio. "Naturalmente il luccichio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

I aimed to preserve the whimsical nature and the wordplay of the original text. How does this version sound to you?

As it was the case with ChatGPT, the passage was not translated properly. The "twinkling" pun was not rendered in a good way.



Tu

Can you please focus more on the passages involving the "twinkling tea" pun? This is the central pun that needs to be translated



Copilot

✓ Generazione di risposte per te in corso...

Certainly! Here's a revised translation that emphasizes the "twinkling tea" pun:

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il luccichio del tè—" "Il luccichio di cosa?" disse il Re. "È iniziato con il tè," rispose il Cappellaio. "Naturalmente il luccichio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

I kept the playful nature and wordplay intact, focusing on the pun involving "twinkling tea." How does this version sound to you?



Tu

Unofrtunately, the words "luccichio" does not start with T, therefore the joke makes no sense in Italian. Could you please try to find a translation that takes into consideration this aspect?



Copilot

✓ Generazione di risposte per te in corso...

I see what you mean. Let's try this:

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il tintinnio del tè—" "Il tintinnio di cosa?" disse il Re. "È iniziato con il tè," rispose il Cappellaio. "Naturalmente il tintinnio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

In this version, "tintinnio" (which means "tinkling" or "jingling") starts with a T, preserving the pun. How does this sound to you?

 **Tu**

Tintinnio is a good translation. However, in the original text the homophones "tea" and "t" are exploited. In Italian the letter "t" is not a homophone of the word "tè". Could you please try to find a solution to this issue?

 **Copilot**

✓ Generazione di risposte per te in corso...

I understand the challenge. Let's try to create a similar wordplay in Italian that maintains the humor and the pun. Here's a revised version:

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il tintinnio del tè—" "Il tintinnio di cosa?" disse il Re. "È iniziato con il tè," rispose il Cappellaio. "Naturalmente il tintinnio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

In this version, "tintinnio" (tinkling) is used to maintain the pun, and the King's response plays on the fact that "tintinnio" starts with a T, similar to the original wordplay. How does this work for you?

 **Tu**

Could you please make the Hatter stutter because he is afraid of the king? In this way he might mispronounce the word "tè", making the sentence "Naturalmente il tintinnio inizia con una T!" meaningful

 **Copilot**

✓ Generazione di risposte per te in corso...


Of course! Here's a revised translation with the Hatter stuttering to mispronounce "tè," making the King's response meaningful:

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il t-t-tintinnio del tè—" "Il tintinnio di cosa?" disse il Re. "È iniziato con il t-t-tè," rispose il Cappellaio. "Naturalmente il tintinnio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

This version incorporates the stutter to create the pun. How does this work for you?

 Tu

Could you please rewrite the sentence in order for the Hatter not to finish the word tè because of the stuttering?

 Copilot

✓ Generazione di risposte per te in corso...

Certainly! Here's the revised translation with the Hatter stuttering and not finishing the word "tè":

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il t-t-tintinnio del t—" "Il tintinnio di cosa?" disse il Re. "È iniziato con il t-t-t—" rispose il Cappellaio. "Naturalmente il tintinnio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

This version should maintain the humor and wordplay effectively. How does this sound to you?

     8 di 10

Thanks to suitable prompts, it was possible to guide the reasoning path of the generative AI system so that it could propose a suitable translation. After managing to find a proper equivalent to “twinkling”, it was possible to recreate the joke, obtaining a translation that reproduced the original pun quite closely.

3.9

AI SECOND EVALUATION

Before getting to the conclusions, the results obtained with generative AI will be presented and scored according to Matusov’s criteria, in order to compare the results obtained by using prompts to the results obtained at the beginning of the study.

CHAT GPT 3.5



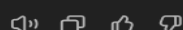
Absolutely! Here's the improved version with additional context to clarify that the *filo* refers to a thread from the Mouse's clothing:

«Mi avevi promesso di raccontarmi la tua storia, lo sai», disse Alice, «e perché odi la C e la D», aggiunse sottovoce, quasi temendo che si offendesse di nuovo.

«La mia è una lunga e triste trama!» disse il Topo, girandosi verso Alice e sospirando.

«Lunga è lunga, ma trama di che?» disse Alice, guardando confusa un filo che pendeva dal vestito del Topo. «Io vedo solo un filo!» continuò, perplessa, mentre il Topo parlava, tanto che la sua idea della *trama* era più o meno così: –

Now, with the phrase *un filo che pendeva dal vestito del Topo* (a thread hanging from the Mouse's clothing), it's clear that Alice is confusing the Mouse's metaphorical "trama" (plot) with the literal *filo* (thread) from its attire, strengthening the fabric-related pun.



Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	1	The translation “e perché odi la C e la D” is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by ChatGPT was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	



Of course! I'll adjust Pat's activity to something more fitting for an Italian context, while maintaining his rustic, dialect-influenced speech. Let's have Pat involved in a more relatable, countryside task, such as digging for vegetables like potatoes or something similar. Here's the revised version:

Subito dopo, si sentì una voce arrabbiata—quella del Coniglio—«Pat! Pat! Dove sei?» E poi una voce che Alice non aveva mai sentito prima, «Eh, so' qui, vostro onore! Sto scavanno pe' le patate!» «Scavando per le patate, davvero!» disse il Coniglio con rabbia. «Vieni qui! Aiutami a uscire da questo pasticcio!» (Suoni di altro vetro che si rompe.) «Ora dimmi, Pat, che cos'è quella cosa alla finestra?» «Ma certo, è un bbraccio, vostro onore!» (Lo pronunciò "bbraccio".) «Un braccio, stupido! Chi ha mai visto un braccio di quelle dimensioni? Ma se riempie tutta la finestra!» «Certo che sì, vostro onore: ma è un bbraccio lo stesso.»

Here, I've replaced the apples with Pat digging for *patate* (potatoes), a common and realistic task for a rustic Italian character, especially one living in a rural area. This makes his activity more plausible within an Italian-speaking context while preserving the tone and humor of the original text.

Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

Could you please repropose the same pun, this time using the verb "impiegare"?

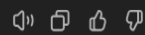


Certainly! Here's the same pun, now incorporating the verb *impiegare* to emphasize Alice's intention to use time productively, while the Hatter suggests that Time is lazy:

Alice sospirò stancamente. «Credo che potresti fare qualcosa di meglio con il tempo,» disse, «piuttosto che impiegarlo in indovinelli senza risposta.»

«Se conoscessi il Tempo bene quanto lo conosco io,» disse il Cappellaio, «sapresti che è inutile cercare di impiegarlo. Lui è un fannullone.»

In this version, *impiegare* (to use or employ) highlights Alice's desire to use time efficiently, while the Hatter humorously suggests that it's pointless to do so because Time is lazy.



Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

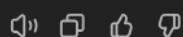
Could you please find a synonym to "pungere"? I would like you to find a verb that could describe better the idea of a bird biting



Certainly! Here's an updated version with a more fitting verb:

«Verissimo,» disse la Duchessa, «i fenicotteri e la senape pizzicano entrambi. E la morale è: "Chi si somiglia si piglia."»

The verb *pizzicare* (to pinch or nip) works well here, as it can describe both the action of a bird like a flamingo lightly biting or nipping, and the sharp, biting taste of mustard. This preserves the playful wordplay in a natural-sounding way.



Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	



Thank you for the clarification! Here's the new version where the Hatter stutters on the word "tè," leading to the King's confusion, while keeping closer to the original "it began with the tea."

"Presentate le vostre prove," ripeté il Re furioso, "o vi faccio giustiziare, che siate nervoso o no."

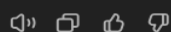
"Sono un pover'uomo, Maestà," cominciò il Cappellaio, balbettando, "—e non avevo nemmeno iniziato il mio tè—non da più di una settimana, o giù di lì—e con il pane e burro che si faceva sempre più sottile—e il t-t-tintinnio del t-t-t—"

"Il t-t-tintinnio del che?" disse il Re.

"È c-c-cominciato col t-t-t..." cercò di spiegare il Cappellaio, senza riuscire a finire.

"Ma certo che comincia con la T!" ribatté il Re seccato. "Mi prendete per uno stupido? Avanti!"

In this version, the Hatter's stuttering makes it impossible for him to finish saying "tè," and the King mistakes the repeated "T" for a comment on the letter itself. This keeps the humor of the original, while adapting it for Italian wordplay.



Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

MICROSOFT EDGE COPILOT



Copilot

✓ Generazione di risposte per te in corso...

Certainly! Let's try to recreate the joke using a synonym for "storia" that could be confused with a physical attribute of the Mouse. Here's a new version:

"Mi avevi promesso di raccontarmi la tua vicenda, sai," disse Alice, "e perché odi – C e D," aggiunse sottovoce, mezza impaurita che si offendesse di nuovo.

"La mia è una lunga e triste vicenda!" disse il Topo, voltandosi verso Alice e sospirando.

"È una lunga fila di baffi, certamente," disse Alice, guardando con meraviglia i baffi del Topo; "ma perché la chiami triste?" E continuava a rimuginarci sopra mentre il Topo parlava, così che la sua idea della vicenda era qualcosa del genere: –

In this version, Alice confuses "vicenda" (which can mean "story" or "event") with "fila di baffi" (row of whiskers), creating a playful misunderstanding. I hope this captures the whimsical nature of the original text! 😊



Esplora

Type of error	Number of errors	Explanation
Severe meaning error	1	Even with the prompts, it is clear that Copilot has not been able to recreate the joke in a suitable way. The meaning of the text is not understandable, and looking at the ST is necessary.
Minor meaning error	1	The translation "e perché odi – C e D" is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by Copilot was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	

Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

 **Tu**

Can you please substitute "mele" with "patate"?

 **Copilot**

✓ Generazione di risposte per te in corso...

Of course! Here's the revised translation with "patate" instead of "mele":

Poi si udì una voce arrabbiata—quella del Coniglio—"Pat! Pat! Dove sei?" E poi una voce che Alice non aveva mai sentito prima, "Certo che so' ccà! Sto scavando pe' truvà e' patate, vossignoria!"

"Scavando pe' truvà e' patate, davvero!" disse il Coniglio arrabbiato. "Vieni qui! Aiutami a uscire da questo pasticcio!" (Suoni di vetri rotti.)

"Ora dimmi, Pat, cos'è quello alla finestra?"

"Certo, è nu' vraccio, vossignoria!" (Lo pronunciava "vraccio.")

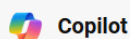
"Un braccio, sciocco! Chi ha mai visto un braccio di quelle dimensioni? Ma guarda, riempie tutta la finestra!"

"Certo che sì, vossignoria: ma è comunque nu' vraccio."

I hope this captures the playful and regional flavor of the original text! 😊

Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

The original pun plays on the personification of time. In English, this is done using the masculine pronoun "him" instead of the neutral "it". However, the Italian language does not offer the possibility to repropose such a pun using pronouns as in the original, since there is no neutral pronoun. Could you please recreate the joke using a new punning mechanism, which is not based on pronouns?



✓ Generazione di risposte per te in corso...

Sure, I can help with that! Here's a version that plays on the concept of "time" in a different way:


Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo facendo domande a cui non ci sono risposte." "Se conoscessi il Tempo come lo conosco io," disse il Cappellaio, "sapresti che non è mai tempo perso. È sempre tempo guadagnato."

This version uses the idea that time is never wasted but always gained, creating a new pun around the value of time. How does that work for you?

Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

 Tu

I would like you to maintain this translation, while changing some minor details. Can you please focus on the passage "flamingoes and mustard both bite"? The verb you used to translate bite is the literal translation of "to bite", which is not suitable. I need a verb that could describe both the "biting" action of a bird and the biting taste of mustard.

 Copilot

✓ Generazione di risposte per te in corso...

Absolutely, I understand the nuance you're looking for. Here's a revised translation that captures the dual meaning of "bite":

"Molto vero," disse la Duchessa: "i fenicotteri e la senape pizzicano entrambi. E la morale è—'Chi si somiglia si piglia.'"

The verb "pizzicano" can describe both the pecking action of a bird and the sharp taste of mustard. How does this version sound to you?

Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	



Tu

Could you please rewrite the sentence in order for the Hatter not to finish the word tè because of the stuttering?



Copilot

✓ Generazione di risposte per te in corso...

Certainly! Here's the revised translation with the Hatter stuttering and not finishing the word "tè":

"Date la vostra testimonianza," ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervoso o no." "Sono un pover'uomo, Maestà," iniziò il Cappellaio con voce tremante, "—e non avevo ancora iniziato il mio tè—non più di una settimana fa—e con il pane e burro che diventano così sottili—e il t-t-tintinnio del t—" "Il tintinnio di cosa?" disse il Re. "È iniziato con il t-t-t—" rispose il Cappellaio. "Naturalmente il tintinnio inizia con una T!" disse il Re bruscamente. "Mi prendete per uno sciocco? Avanti!"

This version should maintain the humor and wordplay effectively. How does this sound to you?



8 di 10

Type of error	Number of errors	Explanation
Severe meaning error	0	
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error		
Tone/register error	0	

Now, a table containing all the results of the evaluation will be presented.

	ChatGPT 3.5	Microsoft Edge Copilot
Severe meaning error	0	1
Minor meaning error	1	1
Unknown word or segmentation error	0	0
Consistency/Term translation error	0	0
Pronoun resolution error	0	0
Locution error	0	0
Omission, insertion, repetition errors	0	0
Severe syntax error	0	0
Minor syntax error	0	0
Tone/register error	0	0

As it can be seen, the results sensibly improved after applying suitable prompts. However, it is still interesting to notice that Copilot performed worse than ChatGPT. While the latter did not commit any Severe meaning error, the former committed one. This seems to suggest that Copilot is less receptive to prompts than GPT. GPT, nevertheless, committed one Minor meaning error. Both of the errors, however, did not cause any problem as far as text comprehension is concerned.

MACHINE TRANSLATION TRAINING

This chapter will be focused on the training of the two Machine Translation providers chosen for this study: RWS Language Weaver and IntoMT. In order to obtain better translations than those obtained in the first stage of the study, the two above-mentioned systems were trained using linguistic data gathered from literary texts.

Machine Translation models can be trained in order to “teach” the machine how to translate texts belonging to certain domains. For example, if a translator were to deal with a legal text, they might use MT models trained with linguistic corpora based on legal texts, in order to work with a machine that would be able to recognize, use and reproduce certain structures and vocabulary. In order to train an MT model, it must be fed with a bilingual file in the .tmx format. In other words, the machine is trained using a translation memory containing linguistic items both in the Source Language and in the Target Language. In this specific case, the file used was an English-Italian translation memory made of roughly 16,000 segments gathered from several different novels. The texts chosen are mainly books written for children or young adults. The memory was created manually using the Alignment function on Trados Studio 2021.

3.10.1

RWS LANGUAGE WEAVER

As far as RWS Language Weaver is concerned, in order to train the provider the integrated function “Adapted Language Pairs” was used. As it is written on the Language Weaver website: “adapting a language pair means customizing an existing generic language pair with a bilingual corpus in the form of a TMX file that corresponds to a specific domain” (<https://www.rws.com/language-weaver/adaptive-machine-translation-for-translators-faq/>) . As already explained above, the bilingual corpus used was created manually aligning a series of novels and short stories. After training the provider, the same texts presented and analyzed in the previous section of the study were translated. The translation was followed both by a manual and automatic evaluation, in order to identify changes and improvements.

English ▾

↔ Italian ▾ Adapted-Thesis-11939 Generic ⓘ

"You promised to tell me your history, you know," said Alice, "and why it is you hate – C and D," she added in a whisper, half afraid that it would be offended again.

"Mine is a long and sad tale!" said the Mouse, turning to Alice, and sighing.

"It is a long tail, certainly," said Alice, looking down with wonder at the Mouse's tail; "but why do you call it sad?" And she kept on puzzling about it while the Mouse was speaking, so that her idea of the tale was something like this: –

— Mi hai promesso di raccontarmi la tua storia, sai, — disse Alice, — e perché tu odi – C e D – aggiunte in un sussurro, per metà paura che si sarebbe offeso di nuovo. La mia è una storia lunga e triste! Disse il topo, volgendosi ad Alice e sospirando. E' certamente una coda lunga, — disse Alice, guardando con stupore la coda del topo; — ma perché la chiami triste? E continuava a sconcertarlo mentre parlava il topo, cosicché la sua idea del racconto era simile a questa

Clear text 488/5000 Rate Translation: ★★★★★

[Click to upload files or drag and drop on the page](#)
[+ Add label](#)
[Copy](#)
[Improve](#)
[Feedback History](#)

Type of error	Number of errors	Explanation
Severe meaning error	2	<p>The translation “ha continuato a sconcertarlo” is a severe meaning error. It did not translate correctly the phrase “she kept on puzzling about it”. The ST phrase meant that Alice kept interrogating herself about the Mouse’s tale. The Italian sentence meant that Alice kept maintaining a behavior that offends and surprises (in a negative way) the Mouse.</p> <p>The system mistranslated the terms “tail” and “tale”. Mistranslation here is used not to refer to the grammar per se, since the grammatical translation is correct. It is used to refer to the mistranslation of the humorous intent of the original ST. The machine translated the two items into the (grammatically correct) Italian equivalents; however, the best way to translate this joke would have been that</p>

		of using two Italian homophones, so that to repropose the original pun.
Minor meaning error	1	The translation “e perché tu odi – C e D” is not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by RWS Language Weaver was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	1	“Half afraid” was translated with “metà paura”. This translation was not acceptable according to Italian syntax, since it had no meaning whatsoever in this context. A more suitable translation would have been “mezza timorosa”, or “mezza impaurita”.
Tone/register error	0	

English
Italian
Adapted-Thesis-11939
Generic

Next came an angry voice—the Rabbit’s—“Pat! Pat! Where are you?” And then a voice she had never heard before, “Sure then I’m here! Digging for apples, yer honour!” “Digging for apples, indeed!” said the Rabbit angrily. “Here! Come and help me out of this!” (Sounds of more broken glass.)

“Now tell me, Pat, what’s that in the window?”

“Sure, it’s an arm, yer honour!” (He pronounced it “arrum.”)

“An arm, you goose! Who ever saw one that size? Why, it fills the whole window!”

“Sure, it does, yer honour: but it’s an arm for all that.”

Poi venne una voce arrabbiata, quella del Coniglio, — Pat! Pat! Dove sei? E poi una voce che non aveva mai sentito prima, “certo che sono qui! Scavare per le mele, vostro Onore!

- Scavare per le mele! Disse il Coniglio arrabbiato. Ecco! Vieni ad aiutarmi! (Suoni di vetro più rotto.)

- Dimmi, Pat, che c'è nella finestra?

— Certo, è un braccio, vostro Onore! (Lo pronunciò “arrum”).

Un braccio, oca! Chi ne ha mai visto uno così grande? Perché, riempie tutta la finestra!

— Certo, è vero, vostro onore; ma è un braccio per tutto questo.

Clear text 536/5000
Rate Translation:
Copy
Improve
Feedback History

Type of error	Number of errors	Explanation
Severe meaning error	1	The humorous device used by Carrol in the original passage was based on Pat being an Irish speaker. The translation neither repropose the original device, nor provided a new device suitable to the Italian readers. This is related to the machine being unable to retrieve the necessary cultural knowledge.
Minor meaning error	0	
Unknown word or segmentation error	1	The word “arrum” being kept as it is was a result of the lack of cultural knowledge on the part of the Machine Translation system. The system being unable to identify “arrum” as the eye dialect rendition of the word “arm” as pronounced by an Irish speaker lead to the system not translating the word.
Consistency/Term translation error	0	

Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	2	<p>“Suoni di vetro” più rotto is wrong. The Italian reader would not get the original idea intended by Carrol: more glass breaking is heard. With this rendition, it seems that the same pieces of glass got broken again into more pieces.</p> <p>The translation “è un braccio per tutto questo” was not idiomatic and too literal.</p>
Tone/register error	0	

English ▾



Italian ▾

Adapted-Thesis-11939

Generic



Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."
 "If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."

Alice sospirò stanca. "Penso che col tempo potresti fare qualcosa di meglio", disse, "che sprecarla nel chiedere indovinelli che non hanno risposte."
 Se tu conoscessi il tempo come me, — disse il Cappellaio, — non parleresti di sprecare il tempo. È lui.

Clear text 245/5000

Rate Translation: ☆☆☆☆☆

Copy

Improve

Feedback History

Click to upload files or drag and drop on the page.

+ Add label

Type of error	Number of errors	Explanation
Severe meaning error	1	The pun was not reproposed in a good way. It was not possible to understand the meaning of the passage without looking at the Source Text.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0,5	As in the previous analysis, the mistranslation of the neutral pronoun "it" into the masculine "lui" would be considered as half of an error. The system could not provide a suitable translation because of the characteristics of the Italian grammar, which does not allow for the use of neutral pronouns.
Locution error	0	

Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	1	The time was referred to using the feminine form “sprecarla” and not the correct “sprecarlo”. However, this did not affect syntax in a meaningful way.
Tone/register error	0	

English
Italian
Adapted-Thesis-11939
Generic

"Very true," said the Duchess: "flamingoes and mustard both bite. And the moral of that is—"Birds of a feather flock together."

"Molto vero", disse la duchessa: "I fenicotteri e la senape mordono entrambi. E la morale di questo è: "Gli uccelli di un gregge di piume insieme".

Clear text
128/5000
Rate Translation:
Copy
Improve
Feedback History

Click to upload files or drag and drop on the page
Add label

Type of error	Number of errors	Explanation
Severe meaning error	1	The literal translation of the verb “to bite” into the Italian equivalent “mordere” was not suitable. The literal translation made it impossible to repropose the original pun in a good way, forcing to read the Source Text to understand completely the joke.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	1	The motto “birds of a feather flock together” was not translated corrected. It should have been used a suitable Italian motto.

Omission, insertion, repetition errors	0	
Severe syntax error	1	The rendition of the motto was not comprehensible in Italian.
Minor syntax error	0	
Tone/register error	0	

English
Italian
Adapted-Thesis-11939
Generic

“Give your evidence,” the King repeated angrily, “or I’ll have you executed, whether you’re nervous or not.”
“I’m a poor man, your Majesty,” the Hatter began, in a trembling voice, “—and I hadn’t begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—”
“The twinkling of the what?” said the King.
“It began with the tea,” the Hatter replied.
“Of course twinkling begins with a T!” said the King sharply. “Do you take me for a dunce? Go on!”

— Di la tua prova, — ripeté il Re arrabbiato, — o ti farò giustiziare, che tu sia nervoso o no.
Io sono un povero uomo, vostra Maestà, — cominciò il Cappellaio con una voce tremante, — e non avevo iniziato il mio tè, non più di una settimana, e che con il pane e il burro che diventavano così sottili, e lo scintillio del tè...
- Il scintillio del che? Disse il Re.
- Cominciava col tè - rispose il Cappellaio.
— Certo che lo scintillio comincia con una T! Disse il Re con decisione. - Mi porti per un dunce? Avanti!

Clear text 502/5000
Rate Translation: ☆☆☆☆☆
Copy Improve Feedback History

Click to upload files or drag and drop on the page
Add label

Type of error	Number of errors	Explanation
Severe meaning error	2	<p>The literal translation of “Do you take me for a dunce” had no meaning in Italian, this resulting in a severe mistranslation error that compromises the meaning of the sentence.</p> <p>The pun present in the Target Text was not translated correctly; the same error committed in the pre-training analysis occurred.</p>
Minor meaning error	0	
Unknown word or segmentation error	1	The word “dunce” was not recognized by the Machine Translation tool, this resulting in a missing translation.
Consistency/Term translation error	0	
Pronoun resolution error	0	

Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

3.10.2

INTENTO MODERNMT

Now, the result obtained with ModernMT will be presented. This Machine Translation tool has been trained using the same exact TM that was used in order to train the RWS Language Weaver. ModernMT was trained using the MT suite made available by Intento ModernMT on the MT Hub; after the training, the ModernMT web interface was used to perform the translations, using the trained MT specifically created for this project. The results were evaluated manually using Matusov's criteria.

Translate from
English

Translate to
Italian

×

“You promised to tell me your history, you know,” said Alice, “and why it is you hate – C and D,” she added in a whisper, half afraid that it would be offended again.

“Mine is a long and sad tale!” said the Mouse, turning to Alice, and sighing.

“It is a long tail, certainly,” said Alice, looking down with wonder at the Mouse’s tail; “but why do you call it sad?” And she kept on puzzling about it while the Mouse was speaking, so that her idea of the tale was something like this: –

↔

“Hai promesso di raccontarmi la tua storia, sai,” disse Alice, “e perché è che odi – C e D,” aggiunse in un sussurro, quasi temendo che si offendesse di nuovo.

“La mia è una storia lunga e triste!” disse il Topo, voltandosi verso Alice e sospirando.

“È una coda lunga, certamente”, disse Alice, guardando con meraviglia la coda del topo; “ma perché lo chiami triste?” E continuava a confonderlo mentre il Topo parlava, in modo che la sua idea del racconto fosse qualcosa del genere: –

488 / 5000

📄

Type of error	Number of errors	Explanation
Severe meaning error	1	The original pun was not reproposed properly. The homophony of the terms “tale” and “tail” was not reproduced in a correct way. This resulted in a pun that did not work properly in Italian.
Minor meaning error	1	The translation “e perché tu odi – C e D” was not acceptable. In the original text Alice used the letters C and D in order not to mention Cats and Dogs, which makes the Mouse nervous. The Italian translation provided by ModernMT was not suitable to make the readers understand this nuance. However, this error did not cause major misunderstandings and it was still possible to understand the text.

145

Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

Translate from
English

Translate to
Italian

×

Next came an angry voice—the Rabbit's —“Pat! Pat! Where are you?” And then a voice she had never heard before, “Sure then I'm here! Digging for apples, yer honour!”

“Digging for apples, indeed!” said the Rabbit angrily. “Here! Come and help me out of this!” (Sounds of more broken glass.)

“Now tell me, Pat, what's that in the window?”

“Sure, it's an arm, yer honour!” (He pronounced it “arrum.”)

“An arm, you goose! Who ever saw one that size? Why, it fills the whole window!”

“Sure, it does, yer honour: but it's an arm for all that.”

Poi arrivò una voce arrabbiata - quella del Coniglio-“Pat! Pat! Dove sei?” E poi una voce che non aveva mai sentito prima: “Certo, allora sono qui! Alla ricerca di mele, il tuo onore!”

Scavare per le mele, davvero! - disse il Coniglio con rabbia. “Ecco! Vieni ad aiutarmi!” (Suoni di altri vetri rotti.)

“Ora dimmi, Pat, cosa c'è nella finestra?”

“Certo, è un braccio, tuo onore!” (L'ha pronunciato “arrum.”)

“Un braccio, oca! Chi ne ha mai visto uno così grande? Perché, riempie tutta la finestra!”

“Certo che sì, il tuo onore: ma è un braccio per tutto questo.”

Type of error	Number of errors	Explanation
Severe meaning error	1	The joke was not reposed correctly. The Machine Translator engine lacked the capacity to grasp the cultural subtleties of the original Source Text, thus being unable to reproduce the joke effectively.
Minor meaning error	0	
Unknown word or segmentation error	1	The word “arrum” was not recognized as a way to imitate and reproduce graphically the pronunciation of the word “arm” by an Irish speaker; ModernMT was not able to produce an acceptable translation.
Consistency/Term translation error	0	

Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	1	Literal translation of “It’s an arm for all that”. The Italian Target Text stucked too much to the original, producing a text that was not immediately understandable. However, this did not make impossible for the reader to grasp the meaning of the phrase.
Tone/register error	0	

Translate from
English

Translate to
Italian

×

Alice sighed wearily. "I think you might do something better with the time," she said, "than waste it in asking riddles that have no answers."

"If you knew Time as well as I do," said the Hatter, "you wouldn't talk about wasting it. It's him."

247 / 5000

↕

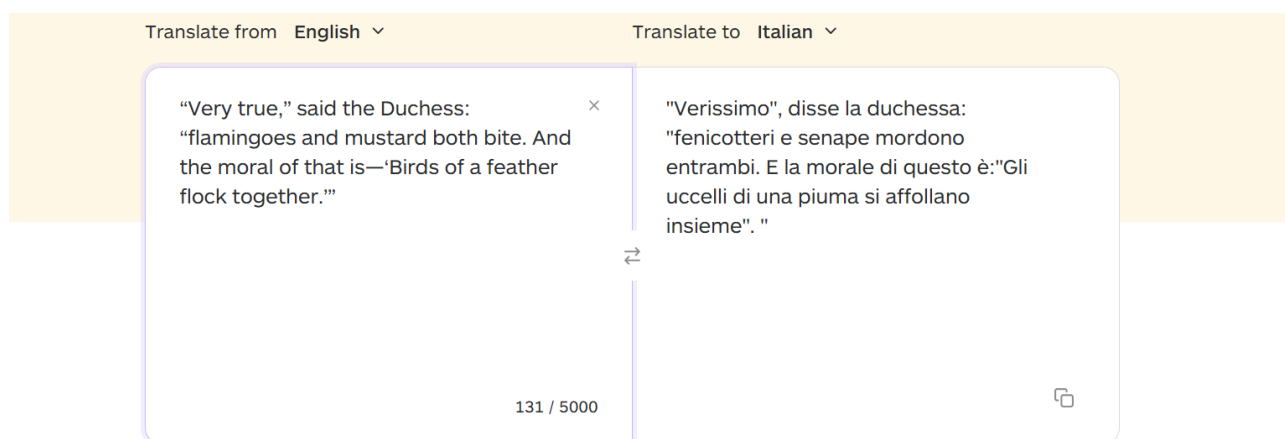
Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo a fare indovinelli che non hanno risposte."

"Se conoscessi il tempo come lo conosco io", disse il Cappellaio, "non parleresti di sprecarlo. È lui."

📄

Type of error	Number of errors	Explanation
Severe meaning error	1	The pun was not reposed in a good way. It was not possible to understand the joke without looking at the Source Text.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0,5	As it has been done before, this error was marked only as half an error, since the Italian grammar does not allow for the use of neutral pronouns.
Locution error	0	

Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	



Type of error	Number of errors	Explanation
Severe meaning error	1	The joke was not translated in an acceptable way. The literal translation of the verb “to bite” with “mordere” did not allow the Italian reader to understand the pun. It would be necessary to look at the Source Text.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	0	
Pronoun resolution error	0	
Locution error	1	The motto “Birds of a feather flock together” should have been translated with a suitable Italian motto.

Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

Translate from
English

Translate to
Italian

Give your evidence," the King repeated angrily, "or I'll have you executed, whether you're nervous or not."

"I'm a poor man, your Majesty," the Hatter began, in a trembling voice, "— and I hadn't begun my tea—not above a week or so—and what with the bread-and-butter getting so thin—and the twinkling of the tea—"

"The twinkling of the what?" said the King.

"It began with the tea," the Hatter replied.

"Of course twinkling begins with a T!" said the King sharply. "Do you take me for a dunce? Go on!"

"Dà la tua testimonianza," ripeté il Re con rabbia, "o ti farò giustiziare, che tu sia nervoso o meno."

"Sono un pover' uomo, Vostra Maestà," iniziò il Cappellaio, con voce tremante, "e non avevo iniziato il mio tè - non più di una settimana o giù di lì - e con il pane e il burro che diventavano così sottili e lo scintillio del tè..."

Il luccichio di cosa? disse il Re.

"È iniziato con il tè," rispose il Cappellaio.

"Certo che lo scintillio inizia con una T!" disse bruscamente il Re. "Mi prendi per un somaro? Vai avanti!"

502 / 5000

Type of error	Number of errors	Explanation
Severe meaning error	1	In this case, meaning error occurred because NMT was not able to recreate the passage satisfactorily. The “mere” translation of the phrases containing the pun on the homophones “tea” and “t” did not recreate the original effect of the joke in the Target Text.
Minor meaning error	0	
Unknown word or segmentation error	0	
Consistency/Term translation error	1	Twinkling was translated first into “scintillio” and then into “luccichio” (then again into “scintillio”).

Pronoun resolution error	0	
Locution error	0	
Omission, insertion, repetition errors	0	
Severe syntax error	0	
Minor syntax error	0	
Tone/register error	0	

MT SECOND EVALUATION

In order to conduct the second evaluation of the Machine Translation tools used in the study, the automatic metrics presented in the previous part were used. The analysis conducted in this section aimed at identifying any possible improvement occurred thanks to the training underwent by both Intento ModernMT and RWS Language Weaver. Before presenting the results obtained with the metrics, the results of the manual evaluation will be presented in the following table.

	Post-training RWS Language Weaver	Post-training Intento ModernMT
Severe meaning error	7	5
Minor Meaning error	1	1
Unknown word or segmentation error	2	1
Consistency/Term translation error	0	1
Pronoun resolution error	0,5	0,5
Locution error	1	1
Omission, insertion, repetition errors	0	0
Severe syntax error	1	0
Minor syntax error	4	1
Tone/register error	0	0

As it can be seen, ModernMT still performed better than RWS Language Weaver. The most notable category of errors is the Severe meaning error: Language Weaver scored 7 points, against the 5 scored by ModernMT. Moreover, it is possible to notice that Language Weaver performed much worse than Intento ModernMT when it comes to Minor syntax errors. Now, the two scoring sessions will be compared.

	Pre-training RWS Language Weaver	Pre-training Intento ModernMT
Severe meaning error	7	5
Minor meaning error	1	2
Unknown word or segmentation error	2	1
Consistency/Term translation error	1	1
Pronoun resolution error	0,5	0,5
Locution error	1	1
Omission, insertion, repetition errors	1	0
Severe syntax error	1	0
Minor syntax error	6	2
Tone/register error	0	0

	Post-training RWS Language Weaver	Post-training Intento ModernMT
Severe meaning error	7	5
Minor Meaning error	1	1
Unknown word or segmentation error	2	1
Consistency/Term translation error	0	1
Pronoun resolution error	0,5	0,5
Locution error	1	1
Omission, insertion, repetition errors	0	0
Severe syntax error	1	0
Minor syntax error	4	1
Tone/register error	0	0

Before analyzing the actual scorings, it is interesting to notice that the training had really positive effects on RWS Language Weaver as far as syntax is concerned. As it was made clear by the tables, the general number of syntax errors is reduced in the second evaluation. Moreover, it was possible to observe a general improvement in the use of verb tenses and modalities. The same could be said for ModernMT: we could observe a slight improvement in Minor syntax errors, where the number was reduced by the training from 2 errors to 1 error. However, ModernMT tended to produce quite similar translation results. From this point of view, the training seemed to be less effective, since this was the only improvement that could be observed. The same could not be said of Language Weaver, where the training improved not only Syntax, but also Consistency: after the training no Consistency errors were committed; the same exact situation could be observed with Omission errors, which were not committed after the training. It was interesting to note that Language Weaver performed worse in terms of Locution errors after the training, committing one more error: the rendition of “Birds of a feather flock together” was

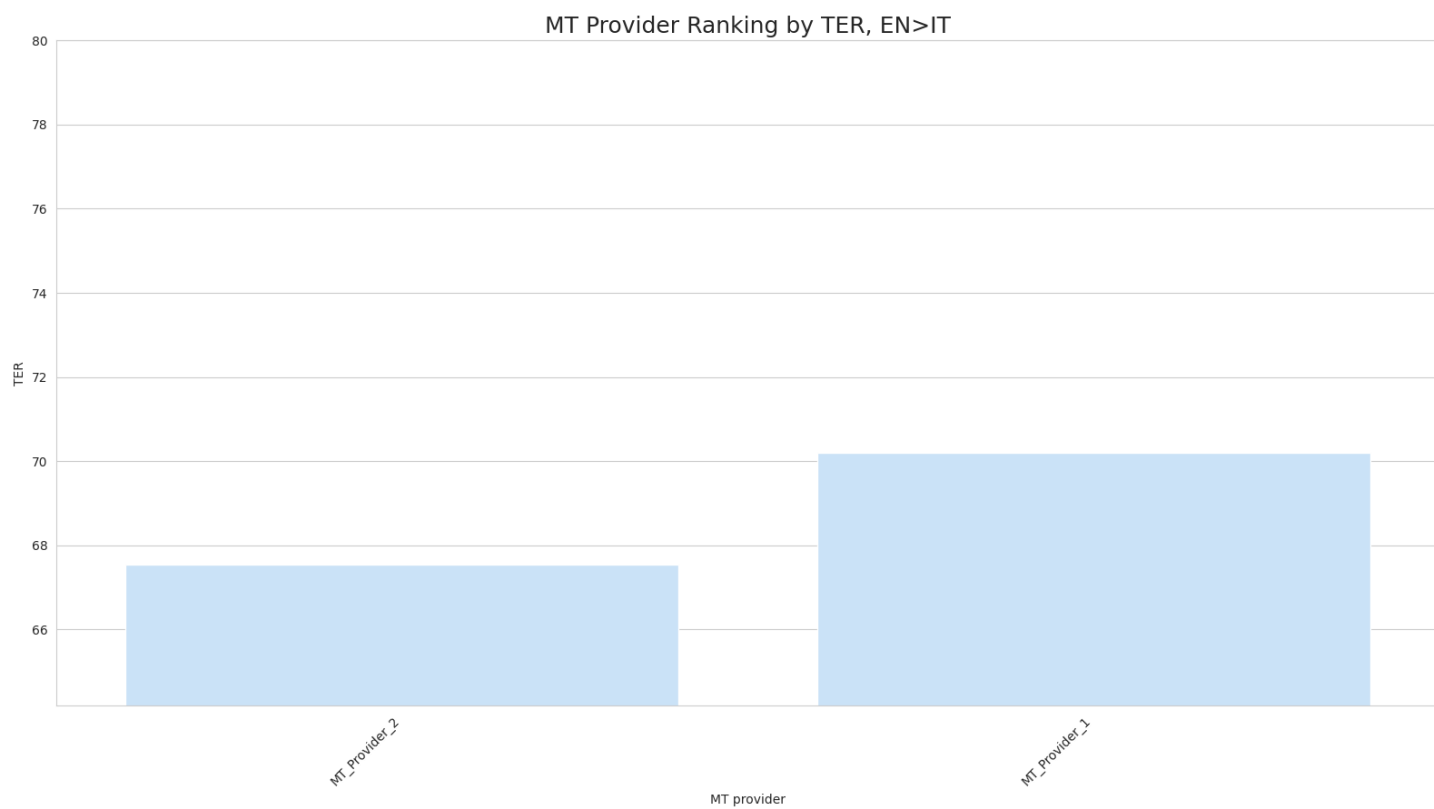
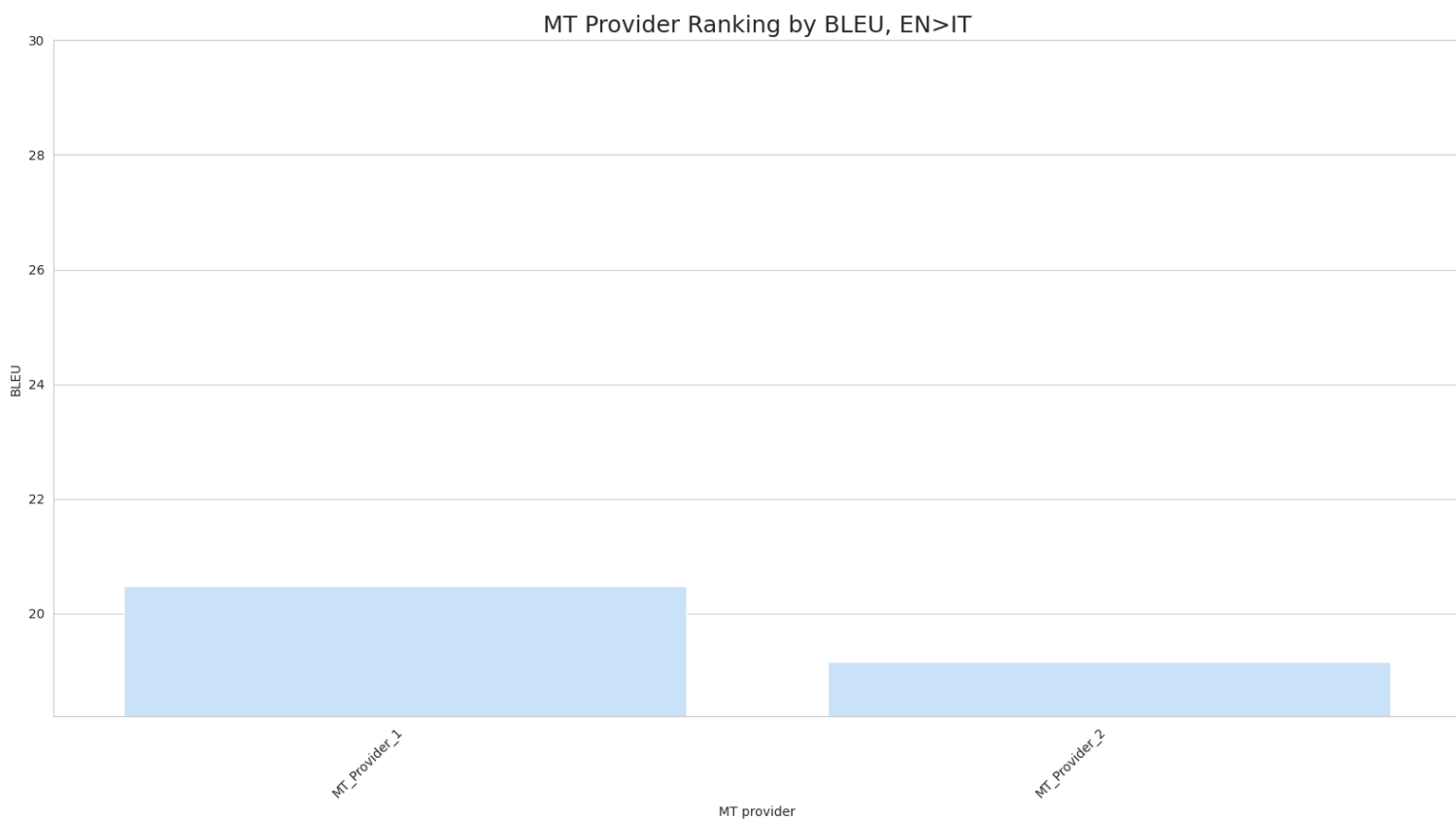
not understandable at all, since the Italian phrase used by the Machine Translation tools had no meaning in Italian.

If we look at the main object of research, we notice that the results are the same both before and after the training. Apparently, training Machine Translation tools with TMs developed for literary translation can have some positive effects on Locution and Syntax, as well as it can help avoiding Omission errors; however, this training does not seem helpful when it comes to puns and wordplay. Probably, the translation of puns and wordplay still requires human interaction in order to be effective.

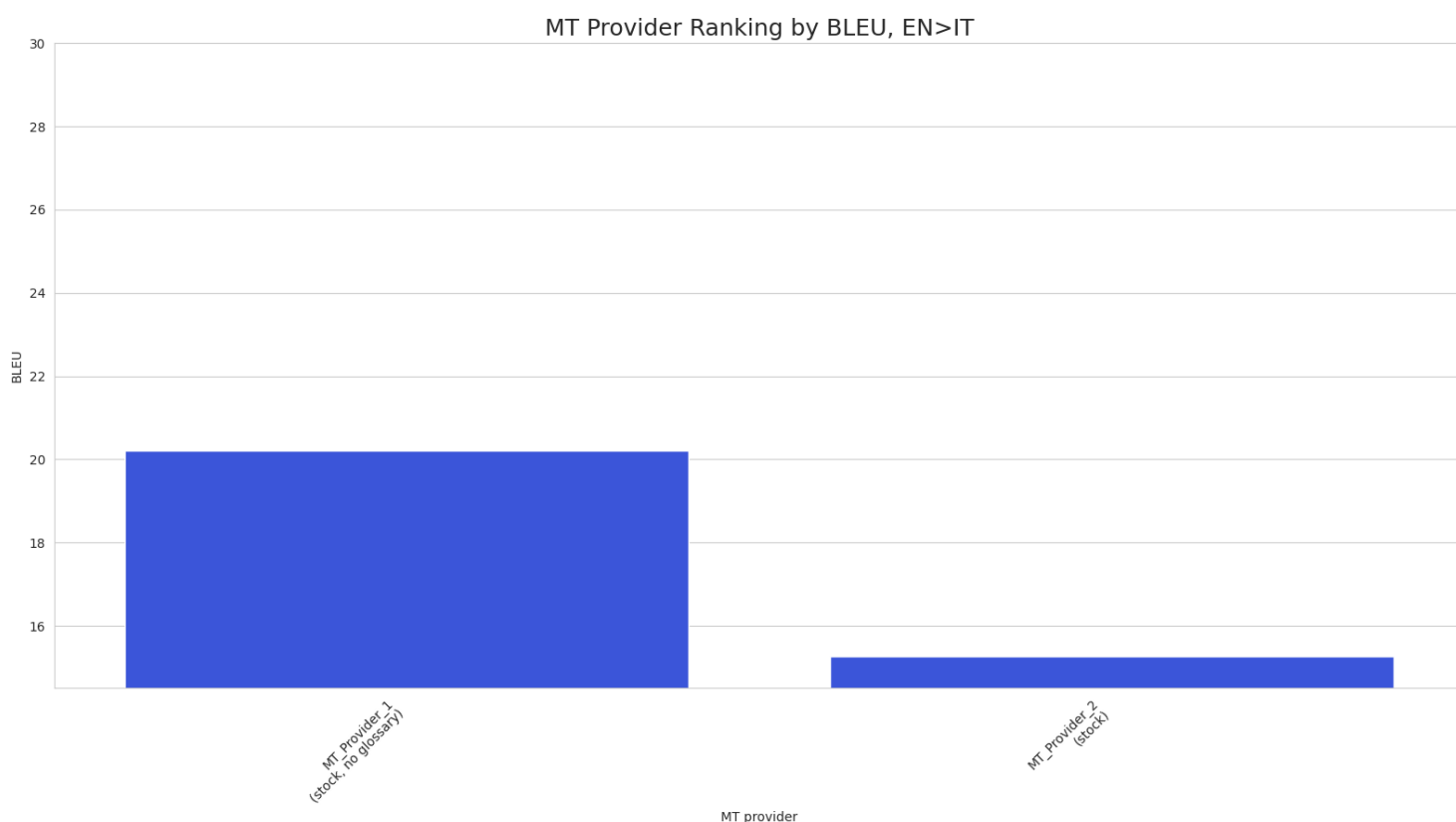
The translations were then evaluated using the metrics already used in the previous section. As already done previously, the main metrics used were BLEU and TER, since they are the most used metrics and they have been demonstrated to be highly comparable. However, all the other metrics will be presented, even though the results will not be analyzed in detail. Again, the reference translation was the one performed by Alessio Ceni, already used in the first part of the study. The MT_Provider_1 refers to Modern MT, while MT_Provider_2 refers to RWS Language Weaver. I will first provide a general view of the results before analyzing them in detail.

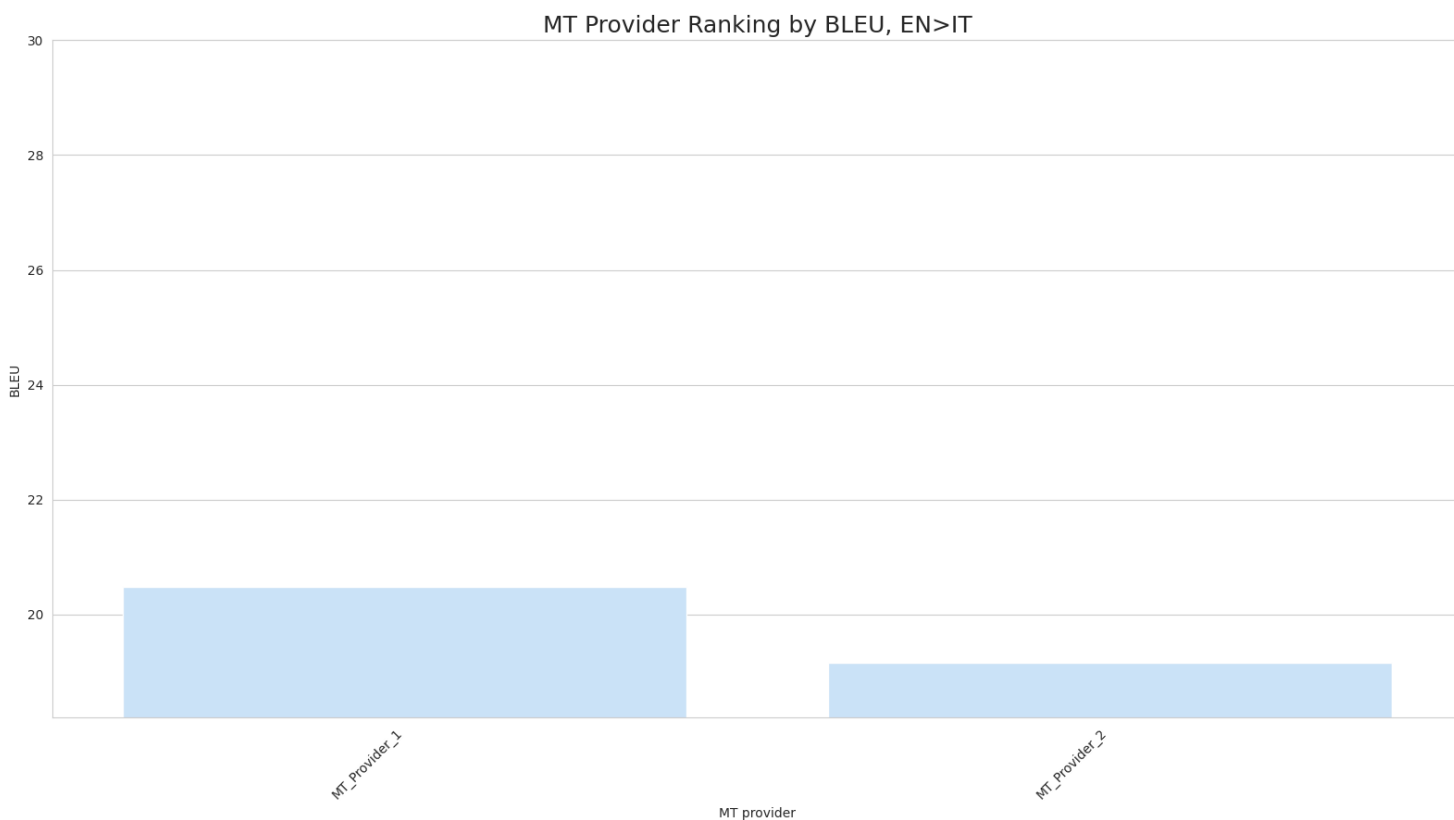
Analyze <small>Select 2 Models to analyze MT customization results. What is this?</small>							<small>+ Add models</small>	<small>↺</small>
MT Models	BERTScore <small>ⓘ</small> ↓	hLEPOR <small>ⓘ</small>	TER <small>ⓘ</small>	BLEU <small>ⓘ</small>	chrF++ <small>ⓘ</small>	COMET <small>ⓘ</small>		
<input type="checkbox"/> MT_Provider_2	0.811338 0.802682 - 0.820312	0.51 0.453 - 0.58	67.55	19.16	45.03	0.68 0.655 - 0.713		
<input type="checkbox"/> MT_Provider_1	0.809768 0.800638 - 0.821245	0.47 0.415 - 0.529	70.20	20.48	44.87	0.71 0.67 - 0.752		

Now the same results will be explored in detail, using the information contained in the downloadable analysis report produced by ModernMT Suite. As already said before, only the results referring to BLEU and TER underwent a detailed analysis.

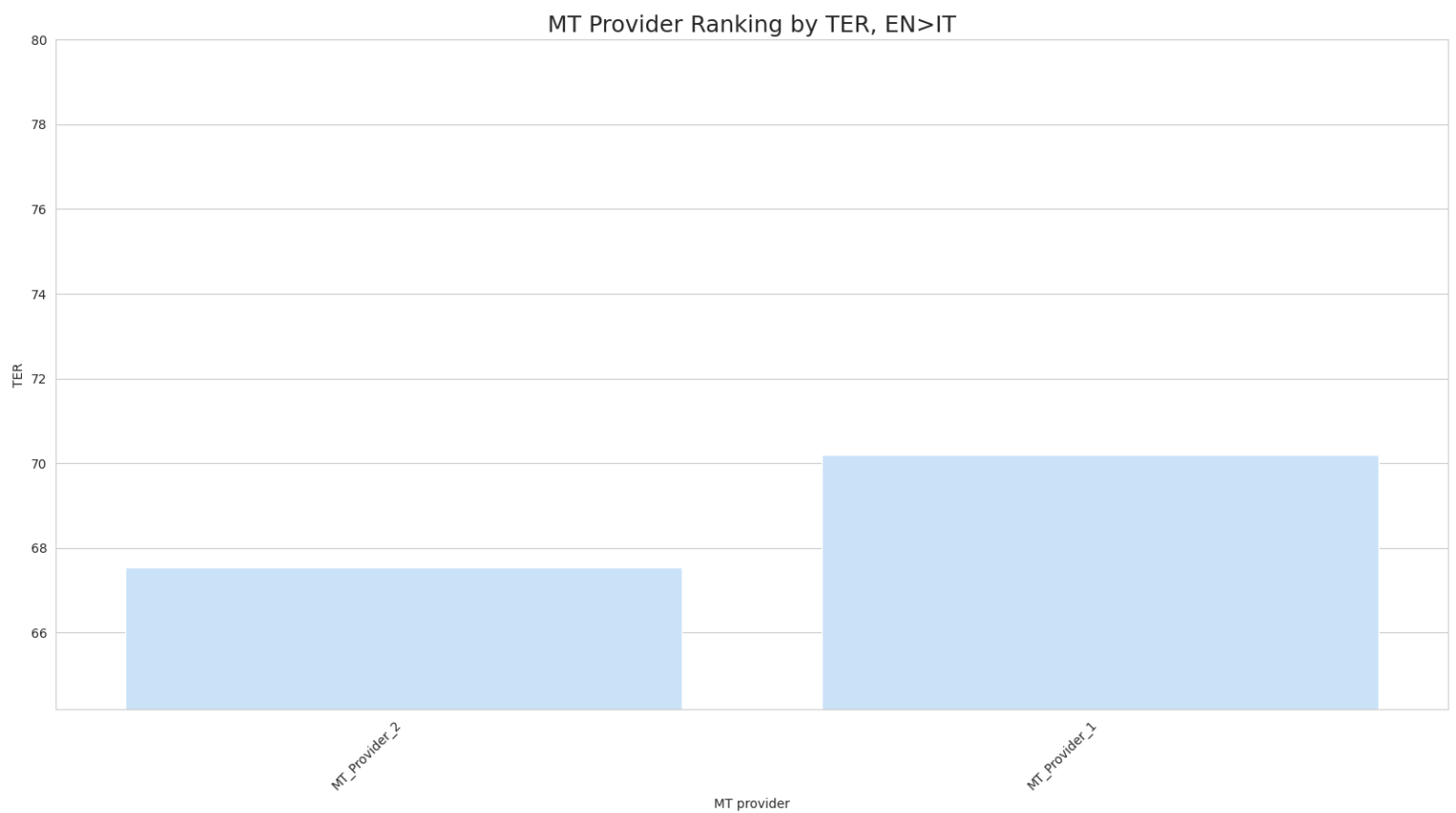
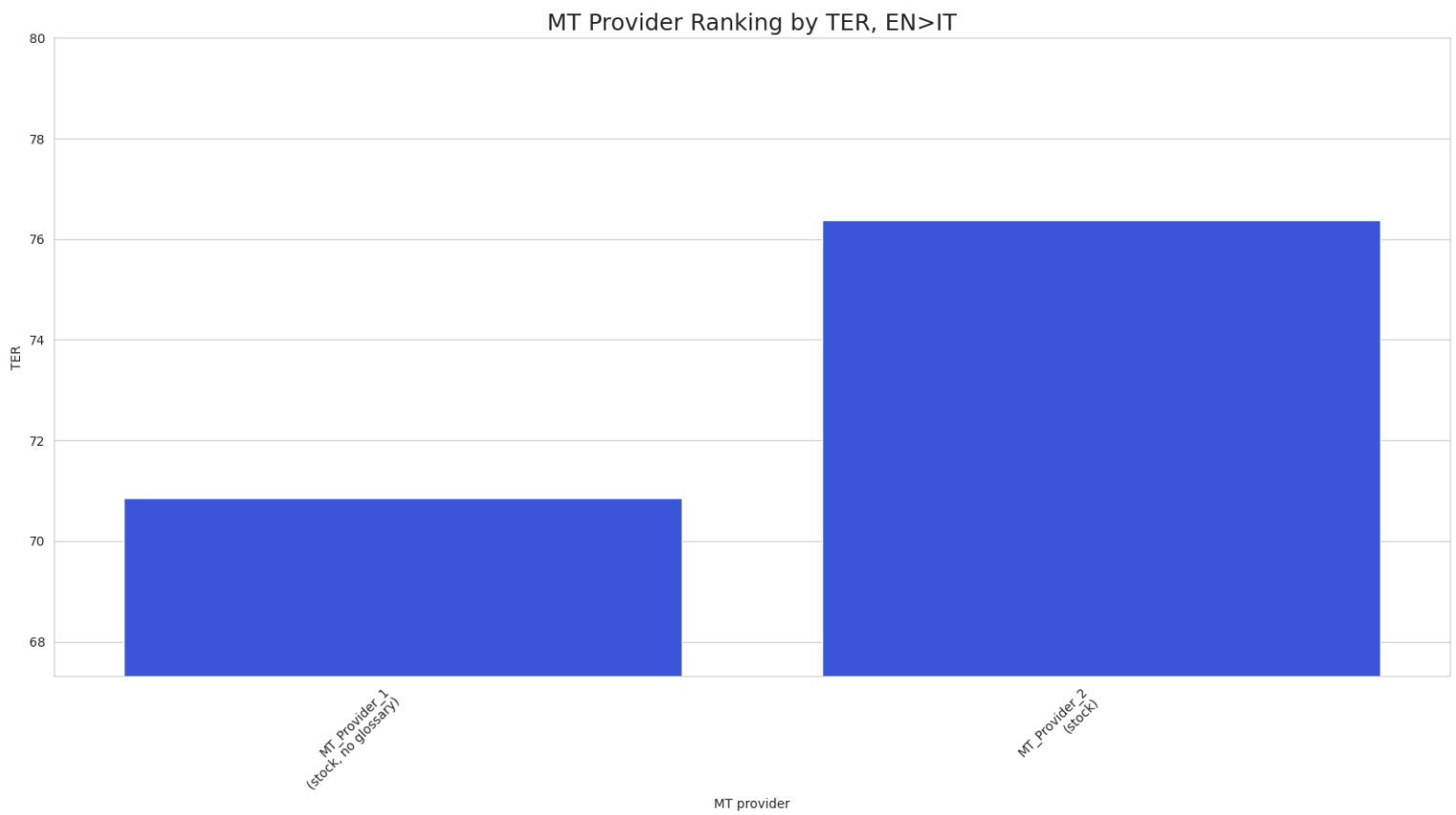


First of all, it is interesting to notice that, according to BLEU, ModernMT performed better, while according to TER better translations were provided by RWS Language Weaver. It is possible that the improvement in the use of verbs and modalities is at the basis of a Target Text closer to a text a native speaker would produce: therefore, the translations produced by Language Weaver might require less changes to get closer to the reference translation than the ones produced by ModernMT. This result seems to support the hypothesis that RWS Language Weaver is more receptive to language training than Intento ModernMT. As already observed thanks to the manual scoring of the translated segments, RWS registered a sensible improvement in syntax. Before focusing on the evaluation of the single segments provided by TER, the results of this metric evaluation will be compared to the previous results obtained. The first image refers to the first scoring, while the second image refers to the post-training scoring.





Taking into account the results obtained with the BLEU metric, it is noticeable that both of the providers improved their scores. Moreover, we could notice that Language Weaver was interested by a greater improvement than ModernMT, again corroborating the hypothesis that Language Weaver seems to react better to the training.



Looking at the results obtained with TER, it was possible to identify the same exact situation observed previously: both of the providers performed better, but RWS Language Weaver on its turn achieved a better score than Intento ModernMT.

As done previously, the scores obtained by each segment according to TER were analyzed and then compared to the previous scores.

segment_id	MT Provider 1	MT Provider 1 TER	MT Provider 2	MT Provider 2 TER
1	"Hai promesso di raccontarmi la tua storia, sai," disse Alice, "e perché è che odi - C e D," aggiunse in un sussurro, quasi temendo che si offendesse di nuovo. "La mia è una storia lunga e triste!" disse il Topo, voltandosi verso Alice e sospirando. "È una coda lunga, certamente", disse Alice, guardando con meraviglia la coda del topo; "ma perché lo chiami triste?" E continuava a confonderlo mentre il Topo parlava, in modo che la sua idea del racconto fosse qualcosa del genere: -	77,91411043	— Mi hai promesso di raccontarmi la tua storia, sai, — disse Alice, — e perché tu odi - C e D - aggiunse in un sussurro, per metà paura che si sarebbe offeso di nuovo. La mia è una storia lunga e triste! Disse il topo, volgendosi ad Alice e sospirando. E' certamente una coda lunga, — disse Alice, guardando con stupore la coda del topo; — ma perché la chiami triste? E continuava a sconcertarlo mentre parlava il topo, cosicché la sua idea del racconto era simile a questa	78,52760736
2	Poi arrivò una voce arrabbiata - quella del Coniglio: "Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima: "Certo, allora sono qui! Alla ricerca di mele, il tuo onore!" Scavare per le mele, davvero! - disse il Coniglio con rabbia. "Ecco! Vieni ad aiutarmi!" (Suoni di altri vetri rotti.) "Ora dimmi, Pat, cosa c'è nella finestra?" "Certo, è un braccio, tuo onore!" (L'ha pronunciato "arrum.") "Un braccio, oca! Chi ne ha mai visto uno così grande? Perché, riempie tutta la finestra!"	67,52136752	Poi venne una voce arrabbiata, quella del Coniglio, — Pat! Pat! Dove sei? E poi una voce che non aveva mai sentito prima, "certo che sono qui! Scavare per le mele, vostro Onore! - Scavare per le mele! Disse il Coniglio arrabbiato. Ecco! Vieni ad aiutarmi! (Suoni di vetro più rotto.) - Dimmi, Pat, che c'è nella finestra? — Certo, è un braccio, vostro Onore! (Lo pronunciò "arrum"). Un braccio, oca! Chi ne ha mai visto uno così grande? Perché, riempie tutta la finestra! — Certo, è vero, vostro onore; ma è un braccio per tutto questo.	65,81196581
3	Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo a fare indovinelli che non hanno risposte." "Se conoscessi il tempo come lo conosco io", disse il Cappellaio, "non parlerei di sprecarlo. È lui."	60,86956522	Alice sospirò stanca. "Penso che col tempo potresti fare qualcosa di meglio", disse, "che sprecarlo nel chiedere indovinelli che non hanno risposte." Se tu conoscessi il tempo come me, — disse il Cappellaio, — non parlerei di sprecare il tempo. È lui.	60,86956522
4	"Verissimo", disse la duchessa: "fenicotteri e senape mordono entrambi. E la morale di questo è: "Gli uccelli di una piuma si affollano	83,33333333	"Molto vero", disse la duchessa: "I fenicotteri e la senape mordono entrambi. E la morale di questo è: "Gli uccelli di un gregge di piume insieme".	83,33333333
5	"Dà la tua testimonianza," ripeté il Re con rabbia, "o ti farò giustiziare, che tu sia nervoso o meno." "Sono un pover' uomo, Vostra Maestà," iniziò il Cappellaio, con voce tremante, "e non avevo iniziato il mio tè - non più di una settimana o giù di lì - e con il pane e il burro che diventavano così sottili e lo scintillio del tè..." Il luocichio di cosa? disse il Re. "È iniziato con il tè," rispose il Cappellaio. "Certo che lo scintillio inizia con una T!" disse bruscamente il Re. "Mi prendi per un	62,13592233	— Dà la tua prova, — ripeté il Re arrabbiato, — o ti farò giustiziare, che tu sia nervoso o no. Io sono un povero uomo, vostra Maestà, — cominciò il Cappellaio con una voce tremante, — e non avevo iniziato il mio tè, non più di una settimana, e che con il pane e il burro che diventavano così sottili, e lo scintillio del tè... - Il scintillio del tè? Disse il Re. - Cominciava col tè - rispose il Cappellaio. — Certo che lo scintillio comincia con una T! Disse il Re con decisione. - Mi porti per un dunce? Avanti!	51,45631068

First of all, it is interesting to notice how segment 3 and 4 received the same score. This scoring marked an improvement for RWS Language Weaver: if previously it scored worse than Intento Modern MT, after the training it managed to score the same. Moreover, it should be noticed that, according to TER, Language Weaver performed better in 2 out of 5 segments. The only segment where TER assigned a

higher score to ModernMT is the first segment. However, the difference between the two scores is not high, since it is roughly more than 1 point.

The scoring obtained before the training will be now presented again in order to compare it to the second scoring.

segment_id	MT_Provider_1 (stock, no glossary)	MT_Provider_1 (stock, no glossary) TER	MT_Provider_2 (stock)	MT_Provider_2 (stock) TER
1	"Hai promesso di raccontarmi la tua storia, sai", disse Alice, "e perché odi - C e D", aggiunse in un sussurro, quasi temendo che si offendesse di nuovo. "La mia è una storia lunga e triste!" disse il Topo, voltandosi verso Alice e sospirando. "È una lunga coda, certamente", disse Alice, guardando con meraviglia la coda del topo; "ma perché lo chiami triste?" E continuava a confonderlo mentre il Topo parlava, in modo che la sua idea del racconto fosse qualcosa del genere: -	76,68711656	"Hai promesso di raccontarmi la tua storia, sai," disse Alice, "e perché sei tu ad odiare - C e D", ha aggiunto in un sussurro, mezzo timore che si sarebbe offeso di nuovo. "La mia è una lunga e triste storia!" Disse il mouse, girandosi verso Alice, e sospirando. "È certamente una coda lunga", disse Alice, guardando in basso con stupore la coda del mouse; "ma perché la chiami triste?" E ha continuato a sconcertarlo mentre il mouse stava parlando, in modo che la sua idea della storia fosse qualcosa di simile a questa: -	78,52760736
2	Poi arrivò una voce arrabbiata - quella del Coniglio: "Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima: "Certo, allora sono qui! Alla ricerca di mele, tuo onore!" "Scavare per le mele, davvero!" disse il Coniglio con rabbia. "Ecco! Vieni e aiutami a venirme fuori!" (Suoni di altri vetri rotti.) "Ora dimmi, Pat, cosa c'è nella finestra?" "Certo, è un braccio, tuo onore!" (Lo pronunciò "arrum.") "Un braccio, oca! Chi ne ha mai visto uno così grande?" Perché, riempie tutta la finestra! "Certo che sì, il tuo onore: ma è un braccio per tutto questo."	69,23076923	Poi è arrivata una voce arrabbiata - quella del Coniglio - "Pat! Pat! Dove sei?" E poi una voce che non aveva mai sentito prima, "certo che sono qui! Scavare per le mele, vostro onore!" "Scavare per le mele, infatti!" Disse il Coniglio con rabbia. "Qui! Vieni ad aiutarmi!" (Suoni di vetro più rotto). "Ora dimmi, Pat, che cosa c'è nella finestra?" "Certo, è un braccio, vostro onore!" (Lo ha pronunciato "arrum".) "Un braccio, tu oca! Chi ne ha mai vista una	74,35897436
3	Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo," disse, "piuttosto che sprecarlo a fare indovinelli che non hanno risposte." "Se conoscessi il tempo come me", disse il Cappellaio, "non parlerei di sprecarlo. È lui."	65,2173913	Alice sospirò stancamente. "Penso che potresti fare qualcosa di meglio con il tempo", ha detto, "piuttosto che sprecarlo nel chiedere indovinelli che non hanno risposte". "Se conoscessi il tempo come me", disse il Cappellaio, "non parlerei di sprecarlo. È lui."	67,39130435
4	"Verissimo", disse la duchessa: "fenicotteri e senape mordono entrambi. E la morale è:" Gli uccelli di una piuma si affollano " "	79,16666667	"Molto vero", disse la duchessa: "Fenicotteri e senape mordono entrambi. E la morale di questo è: "Gli uccelli di una	87,5
5	"Dà la tua prova", ripeté il Re con rabbia, "o ti farò giustiziare, che tu sia nervoso o meno." "Sono un pover' uomo, Vostra Maestà," iniziò il Cappellaio, con voce tremante, "e non avevo iniziato il mio tè - non più di una settimana o giù di lì - e con il pane e il burro che diventavano così sottili e lo scintillio del tè..." "Il luccichio di cosa?" disse il Re. "È iniziato con il tè", rispose il Cappellaio. "Certo che lo scintillio inizia con una T!" disse bruscamente il Re. "Mi prendi per un somaro? Vai avanti!"	64,0776639	"Date le vostre prove", ripeté il Re con rabbia, "o vi farò giustiziare, che siate nervosi o no." "Sono un povero uomo, vostra Maestà", ha iniziato il Cappellaio, con una voce tremante, "e non avevo iniziato il mio tè - non più di una settimana circa - e con il pane e il burro che diventano così sottili - e il scintillio del tè..." "Il scintillio del cosa?" Ha detto il re. "È iniziato con il tè", rispose il Cappellaio. "Naturalmente il scintillio inizia con una T!" Ha detto il re con rabbia, "o vi farò giustiziare, che siate nervosi o no."	76,69902913

First of all, it is possible to notice that the scoring for segment 1 remained unchanged as far as RWS Language Weaver is concerned, while Modern MT received a worse score after the training. However, the training gave very good results, since all the scores improved for all the other segments. It is impressive to notice how positively the training affected Language Weaver: for example, passage 5 was originally scored 76, 69 points, but after the training it obtained a score of 51, 45 points. Such impressive changes did not affect Modern MT, except for passage 4, the scoring of which improved of 5 points after the training.

MT AND AI SECOND COMPARISON

As done previously, a table containing all the results will be presented.

	ChatGPT 3.5	Microsoft Edge Copilot	RWS Language Weaver	Intento Modern MT
Severe meaning error	0	1	6	4
Minor meaning error	1	0	0	1
Unknown word or segmentation error	0	0	2	1
Consistency/Term translation error	0	0	0	1
Pronoun resolution error	0	0	0,5	0,5
Locution error	0	0	2	1
Omission, insertion, repetition errors	0	0	0	0
Severe syntax error	0	0	2	0
Minor syntax error	1	0	4	1
Tone/register error	0	0	0	0
Total Number of errors	2	1	16,5	9,5

The table presented above confirms what was observed after the first phase of the study: generative Artificial Intelligence (in this study represented by ChatGPT and Copilot) performs better than Neural Machine Translation (in this study represented by Intento ModernMT and RWS Language Weaver) when it had to tackle the translation of puns and wordplay. It is important to bear in mind that all of the machines used for the study are based on Artificial Intelligence; however, while

GPT and Copilot represent two examples of generative AI, Language Weaver and ModernMT are not identified as generative AI.

There are some positive results that are worth sharing. First of all, the number of errors committed by generative AI was drastically reduced thanks to prompting. ChatGPT 3.5 scored only 2 errors, while Microsoft Edge Copilot scored just 1. It is important to highlight that Copilot scored 1 Severe meaning error, while ChatGPT scored a Minor meaning error and a Minor Syntax error. Therefore, regarding the quality of the errors and not just their number, ChatGPT 3.5 performed better and it is the system that performed better of all of the 4, followed by Copilot.

Machine Translation providers also improved thanks to MT training. Modern MT performed better than Language Weaver, even though it is worth noting that both of the providers scored less error in this second phase of the evaluation. However, no variation has been observed regarding the Meaning errors, both Severe and Minor. This means that MT training might not be the right path to follow in order to improve the performance of MT providers when applied to puns and wordplay. However, training improved other aspects of the translation, such as Syntax. Thanks to MT training, both RWS Language Weaver and Intento ModernMT provided translations with a better syntax. Language Weaver also obtained a lower score as far as Locution errors are concerned.

The last chapter of the thesis will be dedicated to the conclusions. In this chapter, the results obtained will be taken into consideration in order to understand whether or not the research question has been answered to. Moreover, possible paths of research will be presented.

Conclusions

In this final part of the dissertation, the results obtained will be summarized and analyzed in order to understand whether or not the research question of this paper has been answered to. After briefly reproposing the research question and illustrating the main challenges associated with the main object of research, the results will be presented and analyzed. While discussing the results, the main problems and issues faced during the actual research phase will be described. After that, the results obtained will be used as a basis to present possible further research paths.

The main research question of the paper was the following: “Is it possible to use generative Artificial Intelligence and Machine Translation as an aid to literary translators when it comes to translating complex textual items such as pun and wordplay? Are there strategies directly applicable by translators that could improve the performance of these tools?”. As already discussed in the part of this paper dedicated to the literary review, puns, wordplay and in general humorous devices represent a challenge to translators. This kind of textual instances are characterized by very specific features, that make them very different from other textual elements. In order to tackle the translation of such textual instances, translators need to have a very good knowledge of the grammar, syntax, phonetics and pragmatics of the Source Language and of the Target Language as well. Puns and wordplay rely on a series of linguistic and non-linguistic elements, such as the language used, the context of the humorous instance, or the logical mechanism that makes the joke work well. If we look at one of the jokes analyzed it is immediately clear how translating humorous instances might be challenging. The pun “Mustard and flamingoes both bite” requires a detailed knowledge of the English language in order to be understood and translated. The punning mechanism centers on the verb “to bite”, which is in English is used to describe both the biting action of the flamingoes and the biting taste of mustard. The main issue in this case is that the Italian language does not allow the direct equivalent of this verb (“mordere”) to describe both the biting action of the flamingoes and the biting taste of mustard: it

would be necessary to use two different verbs. At this point, a translator might decide to adopt a Pun-to-non-pun strategy: that is to say, they might decide not to translate the pun. However, in the case in which a translator would want to translate this passage reproposing a pun, they would be facing a challenge. Since it is not possible to resort to the direct equivalent, translators are forced to adapt other strategies, such as proposing another pun, or structure a new joke specifically designed to suit the Target Text.

Having demonstrated that puns, wordplay and other humorous instances in texts represent a challenge to translators, it comes natural to try understanding whether or not translators could be helped in their work when it comes to translating such elements. If in certain texts humor is a secondary element and it can be easily overlooked, there are texts that make humor its central element, such as the novel “Alice’s Adventures in Wonderland” by Lewis Carroll. It might be stated that humor, puns and wordplay are the most important elements of the book, alongside with the plot. Therefore, it is not possible to provide a translation of the novel that completely overlooks its humorous instances, since it would be a really bad translation. This is the reason why the novel has been chosen in order to carry out the research. Five instances of humorous language (puns and wordplay) were extracted from the book for the analysis.

The tools that were tested are generative Artificial Intelligence and Neural Machine Translation. In order to carry out the research I decided to use two interfaces that allow to interact with generative Artificial Intelligence and two providers that implement Neural Machine Translation. These tools were selected in order to test the latest technologies that are used by translators in their everyday professional life. The two generative AI interfaces selected were ChatGPT 3.5 and Microsoft Edge Copilot, and the two Neural Machine Translation providers selected were Intento ModernMT and RWS Language Weaver.

In order to carry out the research the first phase consisted in the translation of the selected segment by every one of the above-mentioned tools. After that, the translations were scored both manually and automatically. The manual evaluation was carried out on the basis of a set of criteria established by Matusov, while the automatic evaluation was carried out using the Intento ModernMT Suite, that allowed to process the translations and score them with different metrics. Although

all of the results obtained by the metrics were presented, only the scores provided by BLEU and TER were discussed. It is important to point out that only the translation performed by Machine Translation providers were evaluated also by the metrics; translated texts generated by generative AI were only scored manually. Manual scoring was used for different reasons: first of all, the impossibility to apply the metrics developed for MT to generative Artificial Intelligence required a form of evaluation that could be used to score all of the four tools in order to compare the results; secondly, manual evaluation was necessary in order to analyze in detail the translation, since it can help spotting error and problems that the automatic metrics cannot identify. For example, while the metrics recorded a general improvement in the performance of RWS Language Weaver after the training, the manual scoring allowed to point out that the improvements interested Syntax and Locution, and not the actual object of research. The scoring (both manual and automated) demonstrated that generative AI and NMT fail to provide acceptable translations for humorous instances in texts. The scoring also demonstrated that generative AI performed better than Neural Machine Translation. The results of the first scoring showed that the system that performed better was Copilot, followed by GPT3.5; RWS Language Weaver performed the worst, while Intento Modern MT performed better than Language Weaver, but worse than Copilot.

The second phase consisted in finding strategies that could be directly applied by translators in order to improve the performance of the selected tools. Therefore, this second phase focused on the possibility of human intervention on the chosen tools in order to improve their performance. As far as generative Artificial Intelligence is concerned, it is possible to interact with the engine through a series of prompts that can be sent using the interface. The prompts might be used to ask the engine to focus only on certain passages of the text, or to ask it to use specific words, or writing styles. Therefore, the prompting strategy was tested in order to understand whether or not correct prompts might actually improve the quality of translation provided by AI. When it comes to Neural Machine Translation, it is possible to train Machine Translation engines using Translation Memories specifically created for the purpose. These memories contain segments extrapolated by texts that refer to a specific domain. For example, a TM made by extrapolating segments from contracts, laws, agreements, and regulations might be used to train a Machine Translation provider in order to use it to translate legal texts. Since the main object

of research were puns and wordplay, the TM was created extrapolating more than 15,000 segments from literary texts (the text being mainly novels and short stories written for children and young adults). Because of the specificities of these instances of humorous language, creating a TM made only of puns, wordplay and jokes would have been impossible. It is not rare that translators resort to strategies different from literal translation when tackling these textual elements. Therefore, the TM was created by incorporating segments from different literary texts, including segments that contained puns, wordplay and examples of humor. The main idea behind this TM was that of creating a Memory that would train the Machine Translation provider to translate a literary text that included instances of humorous language. After the application of these strategies, the translated segments were checked and evaluated again, in order to identify any possible improvement from the previous phase.

Generative AI prompting gave the best results. This whole process consisted in establishing a dialogue with the interface, in order to obtain suitable translation results. After this phase, it could be possible to identify some prompting strategies that could be applied by translators in order to obtain acceptable results. First of all, it is important to point out that ChatGPT 3.5 managed to obtain the best results: after the prompting session all the translations scored 0 Severe meaning errors, while Microsoft Edge Copilot performed worse, with the translation of the first segment that scored 1 Severe meaning error. As far as the first segment is concerned, Copilot did not manage to provide a suitable translation, even after the prompting session. Another segment where GPT 3.5 performed better than Copilot is segment 2; here, GPT 3.5 immediately provided a good solution for the word “arrum”, while Copilot did not do the same. Apart from segment 1, all the other translations performed by Copilot were acceptable. The main finding that emerged is the necessity to provide the system with a detailed context: the more the context is detailed, the more the chances of a suitable translation are high. For example, if the first prompt specified that the system needs to pay attention to puns and wordplay in the segment, it was possible to obtain just with one prompt a result that was better than the one obtained without any prompt. Looking at the translation provided by GPT 3.5 for the first segment, it is clear that humorous instances need to be explained: once the meaning and mechanisms of the joke have been explained to the machine, it is possible to observe improvements in the translation. Among

the suggestions that the translator could give the machine, there is not only information about the context, but also information about the word (for example the verb) that would be advisable to use. Another prompting strategy that could be employed is asking the system to recreate the joke, providing a translation that uses a different joke from the original. This was done with ChatGPT 3.5 in the second segment analyzed, obtaining a good result. The screenshot showing the final result will be presented again. As it can be seen in the screenshot, Pat is not speaking anymore using an Irish pronunciation, but rather he is using a form of Italian language heavily influenced by the regional dialect he speaks. Moreover, the reference to the apples has been removed, since it would have not made any sense at all in the Italian text. The sentence “digging for apples” has been translated into the Italian equivalent of “digging for potatoes”, which is an activity that an Italian countryman would carry out in their garden.



Of course! I'll adjust Pat's activity to something more fitting for an Italian context, while maintaining his rustic, dialect-influenced speech. Let's have Pat involved in a more relatable, countryside task, such as digging for vegetables like potatoes or something similar. Here's the revised version:

Subito dopo, si senti una voce arrabbiata—quella del Coniglio—«Pat! Pat! Dove sei?» E poi una voce che Alice non aveva mai sentito prima, «Eh, so' qui, vostro onore! Sto scavanno pe' le patate!»
 «Scavando per le patate, davvero!» disse il Coniglio con rabbia. «Vieni qui! Aiutami a uscire da questo pasticcio!» (Suoni di altro vetro che si rompe.)
 «Ora dimmi, Pat, che cos'è quella cosa alla finestra?»
 «Ma certo, è un bbraccio, vostro onore!» (Lo pronunciò “bbraccio”).
 «Un braccio, stupido! Chi ha mai visto un braccio di quelle dimensioni? Ma se riempi tutta la finestra!»
 «Certo che sì, vostro onore: ma è un bbraccio lo stesso.»

Here, I've replaced the apples with Pat digging for *patate* (potatoes), a common and realistic task for a rustic Italian character, especially one living in a rural area. This makes his activity more plausible within an Italian-speaking context while preserving the tone and humor of the original text.

Machine Translation training did not obtain good results as far as the research question is concerned. Even after the training, all the Severe and Minor meaning errors remained. However, as it was illustrated, the training managed to solve other issues, leading to results that were not expected. In order to train Intento ModernMT and RWS Language Weaver a Translation Memory was manually

created using the Alignment function made available on Trados Studio 2021. After that, the data were cleaned and used to train the two Machine Translation providers. RWS Language Weaver was trained using the Adapted Language Pair function made available on the online interface; ModernMT was trained via the training option made available on the MT Hub in the Intento ModernMT Suite, then the online interface of Intento ModernMT was used to translate a second time the segments using the trained provider. As already mentioned above, the training did not manage to improve the translations of puns and wordplay, but considerable improvements in Syntax and Locution were observed. The error category that improved the most was Syntax, demonstrating that the linguistic data gathered from literary texts were useful in order to train the MT providers. Even though the training did not improve the quality of the textual instances that were the object of the research, it is still interesting to notice that the quality of literary translation *per se* actually improved. What is even more interesting, is comparing the improvements registered by the two providers. As far as Intento ModernMT is concerned, only a few improvements were recorded; moreover, most of the segments were translated almost in the same way even after the training. The same cannot be said of Language Weaver. It was possible to observe massive improvements in the translations provided by this system. This observation was reinforced by the metrics and, specifically, by TER. The metric discussed now is really interesting, because it measures the number of changes required in order to make a translated text close to a golden standard reference. After the training, the scores assigned to Language Weaver drastically fell, signaling that a considerable inferior number of changes was required if compared to the number of changes required before the training. This result, though unexpected, is quite interesting because it demonstrated that Language Weaver apparently responds better to training than ModernMT.

As already mentioned above, this research lead to interesting results. First of all, it has been demonstrated that a series of correct prompts could be of great use to literary translators when tackling puns and wordplay. Then, the results opened to the possibility of a further analysis of the effect of training on Machine Translation providers. Some issues still need to be pointed out. First of all, training has been demonstrated not to be a good method to tackle translation of puns and wordplay. Even after the training massive post editing on the part of the translator

was required in order to obtain a suitable and acceptable translation. Therefore, more research is needed from this point of view. A major issue experienced when creating the Translation Memory for the training consists of the difficulty in finding instances of humorous language suitable for the creation of a training TM. As already discussed above, usually humorous instances are not translated literally, but the translators need to resort to many different strategies. Sometimes these instances are not translated at all, or in some other cases they are completely transcreated; there might also be cases in which the kind of humorous language is the same, but the execution of the joke is different. To express it in better words, it is possible that an English pun is translated into an Italian pun, but the translator might decide to create a completely new pun, or they might decide to center the pun on a different punning mechanism than the original. For these reasons, a Translation Memory made only of puns, wordplay and other instances of humor is difficult to create: not only would it take a great amount of time, but it would be necessary to check carefully every segment of the memory before deciding to use it for the training. However, this might be an interesting path to follow for future research. It would be possible to create such a Translation Memory, purposefully inserting different kinds of humorous instances in order to train the system on different translation strategies adopted by human translators. A possible way to proceed might be deciding to use in equal parts specific types of humorous instances, paired with translations obtained using different translation strategies. This might open to the possibility of checking whether or not MT providers could be sensitive to learning how to employ different translation strategies in different contexts. Another interesting possibility when it comes to MT training consists of the possibility of checking how MT providers react to the training with different kinds of texts. This research would be interesting in order to understand whether a general training TM could be suitable for different literary genres, or if specific Translation Memories should be created. Following this idea, it might be also interesting to test how MT providers react to training based on different language pairs. As already discussed, RWS Language Weaver reacted better than *Intanto ModernMT* to training was quite an unexpected result; however, this opens to the possibility of a research focused on the comparison of different systems before and after training. Since literary translation requires high levels of cultural knowledge and creativity, it might be very interesting to train and evaluate different models on literary translation.

Another issue experienced during the research has to do with the second phase of research. After training ModernMT using the interface MT Hub available on Intento ModernMT, it was not clear if the trained system could be used directly on that interface. Therefore, it was necessary to use the web interface of Intento ModernMT in order to be sure to use the trained model. Probably, this aspect might be improved, in order to stay on the same interface and avoiding switching MT provider interface during the research. Another possibility would be making smoother and clearer the process to use the trained interface directly in the MT Hub.

Generative AI prompting did not cause many problems and a few issues were met. However, it would be interesting to go further with the research on prompting, in order to identify specific types and categories of prompts that could be used by translators. What has been interesting to analyze was how the two different systems responded differently during the prompting session. After identifying and classifying the different prompts, it might be interesting to apply the same type of prompts in different systems, in order to understand how the systems react. This might be a way to develop a sort of standardized testing for generative AI interfaces: basing on the number of prompts of the same type, it could be possible to observe not only if the system manages to provide a suitable solution, but also how many prompts are needed for different systems to provide an acceptable solution. This testing method could be directly related to the research on explainable AI, that was briefly discussed in the first part of this thesis. As pointed out, one of the main problems with state-of-the-art Artificial Intelligence consists of the difficulties of explaining how these systems take decisions. A standardized test based on prompt types might be a good way to understand how different AI systems work in order to translate a text. For example, it could be possible to categorize prompt types as follows: contextual prompts (for those prompts that aim at providing better context), specific word prompts (for those prompts that aim at explaining in detail which kind of word items the system is asked to use), transcreation prompts (for those prompts that ask the system to recreate a passage)... Such a standardized system might also allow to understand how different systems react to different types of prompts; on its turn, this might help understanding if certain AI interfaces might respond better to specific prompt types.

According to the present study, it seems that one of the main difficulties for Machine Translation is that of identifying the instances of humorous language. In order to

provide a good translation for a pun, it is necessary to identify it as a pun. It would be a good idea that of trying to find a way to support Machine Translation in the process of pun identification. For example, it might be possible to find a way to integrate the tool PunCAT (discussed in the Literature review) into Machine Translation providers. As explained in the Literature Review, PunCAT was developed in order to identify puns in a text, interpreting them and providing possible translation solutions. If it was possible to integrate this tool into MT providers, they would become able to identify puns and to apply specific types of translation when tackling these phrases. Moreover, it might be interesting to develop plugins for Trados Studio specifically designed to identify puns and other instances of humorous language in a text, in order to help translators in identifying them.

The present study gave rather satisfactory results. It demonstrated that generative Artificial Intelligence is a valid ally to translators when tackling instances of humorous language such as puns and wordplay. Nevertheless, it requires high levels of human interaction in order to work well with this kind of textual instances. The same unfortunately cannot be said for Machine Translation. These systems are not able to identify and translate properly puns, not even after a specific training; however, it could be possible to adopt different kinds of training in order to study how MT providers respond to them. The research on this aspect might be expanded and there still is much more that can be done. Despite this aspect, MT training still managed to obtain good results as far as Syntax is concerned, demonstrating that MT could indeed be trained in order to translate literary texts. One of the main issues of this study is the limited quantity of data analyzed. In order to carry out the study, only five instances of humorous language were chosen, translated and analyzed. It is possible that an analysis conducted using much more linguistic data will lead to other results. Moreover, it would be even possible to analyze other instances of humorous language, without being limited to puns and wordplay.

Even though this research did not produce all the expected results, the findings are still very interesting, and open to the possibility of multiple research paths. The world of literary translation is variegated, and the translation of puns and wordplay is just a small aspect of it. The results obtained are very interesting and encouraging, because they demonstrated that generative Artificial Intelligence and Neural

Machine Translation might be helpful tools for literary translators. For sure, their full potential will be expressed if more studies on this topic are carried out.

Bibliography

Alkodimi, Khaled; Alqahtani, Osama Abdulrhman; Al-Wasy, Baleigh Qassim, *Human-AI collaboration in translation and back translation of literary texts*, in “Journal of Social Studies”, Volume 30, Issue 2, July 2024, University of Science and Technology, Yemen, 173-192

Ardi, Havid; Al Hafizh, Muhd; Rezqi, Iftahur; Tuzzikriah, Raihana, *Can Machine Translations Translate Humorous Texts?*, in “Humanus”, Volume 21, Issue 1, 2022, Pusat Kajian Humaniora (Centre for Humanities Studies), FBS Universitas Negeri Padang, 99-112

Attardo, Salvatore, *Universals in Puns and Humorous Wordplay*, in “Cultures and Traditions of Wordplay and Wordplay Research”, Berlin, Boston, De Gruyter, 2018, 89-109

Bouguesmia, Mohammed Tewfik, *Using AI in Translation, a Technological Leap, or a Translator's Nightmare*, in “ALTRALANG Journal”, Volume 2, Issue 2, Faculty of Foreign Languages, University of Oran, 2020, 78-102

Chérargui, Mohamed Amine, *Theoretical Overview of Machine translation*, CEUR Workshop Proceedings, 867, 2012, 160-169

Collins, Brona; Cunningham, Pdraig; Veale, Tony, *An Example-Based Approach to Machine Translation*, in Conference of the Association for Machine Translation in the Americas, Montreal, Canada, 1996, 1-13

Delabastita, Dirk, *Introduction*, in “The translator: studies in intercultural communication”, Volume 2, Number 2, London and New York, Routledge, 1996, 127-139

Dhanani, Farhan; Rafi, Muhammad; Tahir, Muhammad Atif, *FAST-MT Participation for the JOKER CLEF-2022 Automatic Pun and Humour Translation Tasks*, in “CLEF 2022: Conference and Labs of the Evaluation Forum”, Bologna, Italy, September 5–8, 2022

- Ding, Yanzhuo; Liu, Yang; Luan, Huanbo; Sun, Maosong, *Visualizing and understanding neural machine translation*, in “Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics”, Volume 1: Long Papers, Vancouver, Canada, Association for Computational Linguistics, 2017, 1150–1159
- Ermakova, Liana et al., *CLEF Workshop JOKER: Automatic Wordplay and Humour Translation*, in “Advances in Information Retrieval. ECIR 2022. Lecture Notes in Computer Science”, Vol. 13186, Springer, 2022, 355-363
- Eszenyi, Réka; Bednárová-Gibová, Klaudia; Robin, Edina, *Artificial Intelligence, Machine Translation & Cyborg Translators: a Clash of Utopian And Dystopian Visions*, in “Ezikov Svyat”, Volume 21, Issue 2, Faculty of Philology, South-West University “Neofit Rilski”, 2023, 102-113
- Giorgadze, Meri, *Linguistic Features of Pun, its Typology and Classification*, in “European Scientific Journal”, Volume 2, European Scientific Institute, 2014, 271-275
- Görög, Attila, *Quality Evaluation Today: the Dynamic Quality Framework*, in “Translating and the Computer”, Issue 36, 2014, 155-164
- Görög, Attila, *Quantifying and benchmarking quality: the TAUS Dynamic Quality Framework*, in “Revista Tradumàtica: tecnologies de la traducció”, Issue 12, 2014, 443-454
- Gurov, Andrey, *Advances in artificial intelligence in the field of literary translation*, 2023, 1-4
- Haenlein, Michael; Kaplan, Andreas, *A brief history of artificial intelligence: On the past, present, and future of artificial intelligence*, in “California Management Review”, Volume 61, Issue 4, 2019, 5-14
- Hutchins, William John, *Machine translation: a concise history*, 2003, 1-21
- Hutchins, William John, *Reflections on the History And Present State of Machine Translation*, 1995, 1-12
- Koehn, Philipp, *Neural Machine Translation*, in “Statistical Machine Translation”, Draft of Chapter 13, 2017, 5-117

Kolb, Waltraud; Miller, Tristan, *Human-computer interaction in pun translation*, in “Using Technologies for Creative-Text Translation”, New York, Taylor and Francis, 2022, 66-88

Kuzman, Taja; Vintar, Špela; Arčan, Mihael, *Neural Machine Translation of Literary Texts from English to Slovene*, in “Proceedings of the Qualities of Literary Machine Translation”, Dublin, European Association for Machine Translation, 2019, 1-9

Lee, Tong King, *Artificial intelligence and posthumanist translation: ChatGPT versus the translator*, in “Applied Linguistics Review”, University of Hong Kong, 2023, 1-22

Lepage, Yves; Denoual, Etienne, *Purest ever example-based machine translation: Detailed presentation and assessment*, in “Machine Translation”, 2007, 251-282

Li, Aaron; Han, Feng; Wong, Derek Fai; Chao, Lidia Sam; He, Liangye; Lu, Yi; Xing, Junwen; Zeng, Xiaodong, *Language-independent Model for Machine Translation Evaluation with Reinforced Factors*, in “Proceedings of Machine Translation Summit XIV: Posters”, Nice, 2013, 215-222

Low, Peter Alan, *Translating jokes and puns*, in “Perspectives: Studies in Translatology”, Routledge, Volume 19, Issue 1, 2011, 59-70

Matusov, Evgeny, *The Challenges of Using Neural Machine Translation for Literature*, in “Proceedings of the Qualities of Literary Machine Translation”, Dublin, European Association for Machine Translation, 2019, 10-19

Miller, Tristan, *The Punster’s Amanuensis: The Proper Place of Humans and Machines in the Translation of Wordplay*, in “Proceedings of the Human-Informed Translation and Interpreting Technology Workshop”, Varna, Incoma Ltd., 2019, 57-65

Muthukrishnan, Nikesh; Maleki, Farhad; Ovens, Katie; Reinhold, Caroline; Forghani, Behzad; Forghani, Reza, *Brief History of Artificial Intelligence*, in “Neuroimaging Clinics of North America”, Philadelphia, W.B. Saunders, Volume 30, Issue 4, 2020, 393-399

Nagao, Makoto, *A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle*, in “Artificial and Human Intelligence”, New York, Elsevier Science Publishers, 1984, Chapter 11

O'brien, Sharon, *Translation as Human-Computer Interaction*, in “Translation Spaces”, John Benjamins Publishing Company, Vol. 1, Issue 1, 2012, 101-122

Okpor, Margaret Dumebi, *Machine Translation Approaches: Issues and Challenges*, in “IJCSI International Journal of Computer Science Issues”, Vol. 11, Issue 5, N° 2, 2014, 159-165

Papineni, Kishore; Roukos, Salim; Ward, Todd; Zhu, Wei-Jing, *BLEU: a Method for Automatic Evaluation of Machine Translation*, in “Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics”, Philadelphia, Association for Computational Linguistics, 2002, 311-318

Partington, Alan Scott, *A linguistic account of wordplay: The lexical grammar of punning*, in “Journal of Pragmatics”, Volume 41, Issue 9, 2009, 1794–1809

Popova, Olga; Dadić, Petra, *Does AI Have a Sense of Humor? CLEF 2023 JOKER Tasks 1, 2 and 3: Using BLOOM, GPT, SimpleT5, and More for Pun Detection, Location, Interpretation and Translation*, in “Working Notes of the Conference and Labs of the Evaluation Forum”, Thessaloniki, 2023, 1888-1908

Popović, Maja, *Class error rates for evaluation of machine translation output*, in “Proceedings of the Seventh Workshop on Statistical Machine Translation”, Association for Computational Linguistics, 2012, 71-75

Popović, Maja, *CHRF: character n-gram F-score for automatic MT evaluation*, in “Proceedings of the Tenth Workshop on Statistical Machine Translation”, Association for Computational Linguistics, 2015, 392-395

Post, Matt, *A Call for Clarity in Reporting BLEU Scores*, in “Proceedings of the Third Conference on Machine Translation: Research Papers”, Association for Computational Linguistics, 2018, 186–191

Raskin, Victor, *Linguistic heuristics of humor: a script-based semantic approach*, in “International Journal of the Sociology of Language”, Volume 1987, N° 65, 1987, 11-26

Raskin, Victor; Hempelmann, Christian F.; Taylor, Julia M., *How to Understand and Assess a Theory: The Evolution of the SSTH into the GTVH and Now into the OSTH*, in “Journal of Literary Theory”, De Gruyter, Volume 3, Issue 2, 2010, 285-312

Rei, Ricardo; Stewart, Craig; Farinha, Ana C.; Lavie, Alon, *COMET: A Neural Framework for MT Evaluation*, in “Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)”, Association for Computational Linguistics, 2020, 2685-2702

Rei, Ricardo; Farinha, Ana C.; Zerva, Chrysoula; Van Stigt, Daan; Stewart, Craig; Ramos, Pedro G.; Glushkova, Taisiya; Martins, André F.T.; Lavie, Alon, *Are References Really Needed? Unbabel-IST 2021 Submission for the Metrics Shared Task*, in “Proceedings of the Sixth Conference on Machine Translation”, Association for Computational Linguistics, 2021, 1030–1040

Risden, Edward L., *Script-based Semantic Theory of Humor and the Old English Riddles*, in “Enarratio”, Ohio State University. Medieval Association of the Midwest, Volume 8, 2001, 61-70

Rivera-Trigueros, Irene, *Machine translation systems and quality assessment: a systematic review*, in “Language Resources and Evaluation”, Volume 56, Issue 2, 2022, 593-619

Sahari, Yousef; Al-Kadi, Abdu M. Talib; Ali, Jamal Kaid Mohammed, *A Cross Sectional Study of ChatGPT in Translation: Magnitude of Use, Attitudes, and Uncertainties*, in “Journal of Psycholinguistic Research”, Volume 52, Issue 6, 2023, 2937-2954

Škobo, Milena; Petričević, Vedran, *Navigating the Challenges and Opportunities of Literary Translation in the Age of AI: Striking a Balance Between Human Expertise and Machine Power*, Društvene i humanističke studije (Online), Vol. 8, N° 2, 2023, 317-336

Sloculn, Jonathan, *A Survey of Machine Translation: its History, Current Status, and Future Prospects*, in “Computational Linguistics”, Volume 11, N° 1, 1985, 1-17

Snover, Matthew; Dorr, Bonnie; Schwartz, Richard; Micciulla, Linnea; Makhoul, John, *A Study of Translation Edit Rate with Targeted Human Annotation*, in “Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers”, Cambridge, Association for Machine Translation in the Americas, 2006, 223-231

Somers, Harold, *Review Article: Example-based Machine Translation*, in “Machine Translation”, USA, Kluwer Academic Publishers, Volume 14, 1999, 113-157

Soricut, Radu; Bach, Nguyen; Wang, Ziyuan, *The SDL Language Weaver Systems in the WMT12 Quality Estimation Shared Task*, in “Proceedings of the 7th Workshop on Statistical Machine Translation”, Madison, Omnipress Inc., 2012, 145–151

Stahlberg, Felix, *Neural Machine Translation: A Review*, in “Journal of Artificial Intelligence Research”, San Francisco, AI Access Foundation, Volume 69, 2020, 343-418

Stein, Daniel, *Machine translation: Past, present and future*, in “Language technologies for a multilingual Europe TC3 III”, Berlin, Language Science Press, 2018, 5-17

Sumita, Eiichiro; Iida, Hitoshi; Kohyama, Hideo, *Translating with Examples: A New Approach to Machine Translation*, in “Proceedings of the 3rd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language”, Linguistics Research Centre, University of Texas at Austin, USA, 1990, 203-212

Tang, Jie; Kan, Min-Yen; Zhao, Dongyan; Li, Sujian; Zan, Hongying, *Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges*, in “Natural Language Processing and Chinese Computing: 8th CCF International Conference”, Berlin, Springer-Verlag, 2019, 563-574

Toosi, Amirhosein; Bottino, Andrea G.; Saboury, Babak; Siegel, Eliot; Rahmim, Arman, *A Brief History of AI: How to Prevent Another Winter (A Critical Review)*, in “PET Clinics”, Philadelphia, W.B. Saunders, Volume 16, Issue 4, 2021, 449-469

Toral, Antonio; Way, Andy, *Machine-assisted translation of literary text: A case study*, in “Translation Spaces”, John Benjamins Publishing Company, Volume 2, Issue 4, 2015, 240-267

Turcato, Davide; Popowich, Fred, *What is Example-Based Machine Translation?*, in “Recent Advances in Example-Based Machine Translation. Text, Speech and Language Technology”, Dordrecht, Springer Dordrecht, Volume 21, 2003, 59-81

Turian, Joseph P.; Shen, Luke; Melamed, I. Dan, *Evaluation of Machine Translation and its Evaluation*, in “Proceedings of Machine Translation Summit IX: Papers”, New Orleans, 2003

Vandaele, Jeroen, *Wordplay in Translation*, “Handbook of Translation Studies”, Vol. 2, John Benjamins Publishing Company, 2011, 180-183

Wang, Haifeng; Wu, Hua; He, Zhongjun; Huang, Liang; Church, Kenneth Ward, *Progress in Machine Translation*, in “Engineering”, Volume 18, Elsevier, 2022, 143-153

White, John S.; O'connell, Theresa A; Carlson, Lynn M, *Evaluation Of Machine Translation*, in “Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey”, March 21-24, 1993, 206-210

Winter-Froemel, Esme; Thaler, Verena; Demeulenaere, Alex, *The dynamics of wordplay and wordplay research: Approaches, contexts and traditions*, in “Cultures and Traditions of Wordplay and Wordplay Research”, Berlin, Boston, De Gruyter, 2018, 1-19

Xu, Feiyu; Uszkoreit, Hans; Du, Yangzhou; Fan, Wei; Zhao, Dongyan; Zhu, Jun, *Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges*, in “Natural Language Processing and Chinese Computing”, NLPCC 2019, Lecture Notes in Computer Science, Vol. 11839, Springer, 2019, 563-574

Zhang, Chengzhi; Seong, Goh Sang; Muhammad, Suzana Hj., *Restructuring a New Approach of Pun Translation Strategy*, in “Theory and Practice in Language Studies”, London, Academy Publications, Volume 4, N° 12, 2014, 2498-2506

Zhang, Tianyi; Kishore, Varsha; Wu, Felix; Weinberger, Kilian Q.; Artzi, Yoav, *BERTScore: Evaluating Text Generation with BERT*, in “ArXiv”, 2019, 1-43

Harrison, Tomvas (editor), *chrF – CHaRacter-level F-score*,

<https://machinetranslate.org/chrF>, 20/09/2024

ModernMT, ModernMT Homepage, <https://www.modernmt.com/>, 8/08/2024

Nagy, Oana (revisor), *Qualitivity*, <https://community.rws.com/product-groups/trados-portfolio/rws-appstore/w/wiki/2251/qualitivity>, 20/09/2024

RWS, Adaptive machine translation for translators – FAQ, <https://www.rws.com/language-weaver/adaptive-machine-translation-for-translators-faq/>, 22/09/2024

RWS, Language Weaver Government Solutions, <https://www.rws.com/language-weaver/government-solutions/>, 8/08/2024

Unbabel, COMET,

[https://unbabel.com/research/comet/#:~:text=COMET%20\(Crosslingual%20Optimized%20Metric%20for,\(such%20as%20MQM%20scores\)](https://unbabel.com/research/comet/#:~:text=COMET%20(Crosslingual%20Optimized%20Metric%20for,(such%20as%20MQM%20scores)), 20/09/2024

Summary

La presente tesi di laurea ha l'obiettivo di descrivere una ricerca condotta per valutare le prestazioni dei modelli di traduzione automatica neurale e della Intelligenza Artificiale generativa quando applicati a contenuti umoristici nei testi letterari, principalmente giochi di parole. Per condurre la ricerca sono stati selezionati cinque esempi di giochi di parole e contenuti umoristici dal romanzo di Lewis Carroll "Alice's Adventures in Wonderland", questi elementi sono poi stati tradotti utilizzando due motori per la traduzione automatica (RWS Language Weaver e Intento ModernMT) e due interfacce per utilizzare dei software di Intelligenza Artificiale generativa (Microsoft Edge Copilot e ChatGPT 3.5). In seguito alla traduzione, sono state condotte una serie di valutazioni sui risultati della traduzione, in modo tale da identificare le problematiche emerse durante il processo di traduzione con l'obiettivo di sviluppare possibili rimedi. La seconda fase dello studio si è concentrata sull'applicazione di questi rimedi: l'addestramento dei motori di traduzione automatica e il prompting per le interfacce sviluppate per l'utilizzo dei software di Intelligenza Artificiale generativa. In seguito all'applicazione dei rimedi, i contenuti umoristici scelti sono stati sottoposti a una seconda traduzione, la quale è stata valutata applicando gli stessi criteri della valutazione precedente per verificare l'efficacia dei rimedi identificati.

La traduzione degli elementi umoristici all'interno dei testi tende a porre diverse sfide ai traduttori. La traduzione di questi particolari elementi testuali, infatti, richiede non solo un'approfondita conoscenza della Source Language, ma anche del retroterra culturale del testo e dell'autore. Nonostante questo, ad oggi i traduttori che si approcciano a questa sfida possono contare su una serie di strumenti tecnologici che li possono aiutare. Come discusso in altre ricerche, i traduttori utilizzano quotidianamente nel loro lavoro vari strumenti tecnologici, come i CAT tools; per questo motivo, sarebbe interessante investigare l'effettiva utilità di strumenti come l'Intelligenza Artificiale generativa o la traduzione automatica per la traduzione del linguaggio umoristico. Un tentativo di studiare le possibili applicazioni dell'IA alla traduzione di elementi umoristici è stato fatto durante la serie di conferenze tenutesi in occasione del CLEF nel 2022 e 2023. Nel 2022 un

team di ricercatori ha tentato di valutare le capacità di comprensione del linguaggio umoristico da parte di traduttori e di modelli computerizzati; come dimostrato in occasione delle ricerche svolte nel 2022 e in seguito nel 2023, nonostante i modelli computerizzati basati sull'IA abbiano raggiunto elevati livelli di complessità, sono comunque necessari dei miglioramenti per permettere a questi software di svolgere in maniera accettabile compiti come la traduzione dei giochi di parole. Per quanto riguarda la traduzione automatica, è stato dimostrato come questo strumento incontri diverse difficoltà quando si tratta di tradurre esempi di linguaggio creativo. Strumenti e software come la traduzione automatica e l'Intelligenza Artificiale sono sicuramente utili ai traduttori, ma presentano una serie di limiti, soprattutto quando impiegati per tradurre esempi di linguaggio creativo o umoristico. Laddove ricerche precedenti hanno messo in luce i limiti di questi strumenti, il presente studio punta a identificare dei possibili metodi per superare questi limiti. La ricerca ha impiegato due tipi di strumenti: la traduzione automatica neurale e l'Intelligenza Artificiale generativa. Gli esempi di linguaggio umoristico utilizzati per l'analisi provengono dal romanzo "Alice's Adventures in Wonderland". Gli strumenti sopra elencati sono stati scelti in quanto sono tra le soluzioni più sviluppate e sofisticate, al punto da essere utilizzate da diversi traduttori professionisti per gli scopi più disparati. La presente ricerca adotta sia un metodo quantitativo, che un metodo qualitativo. Il metodo quantitativo sarà applicato tramite la scelta di un preciso numero di esempi di linguaggio umoristico da analizzare; il metodo qualitativo sarà applicato durante la fase di analisi, che punta a valutare i risultati ottenuti durante la fase di traduzione individuando gli errori commessi dai sistemi utilizzati. La prima fase della ricerca si concentrerà sulla valutazione dei risultati ottenuti dagli strumenti impiegati per la traduzione; in seguito a questa analisi preliminare, la seconda parte della tesi si concentrerà sull'applicazione di possibili strategie sviluppare per migliorare la performance degli strumenti utilizzati. In seguito all'applicazione di queste strategie, i risultati traduttivi saranno valutati di nuovo. I criteri di valutazione saranno gli stessi applicati nella prima fase, in modo tale da paragonare la prima e la seconda fase di valutazione. L'obiettivo sarà quello di confermare l'effettiva utilità delle strategie identificate per il miglioramento della performance degli strumenti oggetto di indagine. Prima di descrivere la ricerca, la tesi si concentrerà sulla storia e sugli sviluppi nel tempo di traduzione automatica e Intelligenza Artificiale.

La storia della traduzione automatica e dei suoi sviluppi è stata oggetto di diversi studi. Nonostante la traduzione automatica sia ormai diventata parte della vita quotidiana, la sua storia e il suo sviluppo non sono stati semplici come si potrebbe pensare. L'idea di una "traduzione automatica" venne sviluppata per la prima volta nel 1947, anche se il concetto di una macchina che potesse tradurre i testi in modo autonomo venne concepito una decina di anni prima in Francia e in Unione Sovietica. Tuttavia, l'effettiva ricerca per lo sviluppo di una traduzione automatica cominciò solamente negli Anni '40. Dopo circa un decennio di ricerche, i primi successi vennero ottenuti nel 1954, quando una macchina riuscì a completare la traduzione di un testo dal russo all'inglese. Tra gli Anni '50 e gli Anni '60 vennero poi sviluppati i tre approcci principali allo sviluppo della traduzione automatica: la traduzione diretta, il modello interlingua e il metodo transfer. Il metodo della traduzione diretta prevede la traduzione dal Testo di partenza al Testo di arrivo in seguito alla rimozione di tutte le inflessioni morfologiche da una stringa di testo, in modo tale da lavorare solo sulla forma base delle parole; in seguito, un dizionario bilingue è utilizzato per identificare le corrispondenze tra le parole del Testo di arrivo e quelle che devono essere utilizzate nel Testo di partenza. Il modello interlingua è diviso in due parti: la prima fase prevede l'analisi del Testo di partenza, il cui contenuto viene in seguito rappresentato utilizzando un linguaggio intermedio, chiamato appunto "interlingua". Questo linguaggio intermedio viene poi utilizzato come base per generare le frasi che compongono il Testo di arrivo. Il metodo transfer prevede invece l'utilizzo di una rappresentazione del Testo di partenza per generare il Testo di arrivo; inizialmente il Testo di partenza è convertito in una rappresentazione che mantiene diverse caratteristiche (dalla sintassi alla semantica) del testo da tradurre, in seguito questa rappresentazione viene convertita in un'analoga rappresentazione per il Testo di arrivo, dalla quale questo testo è poi generato. Dopo anni di ricerca, nel 1966 il comitato ALPAC suggerì di bloccare gli studi sulla traduzione automatica, in quanto non si erano ancora raggiunti dei risultati che potessero esser ritenuti soddisfacenti. Nonostante il parere dell'ALPAC, gli studi proseguirono negli Anni '70. In questo periodo vennero sviluppato diversi sistemi per la traduzione automatica, come SYSTRAN1 (sviluppato per la traduzione russo-inglese) e WEATHER (sviluppato per la traduzione inglese-francese delle previsioni meteo). A partire dagli Anni '80 l'interesse per questo campo di ricerca si riaccese, con i finanziamenti che

cominciarono ad arrivare anche da parte di aziende private. In questo periodo, cambiò anche l'approccio allo sviluppo della traduzione automatica: se in precedenza l'attenzione era rivolta alle caratteristiche sintattiche del testo da tradurre, ora l'attenzione veniva rivolta agli aspetti semantici del testo, con il modello interlingua che cominciò a guadagnare popolarità. Inoltre, gli Anni '80 furono anche il decennio in cui cominciò la ricerca sull'approccio statistico, alla base della Statistic Machine Translation (SMT). Gli Anni 2000 sono un momento molto importante nel campo della traduzione automatica, in quanto vedono l'emergere dell'approccio basato sul corpus, che diventa l'approccio principale sostituendo l'approccio basato sulle regole. Questo tipo di approccio è basato sull'utilizzo di dizionari bilingue e regole scritte manualmente, utilizzati per tradurre dalla lingua di partenza a quella di arrivo. Le regole venivano utilizzate per vari scopi, come il trasferimento lessicale, la gestione degli elementi morfologici del testo, o l'analisi sintattica. L'approccio basato sulle regole era l'elemento fondamentale della Rule-based Machine Translation (RBMT), che effettuava il processo di traduzione in due fasi: un'analisi sintattica, morfologica e semantica del Testo di partenza, seguita dalla generazione del Testo di arrivo tramite una conversione strutturale basata su strutture interne. Il sistema in questione prevedeva l'utilizzo di dizionari e grammatiche creati manualmente dai linguisti, i quali avevano anche il compito di scrivere manualmente le regole che il sistema avrebbe dovuto seguire per funzionare correttamente: tutto ciò rendeva l'approccio Rule Based particolarmente lento e dispendioso.

A partire dai primi Anni 2000, l'approccio Rule Based venne sostituito da un approccio basato sull'utilizzo di corpora, soprattutto bilingue. Questo tipo di approccio è alla base di tre modelli di traduzione automatica: Example-based Machine Translation (EBMT), Statistical Machine Translation (SMT) e Neural Machine Translation (NMT). Il modello EBMT fu proposto per la prima volta negli Anni '80; il problema più grande riscontrato è la difficoltà per questo tipo di modello di ottenere buoni risultati quando i corpora utilizzati non presentano frasi particolarmente simili. Il modello SMT fu proposto negli Anni '90 ed è basato sulla capacità della macchina di apprendere in maniera autonoma come tradurre sulla base di una grande quantità di dati statistici; con questo modello, la macchina è in grado di stimare (utilizzando il Teorema di Bayes) la probabilità che una determinata frase possa essere tradotta con un'altra frase presente in uno dei corpora

bilingue utilizzati per l'apprendimento. In poche parole, il modello SMT funziona sulla base di corpora bilingue contenenti frasi che rappresentano le possibili traduzioni dalla lingua di partenza a quella di arrivo. A sua volta, questo modello si divide in due sotto modelli: Word-based SMT, che prende in considerazione le singole unità lessicali; Phrase-based SMT, che prende in considerazione i sintagmi e non le singole parole. In generale, questo sistema presenta degli enormi vantaggi: ad esempio, la possibilità di apprendere sulla base di una serie di dati statistici permette al sistema SMT di lavorare anche con lingue sconosciute, in quanto non sono necessari dizionari specifici o regole appositamente scritte. Dagli studi emerge tuttavia che mentre l'approccio SMT ottiene risultati decisamente migliori quando si tratta di disambiguare le soluzioni traduttive o nella scelta delle parole corrette, l'approccio RBMT ottengono risultati migliori per quanto riguarda l'ordine delle parole, la sintassi e la coerenza del Testo di arrivo. I sistemi SMT divennero progressivamente sempre più popolari, fino al lancio di un motore di traduzione automatica basato proprio su questo tipo di approccio da parte di Google nel 2006. L'approccio Example-based utilizza dati provenienti da corpora linguistici in modo tale che la macchina identifichi o "ricordi" degli esempi di traduzione provenienti dai corpora in questione. Il processo di traduzione si svolge estraendo e selezionando sintagmi o gruppi di parole equivalenti dai corpora utilizzati per l'apprendimento. Fondamentalmente, l'approccio EBMT funziona sulla base di analogie tra i dati linguistici estrapolabili dai corpora. Il metodo in questione è stato sviluppato cercando di riproporre nel modo più fedele possibile il processo di traduzione di un essere umano: la prima fase della traduzione consiste in un'analisi della frase, seguita dal calcolo della similitudine tra i segmenti e i gruppi di parole sulla base dei dati presenti nei corpora. L'ultima fase prevede la ricomposizione dei dati linguistici ottenuti tramite il processo dell'analogia, che porta alla creazione del Testo di arrivo. La Rule-based Machine Translation è stata sviluppata per permettere la traduzione tra lingue dissimili, come l'inglese e il giapponese, in quanto la similitudine delle due lingue (dal punto di vista strutturale e grammaticale) non è necessaria per il corretto funzionamento della macchina. Il problema principale presentato da questo tipo di modello risiede negli specifici requisiti dei corpora che devono essere utilizzati, in quanto l'allineamento tra la lingua di partenza e la lingua di arrivo non deve presentare errori; inoltre, è

necessario identificare quali blocchi di parole il sistema dovrebbe analizzare per ottenere dei buoni risultati.

Il modello attualmente più utilizzato è la Neural Machine Translation, che segna un passo avanti rispetto ai due modelli precedenti. La ricerca su questo tipo di modello cominciò già negli ultimi anni del XX secolo, ma le limitazioni tecniche imposero di sospendere le ricerche, che ripresero nel XXI secolo. In questi anni si è assistito all'integrazione dei modelli NMT all'interno dei modelli SMT; oltre a questo, negli anni tra il 2013 e il 2015 ha ripreso vita il filone di ricerca interamente dedicato allo sviluppo di modello esclusivamente NMT, con il rilascio del primo motore di traduzione automatica NMT proprio nel 2015 da parte di Baidu. Similmente ai modelli EBMT, questo modello cerca di replicare i passaggi che un traduttore umano seguirebbe nel processo traduttivo. Una prima fase prevede la "lettura" della frase del Testo di partenza, in seguito viene generato un Testo di arrivo in base alla comprensione della frase originale da parte della macchina. Questo tipo di approccio non richiede nessun tipo di regola scritta manualmente, così come non richiede vocabolari compilati dai linguisti o modifiche al database dei corpora. Il modello in questione apprende sulla base del cosiddetto "deep learning", ovvero la capacità di apprendere in maniera diretta e indipendente dai dati caricati sul sistema. L'approccio NMT sviluppa ulteriormente il sistema alla base dei modelli SMT: queste macchine sono basate sui "modelli lineari", ovvero modelli in cui un determinato numero di elementi che rappresentano potenziali traduzioni di una frase sono pesate in base parametri specifici; la traduzione è data da un punteggio generale ottenuto in base a questi parametri. Il sistema NMT si basa sui modelli lineari, ma aggiunge una serie di livelli ulteriori ai due presenti nelle macchine SMT. Questi livelli in più sono definiti "nodi nascosti" e sono stati sviluppati appositamente per ridurre o eliminare gli interventi ingegneristici volti ad identificare dei pattern utili alla traduzione: grazie ai nodi nascosti la macchina scopre questi pattern in autonomia ed è quindi in grado di imparare da sola. I modelli neurali generalmente sono formati da due componenti: un network di codificazione e un network di decodificazione. Il primo mappa la frase del Testo di partenza, generando un vettore che viene utilizzato dal network di decodificazione per produrre la traduzione. Questi vettori vengono utilizzati dal modello NMT per comprendere meglio la lingua del Testo di partenza. Dato che non è necessario nessun intervento umano per il suo funzionamento, la traduzione automatica

neurale si è imposta come modello predominante. Per questo motivo, il presente studio utilizza due motori per la traduzione automatica basati su modelli di traduzione automatica neurale: RWS Language Weaver e Intento ModernMT.

Una parte molto importante del presente studio è rappresentata dalla valutazione delle traduzioni ottenute tramite i motori sopracitati. La valutazione della traduzione automatica è un filone di studi che va avanti da molti anni, con le prime ricerche condotte a partire dagli Anni '90. L'obiettivo principale di quest'area di ricerca è lo sviluppo di criteri oggettivi per la valutazione degli output forniti dai motori di traduzione automatica, in modo tale da ridurre l'elemento soggettivo nella valutazione di queste traduzioni. Nei decenni successivi, sono state sviluppate diverse metriche, in quanti l'interesse in quest'area di ricerca cresceva sempre di più. La necessità di sistemi di valutazione oggettivi è cresciuta anche grazie al fenomeno della globalizzazione. Se da un lato è appurato che la qualità della traduzione automatica non equivale a quella di una traduzione umana, la globalizzazione ha reso necessario produrre un gran numero di traduzioni in poco tempo, cosa che non lascia tempo per il controllo della qualità della traduzione da parte di un essere umano; di conseguenza, è emersa la necessità di utilizzare in maniera massiccia dei motori di traduzione automatica, i quali devono essere valutati in maniera oggettiva per poter essere sicuri della buona qualità delle traduzioni. Una delle metriche di valutazione più utilizzate è BLEU, basata sul paragone tra il testo tradotto e un testo di riferimento prodotto da un traduttore umano. Per la valutazione degli output il presente studio utilizzerà le metriche presenti nella suite fornita da Intento Modern MT: BLEU, TER, hLEPOR, COMET, chrF++, and BERTScore.

BLEU è l'acronimo di "Bilingual Evaluation Understudy". L'obiettivo principale dietro lo sviluppo di questa metrica era quello di sviluppare un sistema di valutazione che fosse il più simile possibile alla valutazione umana. BLEU è diventata rapidamente la metrica di valutazione più utilizzata grazie alla sua facilità di computazione e alla somiglianza con i giudizi di revisori umani. L'idea alla base della metrica è quella di paragonare la traduzione effettuata dalla macchina con quella effettuata da un essere umano. Di solito, delle buone traduzioni tendono ad avere un numero relativamente alto di parole in comune con la traduzione usata come riferimento. La metrica identifica queste parole in comune e assegna un

punteggio tra 0 e 1: più alto è il punteggio più simile è la traduzione al riferimento. Per essere più precisi, questa metrica paragona dei gruppi di parole, noti come *n*-gram: gli *n*-gram della traduzione effettuata dalla macchina vengono paragonati a quelli del riferimento, contando il numero di raggruppamenti uguali. L'algoritmo in base al quale BLEU funziona usa come valore principale la misura della precisione, grazie alla quale i raggruppamenti di parole vengono contati solamente una volta. Per migliorare la performance di BLEU è stata sviluppato il codice SacreBLEU, che risolve alcuni problemi. Il codice in questione è stato implementato nella versione di BLEU utilizzata per il presente studio.

La metrica TER (Translation Edit Rate) è descritta come più intuitiva rispetto alle altre metriche, in quanto è stata sviluppata per calcolare la quantità di modifiche necessarie per far sì che il segmento tradotto e valutato sia identico al riferimento utilizzato dalla metrica. Fondamentalmente TER determina quante correzioni dovrebbe apportare un traduttore umano per ottenere una traduzione corretta. Dato che TER misura il numero minimo di modifiche necessarie, più alto è il punteggio assegnato, meno accettabile è la traduzione analizzata. È importante sottolineare come tutte le modifiche abbiano lo stesso punteggio, quindi non ci sono modifiche che vengono considerate più influenti di altre. Per rendere la metrica più oggettiva, sono stati apportati dei miglioramenti, giungendo allo sviluppo della metrica nota come HTER (Human-targeted Translation Edit Rate), che viene applicata nel presente studio. Come è stato dimostrato, i punteggi assegnati da BLEU e TER tendono ad essere particolarmente vicini: di conseguenza, queste saranno le metriche utilizzate per la valutazione nella presente ricerca.

hLEPOR è l'acronimo di "harmonic mean of enhanced Length Penalty, Precision, *n*-gram Position difference Penalty and Recall". Questa metrica prende in esame il numero di *n*-gram simili presente nella traduzione prodotta dalla traduzione automatica e nel testo usato come riferimento, assegnando un punteggio che varia da 0 (il minimo) a 1 (il massimo). Questa metrica è stata sviluppata per ovviare a un problem che, secondo gli studiosi, le altre metriche presentano: ovvero, la presenza di combinazioni linguistiche con cui le metriche non ottengono performance ottimali. La metrica hLEPOR implementa una serie di penalità inedite, come la Enhanced Length Penalty (applicata sia alle frasi troppo brevi che a quelle troppo lunghe), o la *N*-gram Position Difference Penalty (che paragona l'ordine

delle parole del testo di riferimento con quello del testo prodotto dalla traduzione automatica).

La metrica COMET (Crosslingual Optimized Metric for Evaluation of Translation) assegna un punteggio sulla base di informazioni ricavate sia dal Testo di partenza, che dal testo di riferimento. Secondo gli sviluppatori, un grande punto di forza di COMET risiede nella vicinanza tra le sue valutazioni e le valutazioni prodotte da valutatori umani, i quali potrebbero addirittura essere sostituiti dalla metrica. COMET rappresenterebbe un passo avanti rispetto alle altre metriche, dato che non prende in considerazione solamente le caratteristiche lessicali; inoltre, implementa due modelli distinti, definiti Estimator model e Translation Ranking model. Questi modelli permettono di operare una regressione diretta sul punteggio di qualità assegnato e di minimizzare la distanza tra quella che viene considerata un'ipotesi migliore e sia il testo usato come riferimento, sia il Testo di partenza.

chrF++ (CHaRacter-level F-score++) è stata introdotta nel 2015 e assegna un punteggio in base alla somiglianza tra l'output prodotto da un motore per la traduzione automatica e un testo di riferimento. Questa metrica, a differenza di altre che applicano lo stesso metodo, è basata sugli n -gram dei caratteri, non sugli n -gram di parole.

La metrica BERTScore (Bidirectional Encoder Representations from Transformers) è stata sviluppata per risolvere alcuni problemi presentati dalle altre metriche basate sugli n -gram. Innanzitutto, le altre metriche si fermano a un'analisi superficiale, in quanto si limitano a fornire una valutazione sulla base del numero di raggruppamenti uguali nel testo analizzato e nel testo di partenza; questo impedisce ad esempio di valutare correttamente una traduzione sotto forma di parafrasi, dato che non vengono rilevati n -gram simili. Inoltre, le altre metriche penalizzano i cambiamenti dell'ordine delle parole che influiscono pesantemente sulla semantica, rendendo la semantica del testo analizzato diversa da quella del testo di riferimento. Per ovviare a questi problemi, BERT effettua due valutazioni: vengono calcolate sia la capacità di richiamare il testo di riferimento, sia la precisione della traduzione.

Negli ultimi anni l'Intelligenza Artificiale è stata applicata al campo della traduzione: la Neural Machine Translation è l'esempio più recente di questo tipo di applicazioni. Secondo alcune ricerche, l'Intelligenza Artificiale, se usata correttamente potrebbe portare a risultati addirittura migliori di quelli ottenuti con i motori per la traduzione automatica. Così come la traduzione automatica, anche

questo tipo di software è caratterizzato da una storia particolarmente lunga e complessa. Le origini dell'IA infatti sono collocabili negli Anni '40, quando Alan Turing ideò una macchina per decodificare i messaggi in codice dei nazisti. In seguito alla creazione della macchina, proprio Turing cominciò a sviluppare l'idea alla base dell'Intelligenza Artificiale, tanto che proprio a lui si deve uno dei test tutt'ora usati per determinare il livello di "intelligenza" di una macchina: il cosiddetto "Test di Turing". Sulla base delle ricerche condotte da Turing, nel 1943 venne proposto un primo modello di Intelligenza Artificiale: una macchina in grado di apprendere in maniera autonoma grazie a una struttura interna costituita da neuroni artificiali che replicava il funzionamento dei neuroni presenti nel cervello umano. Nel 1958 venne proposta una versione migliorata di questo modello, chiamata Perceptron, in grado di apprendere grazie a un processo di trial and error.

Il termine "Intelligenza Artificiale" venne utilizzato per la prima volta nel 1956, anno a partire dal quale questo campo di ricerca progredì grazie a una serie di successi, come il programma ELIZA, un primo tentativo di creazione di un chatbot. Tra gli Anni '70 e gli Anni '80 ci fu una brusca interruzione dei fondi dedicati alla ricerca sull'IA. Nonostante questo e nonostante il calo dell'interesse nei confronti della ricerca sull'IA, questi decenni videro comunque alcuni importanti passi avanti, come l'algoritmo di discesa del gradiente, usato per ottimizzare i network neurali, o l'algoritmo di retropropagazione dell'errore, che migliora le capacità di apprendimento delle reti neurali.

Le reti neurali rappresentano uno spartiacque nella storia dell'IA, in quanto proprio lo sviluppo di questo network permise ai sistemi di affrontare e risolvere problemi più complessi; grazie a queste reti, infatti, non era più necessario basare i sistemi su regole scritte a mano dai programmatori, cosa che rendeva l'IA efficace solo nel caso in cui il problema da risolvere richiedeva ragionamenti formali. Laddove la formalizzazione non era necessaria, l'IA non era in grado di portare a termine il compito, almeno fino allo sviluppo delle reti neurali, grazie alle quali il sistema poteva autonomamente apprendere da dati esterni e adattarsi alla situazione. La ricerca sulle reti neurali esplose negli Anni '90. L'esempio migliore dei risultati di questa ricerca è lo sviluppo di Deep Blue, un software di Intelligenza Artificiale sviluppato per giocare a scacchi; Deep Blue fu in grado di battere il Gran Maestro e campione del mondo Garry Kasparov.

Negli ultimi anni, i miglioramenti nell'ambito dell'immagazzinamento dei dati e l'introduzione della GPU hanno permesso all'Intelligenza Artificiale di svilupparsi ulteriormente. L'algoritmo di apprendimento funziona tanto più efficacemente quanti più dati gli sono forniti; di conseguenza, le migliorate capacità dei sistemi informatici nella memorizzazione ed elaborazione dei dati hanno permesso all'IA di registrare grandi sviluppi. Attualmente, il filone di ricerca principale si concentra sul "Deep Learning", un'architettura di sistema introdotta nel 2012 in grado di addestrare più livelli di neuroni simultaneamente. I sistemi basati sulle reti neurali hanno ottenuto grandi successi. Ad esempio, nel 2016 il software Deep Mind's AlphaGo è riuscito a battere il campione mondiale del gioco da tavolo Go, per anni considerato troppo complesso perché una macchina riuscisse a imparare a giocarci. Un grande problema associato allo sviluppo dell'Intelligenza Artificiale risiede nell'effettiva possibilità di spiegarne il funzionamento. I sistemi di Intelligenza Artificiale utilizzati oggi e basati sui Deep Neural Networks hanno raggiunto livelli di complessità talmente alti che risulta particolarmente difficile comprendere come prendano decisioni. Dato che l'IA sarà applicata negli anni a venire a una serie di settori fondamentali (come il settore medico), è molto importante cercare di identificare e spiegare i processi che portano ad effettuare determinate scelte. L'IA è dunque un campo di ricerca particolarmente interessante, che promette di ottenere risultati particolarmente promettenti. In particolar modo, le elevate possibilità di interazione uomo-macchina rendono questo strumento particolarmente utile dal punto di vista della traduzione. Infatti, mentre i sistemi di traduzione automatica sono limitati alla sola traduzione del testo, non dando grandi possibilità di interazione da parte del traduttore umano, le interfacce per la comunicazione con i sistemi IA permettono di instaurare un dialogo con la macchina, offrendo la possibilità di usare una serie di prompt per suggerire le modifiche necessarie per ottenere la traduzione desiderata.

Come già detto, traduzione automatica e Intelligenza Artificiale sono oggi ampiamente utilizzate dai traduttori, tanto che la traduzione può ormai essere considerata una sorta di interazione uomo-computer. Ai fini di questa tesi, verranno analizzati i tentativi di applicare traduzione automatica neurale e IA generativa alla traduzione letteraria, considerata la forma di traduzione più complessa per i traduttori umani. Un testo letterario pone molte difficoltà ai traduttori umani, come ad esempio l'interpretazione delle espressioni idiomatiche, la necessità di

preservare lo stile letterario, o di rendere in maniera appropriata le sfumature linguistiche. Un elemento di difficoltà aggiuntivo sono i giochi di parole e in generale il linguaggio umoristico che certi testi presentano. Il linguaggio umoristico generalmente è strettamente legato alla cultura e alle caratteristiche grammaticali di una lingua, per questo motivo la sua traduzione richiede competenze culturali e creative che di solito una macchina non possiede.

La traduzione automatica è stata applicata alla traduzione di vari testi letterari in diversi studi. Le ricerche condotte hanno concluso che ci sono prospettive incoraggianti per questa tipologia di studio. Nel 2022 è stato condotto uno studio sulla traduzione automatica applicata ai testi umoristici da parte di alcuni ricercatori dell'università di Padang, in Indonesia. L'obiettivo della ricerca era identificare gli errori commessi dalle macchine utilizzate per tradurre i testi, in modo tale da determinare i motivi per cui i motori di traduzione automatica utilizzati non riuscissero a produrre una traduzione accettabile dei testi presi in esame. I ricercatori crearono quattro categorie di errore: Errori lessicali, Errori sintattici, Errori semantici ed Errori pragmatici; stando ai risultati della ricerca, la tipologia di errore più comune era quella legata alla sintassi. Stando alle conclusioni dello studio, sembra che i motori per la traduzione automatica che implementano dei sistemi di Neural Machine Translation siano quelli più efficaci. Se questa ricerca si è concentrata sull'identificazione degli errori commessi, lo studio descritto in questa tesi cercherà di effettuare un passo avanti, individuando un potenziale metodo per evitare che la macchina commetta degli errori in fase di traduzione. Per condurre il presente studio sono stati usati motori per la traduzione automatica basati sui modelli neurali, in quanto rappresentano l'ultimo ritrovato della ricerca in questo campo. Un altro studio che si è concentrato sulla possibilità di applicare la traduzione automatica neurale alla traduzione letteraria è lo studio condotto da Evgenij Matusov, descritto in un articolo del 2019. Lo studio in questione è stato svolto addestrando dei motori per la traduzione automatica, dimostrando che il post editing delle traduzioni ottenute con dei motori addestrati è più veloce rispetto al post editing delle traduzioni ottenute con motori privi di addestramento. Oltre a questo, la ricerca condotta da Matusov ha portato all'elaborazione di un nuovo sistema di categorizzazione degli errori commessi dai motori per la traduzione automatica; questo sistema per la valutazione umana è applicato anche nella presente ricerca. Le categorie di errore individuate da Matusov sono: Grave errore

di significato; Errore di significato non grave; Parola sconosciuta o errore di segmentazione; Errore di consistenza/traduzione del termine; Errore di risoluzione del pronome; Errore di locuzione; Errori di omissione, inserzione, ripetizione; Grave errore di sintassi; Errore di sintassi non grave; Errore di tono/registro. La ricerca condotta da Matusov ha prodotto risultati particolarmente interessanti, in quanto ha dimostrato innanzitutto la necessità di una valutazione umana dei risultati prodotti dalla traduzione automatica. Inoltre, Matusov ha elaborato un sistema di valutazione particolarmente interessante, che verrà applicato nel presente studio.

Dal punto di vista dell'applicazione dell'IA alla traduzione dei testi letterari, lo scoglio più grande da superare consiste nella produzione di una traduzione fedele al testo originale in grado di mantenere le sfumature artistiche del testo. È evidente dunque che Intelligenza Artificiale e traduzione automatica affrontano gli stessi problemi quando applicate alla traduzione di un testo letterario. La differenza principale risiede nel maggiore livello di interazione uomo-macchina reso possibile dalle interfacce di Intelligenza Artificiale generativa: per poter funzionare correttamente, è necessario che gli utenti di queste interfacce sappiano utilizzare i comandi corretti (una pratica che prende il nome di “prompt engineering”). Diversi studi hanno dimostrato che l'Intelligenza Artificiale ha fatto progressi nel campo della traduzione letteraria; ad esempio, uno studio condotto da Škobo e Petričević ha dimostrato che l'IA è in grado di identificare correttamente gli errori di traduzione presenti all'interno di un testo letterario, giungendo a un livello di valutazione molto simile alla valutazione umana. Se l'IA è in grado di identificare correttamente gli errori di traduzione, allora è possibile che sia riuscita ad apprendere come effettuare una traduzione di tipo letterario. Un altro studio condotto da Alkodimi, Alqahtani, e Al-Wasy si è concentrato sulla traduzione e retroversione di alcuni testi letterari utilizzando degli strumenti per la traduzione basati sull'Intelligenza Artificiale. I promettenti risultati ottenuti dimostrano che l'IA può essere di grande aiuto nel processo di traduzione letteraria, soprattutto per quanto riguarda la retroversione. Gli strumenti basati sull'IA sono di grande aiuto in quanto non solo sono in grado di comprendere il contesto del Testo di partenza, ma sono anche in grado di identificare gli errori commessi in fase di traduzioni, correggendoli se necessario. Oltre a questo, l'IA può avere un ruolo importante dal punto di vista dell'insegnamento; come dimostrato da una ricerca condotta nel 2023, gli studenti di varie discipline (come linguistica, letteratura o traduzione)

preferiscono di gran lunga strumenti come ChatGPT rispetto a strumenti come Google Translate come supporto per la traduzione.

Nonostante gli ottimi livelli raggiunti da traduzione automatica e Intelligenza Artificiale quando applicate alla traduzione, i giochi di parole e, in generale, il linguaggio umoristico tendono a rappresentare un elemento di difficoltà non certo secondario.

Il linguaggio umoristico è uno dei meccanismi linguistici più studiati. Attualmente, le due teorie più accreditate e utilizzate per l'analisi dei meccanismi umoristici sono la Script-based Semantic Theory (SSTH) e la Generalized Theory of Verbal Humour (GTVH). Secondo la SSTH, un testo può essere considerato umoristico nel caso in cui il testo in questione sia compatibile (anche solo in parte) con due diversi "script", i quali sono opposti da un ben preciso punto di vista. Le opposizioni possono essere situazione/irreale, buono/cattivo, sesso/no sesso, denaro/no denaro, oppure vita/morte. Secondo questa teoria, un testo umoristico presenta cinque componenti principali: lo spostamento da un modo di comunicazione diretto a un modo di comunicazione che ha un sottotesto implicito; il testo vero e proprio; due *script* compatibili con il testo che si sovrappongono parzialmente; un'opposizione tra i due *script*; un "interruttore" che permette di spostarsi da uno *script* all'altro. La teoria GTVH non fa altro che espandere questi cinque componenti, in modo da raggiungere una comprensione migliore dei testi umoristici e del loro funzionamento. Secondo questa teoria, i parametri che definiscono il testo umoristico sono sei: Opposizione degli *script*; Meccanismo logico; Situazione; Target; Strategia narrativa; Lingua. Come è possibile notare, il linguaggio umoristico è particolarmente complesso e stratificato, motivo per cui presenta una serie di problemi ai traduttori umani e in particolare alle macchine.

I giochi di parole sono presenti nella nostra vita quotidiana, non solo all'interno dei testi letterari, comparando in maniera spontanea, così come in contesti accuratamente studiati, come le pubblicità. I giochi di parole sono un'area linguistica particolarmente complessa, dato che si presentano sotto molte forme, come anagrammi, palindromi o acrostici. Il gioco di parole è quindi prevalentemente grafemico; esiste però una particolare tipologia di gioco di parole, identificata in inglese come "pun". Questo tipo di gioco di parole prevede due sensi diversi per l'interpretazione del termine che fa scattare il gioco, ma non necessariamente la presenza di due termini diversi. Sono presenti *pun* generate dalla

doppia lettura della parola, così come *pun* che derivano da un'ambiguità sintattica o morfologica. Dato che sono basati su caratteristiche estremamente specifiche di una lingua, *pun* e giochi di parole rappresentano un elemento testuale particolarmente complesso da tradurre, anche se sarebbe scorretto considerarli intraducibili. Si possono infatti applicare diverse strategie per la loro traduzione. Innanzitutto, sarebbe preferibile un approccio testuale, che prenda in considerazione questo elemento linguistico come parte di un testo, senza tentare di tradurlo "in isolamento". In particolare, Delabastita ha formalizzato una serie di strategie applicabili alla traduzione delle *pun*: Da *pun* a *pun*; Da *pun* a *non-pun*; Da *pun* a *punoide*; Traduzione zero; Copia diretta; Trasferimento; Aggiunta; Aggiunta di nuovo materiale testuale; Tecnica editoriale. Come risulta evidente, la complessità delle *pun* e il gran numero di tecniche traduttive applicabili rende il compito del traduttore particolarmente difficile quando si approcciano questo tipo di elementi testuali. È necessaria non solo una profonda conoscenza della lingua di partenza, ma anche del contesto in cui il Testo di partenza è stato prodotto; di conseguenza, viene naturale chiedersi se una macchina possa essere in grado di portare a termine un compito così complesso.

Sono stati svolti diversi studi focalizzandosi sulla capacità da parte degli strumenti tecnologici di tradurre correttamente giochi di parole, *pun* e altri tipi di linguaggio umoristico. Alcuni di questi studi hanno portato anche allo sviluppo di applicativi particolarmente interessanti. Uno di questi è PunCAT, uno strumento sviluppato per l'individuazione delle *pun* all'interno di un testo. Sebbene il programma non sia sviluppato per tradurre automaticamente gli elementi testuali identificati, fornisce comunque un grande aiuto ai traduttori, dato che permette di esplorare i vari significati associabili a una *pun*. Altri studi sono stati condotti in occasione del percorso JOKER, organizzato nell'ambito del progetto CLEF nel 2022 e 2023. Questi studi si sono concentrati sull'identificazione traduzione delle *pun*, in modo tale da identificare delle strategie per migliorare la traduzione automatica e l'Intelligenza Artificiale quando applicate alla traduzione delle *pun*. Stando ai risultati ottenuti dalle ricerche del 2022, strumenti come la traduzione automatica sono in grado di identificare i giochi di parole senza problemi, i quali emergono nel momento in cui questi elementi testuali devono essere correttamente interpretati e tradotti. Risultati simili sono stati ottenuti nel 2023, quando le ricerche si sono concentrate sull'Intelligenza Artificiale. I risultati di queste ricerche hanno

dimostrato che in alcuni casi Intelligenza Artificiale e traduzione automatica sono in grado di ottenere risultati accettabili, ad esempio nel caso in cui la *pun* abbia una traduzione esatta nella lingua di arrivo.

Affinché il presente studio metta in difficoltà gli strumenti utilizzati verranno prese in esame solamente quegli esempi di *pun* e giochi di parole particolarmente complessi (ad esempio non verranno tradotti elementi testuali che hanno un equivalente esatto nella lingua di arrivo). In particolar modo, verranno presi in esami quei casi in cui l'elemento umoristico è legato all'intero contesto e quei casi in cui l'elemento testuale umoristico presenta delle caratteristiche tipiche della lingua inglese, tali per cui la riproposizione in italiano risulta particolarmente difficile.

Le strategie di miglioramento che saranno testate sono strategie direttamente applicabili dai traduttori, senza ricorrere a strategie che richiedono interventi di ingegneria informatica sui sistemi analizzati. Per migliorare la forma di interazione uomo-computer devono necessariamente essere identificate delle strategie che permettano l'intervento diretto sugli output e sulla macchina.

La fase di analisi preliminare degli elementi linguistici presi in esame punta a spiegare i motivi dietro la scelta degli esempi tradotti e la difficoltà nella loro traduzione. Ai fini dell'analisi sono stati selezionati cinque esempi di giochi di parole, *pun* e linguaggio umoristico dal romanzo "Alice's Adventures in Wonderland". Per ottenere risultati particolarmente significativi, sono stati evitati quegli esempi che permettono una facile traduzione; ovvero, quei casi in cui la traduzione non avrebbe presentato particolari problematiche, dato che il meccanismo umoristico non è basato su nessuna peculiarità morfologica, lessicale o culturale.

In questa sezione gli esempi utilizzati per la ricerca sono stati analizzati ed è stata proposta una possibile traduzione cercando di applicare le strategie elaborate da Delabastita. I casi di linguaggio umoristico presi in analisi presentano dunque determinate caratteristiche che rendono particolarmente difficile la loro traduzione. Nell'ordine, queste caratteristiche sono:

- la presenza delle parole omofone *tail* e *tale*. L'omofonia presente in inglese non è riproducibile traducendo in italiano, in quanto gli equivalenti di queste parole (rispettivamente coda e racconto/storia) non sono omofoni. Non è quindi

possibile ricorrere alla strategia che prevede la traduzione della *pun* del Testo di partenza con una *pun* equivalente nel Testo di arrivo

- la presenza di una rappresentazione grafemica del particolare tipo di pronuncia della parola *arm* da parte di parlanti con accento irlandese. Per poter comprendere correttamente questo passaggio è necessario possedere precise conoscenze culturali, senza le quali il linguaggio umoristico non risulta comprensibile
- la presenza in lingua inglese della categoria neutra del pronome. La lingua inglese infatti non comprende solamente l'utilizzo dei pronomi di genere maschile e femminile, ma anche di genere neutro. La loro assenza in italiano rende particolarmente difficile la traduzione del passaggio in questione
- la possibilità di utilizzare il verbo inglese *to bite* per descrivere azioni diverse. Infatti, questo verbo descrive sia l'azione di *beccare* da parte di un uccello, sia il *pizzicare* della senape. Dato che le diverse azioni sono descritte in italiano da verbi diversi e non da un unico verbo, risulta chiaro come una traduzione italiana che riproponga lo stesso schema della *pun* originale sia complessa
- la presenza di un omofonia tra il suono [t] e la parola *tea*. Dato che in italiano il suono [t] non è un omofono della parola *tè*, non è possibile mantenere la stessa struttura della *pun* originale

La terza parte dello studio si è concentrata sull'effettiva traduzione e analisi degli esempi di linguaggio umoristico selezionati. I testi selezionati sono stati tradotti dall'inglese all'italiano utilizzando i motori per la traduzione automatica Intendo Modern MT e RWS Language Weaver e le interfacce Microsoft Edge Copilot e ChatGPT 3.5. In seguito alla traduzione, i risultati ottenuti sono stati analizzati e valutati sia utilizzando delle metriche automatiche (principalmente BLEU e TER), sia utilizzando i criteri di Matusov per una valutazione umana manuale. La valutazione manuale è stata effettuata non solo perché necessaria per avere un'idea più chiara e precisa dell'effettiva performance dei sistemi, ma anche perché le metriche in questione sono state sviluppate per la valutazione degli output prodotti dalla traduzione automatica, non per valutare le traduzioni prodotte dall'IA generativa. Per quanto riguarda i criteri di Matusov, le categorie di errore riferite al Significato (Severe e Minor Meaning Error) sono state rivisitate in modo tale da

essere applicabili agli esempi di testo analizzati: verranno considerati errori di significato quelle traduzioni che non ripropongono il Testo di partenza in modo tale da mantenere intatto il contenuto umoristico del testo.

In seguito, sono state effettuate la traduzione e valutazione secondo i criteri di Matusov degli output della traduzione automatica. La valutazione di Intento ModernMT e RWS Language Weaver ha portato ai seguenti risultati:

- Language Weaver ha ottenuto un punteggio peggiore rispetto a ModernMT, dato che ha commesso 6 errori gravi per nell'ambito del significato, contro i 4 di ModernMT
- Mentre Intento Modern MT non ha commesso errori relativi alla Sintassi, RWS Language Weaver ne ha commesso 1

In generale, è possibile notare come la performance di Language Weaver sia stata inferiore rispetto a quella di ModernMT, nonostante entrambi i motori abbiano avuto delle difficoltà nella traduzione del primo e terzo segmento. Inoltre, entrambi i software non sono stati in grado di tradurre correttamente il Testo di partenza, mantenendo intatto l'elemento umoristico.

Per quanto riguarda la valutazione tramite le metriche, sono state utilizzate tutte le metriche contenuti nella suite MT Hub fornita da Intento ModernMT; tuttavia, solo BLEU e TER sono state prese in considerazione. La traduzione usata come riferimento è la traduzione del romanzo realizzata da Alessandro Ceni, utilizzata per l'edizione kindle di "Alice nel Paese delle Meraviglie", pubblicata nel 2024. Dalle metriche emerge la concordanza tra BLEU e TER, dato che entrambe hanno assegnato un punteggio maggiore a ModernMT; questo significa non solo che le traduzioni di ModernMT presentano un maggiore numero di n -gram in comune con il riferimento, ma anche che queste traduzioni richiedono un numero minore di modifiche per avvicinarsi al testo di Ceni. Anche l'analisi dei punteggi assegnati da TER ai singoli segmenti è interessante, dato che ModernMT ha ottenuto risultati migliori in ogni singolo segmento.

Per condurre l'analisi e la valutazione dei risultati ottenuti dall'Intelligenza Artificiale, non sarà dato alla macchina nessun tipo di *prompt* particolare. La performance ottenuta dalle due macchine è molto simile, tenendo presente che anche gli *output* prodotti sono molto simili. Se non si prende in considerazione un

errore minore relativo alla sintassi commesso da GPT, si può affermare che le due intelligenze artificiali abbiano ottenuto una performance molto simile.

Dal punto di vista di una valutazione complessiva dei quattro software utilizzati, emerge chiaramente che l'IA generativa ha ottenuto una performance nettamente migliore rispetto alla traduzione automatica.

I risultati ottenuti in questa prima fase saranno usati come paragone per la seconda valutazione degli strumenti presi in esame.

L'ultima parte della ricerca si è concentrata sull'applicazione delle strategie per il miglioramento degli *output*: l'addestramento dei motori per la traduzione automatica e il *prompting* per le interfacce di Intelligenza Artificiale generativa.

Per quanto riguarda l'IA generativa, è stata applicata la pratica del *prompt engineering*, consistente nel fornire alla macchina dei *prompt* specifici per far sì che produca il risultato sperato. In seguito alla questa fase, i risultati delle traduzioni sono stati rivalutati utilizzando i criteri di Matusov già precedentemente applicati. I risultati sono stati soddisfacenti: in generale è possibile notare un miglioramento nella performance sia di ChatGPT 3.5, che di Microsoft Edge Copilot. GPT grazie ai prompts non ha commesso nessun errore grave relativo al significato, ma solo un errore minore; Copilot ha comunque commesso un errore grave relativo al significato. Sebbene la performance di Copilot sia effettivamente migliorata, il sistema sembra comunque meno sensibile ai *prompt* rispetto a GPT.

Per quanto riguarda l'addestramento dei motori per la traduzione automatica, è necessario specificare che questi motori possono essere addestrati con specifiche memorie di traduzione, le quali non sono altro che corpora paralleli di testi in una lingua di partenza e in una lingua di arrivo. Questi corpora contengono testi relativi a un dominio ben preciso e specifico, come ad esempio l'ambito legale. La memoria di traduzione usata nel presente studio per l'addestramento dei motori di traduzione automatica è stata creata manualmente utilizzando la funzione "Alignment" presente in Trados Studio 2021; la memoria consta di circa 16 000 segmenti. I testi tradotti con i motori addestrati sono stati valutati sia manualmente, sia utilizzando le metriche.

In base ai risultati ottenuti tramite la valutazione manuale, è possibile confermare come ModernMT ottenga una performance migliore rispetto a RWS Language Weaver, che ha totalizzato 6 errori gravi relativi al significato contro i 4 totalizzati

da ModernMT. L'addestramento ha permesso di ottenere risultati lievemente migliori per quanto riguarda la sintassi e la locuzione; tuttavia, gli elementi umoristici all'interno dei testi non sono stati tradotti in maniera corretta nonostante l'addestramento. Dai risultati ottenuti è comunque possibile notare come RWS Language Weaver sia più sensibile all'addestramento, al punto da ottenere risultati nettamente migliori dal punto di vista della sintassi, della consistenza e della locuzione rispetto ai risultati ottenuti nella prima fase dello studio.

La valutazione tramite le metriche è stata svolta seguendo lo schema adottato nella prima fase dello studio: nonostante tutte le metriche siano state presentate, solo i risultati di BLEU e TER sono stati presi in considerazione. Innanzitutto, è interessante notare come le metriche non siano concordi: mentre BLEU ha valutato meglio i risultati ottenuti da ModernMT, TER ha assegnato un punteggio migliore a RWS Language Weaver. Si può dunque concludere che i segmenti tradotti da Language Weaver sono più simili alla traduzione usata come riferimento: il motivo per cui TER ha assegnato il punteggio migliore indicherebbe che sono necessari meno cambiamenti per portare il segmento tradotto ad essere uguale a quello di riferimento. Questa potrebbe essere un'ulteriore conferma del fatto che Language Weaver recepisce meglio l'addestramento: i miglioramenti nella sintassi, nella locuzione e nella consistenza del testo hanno reso questi segmenti più simili a una traduzione prodotta da un essere umano. In generale, i punteggi assegnati dalle metriche sostengono l'ipotesi che RWS Language Weaver reagisca meglio all'addestramento rispetto a ModernMT. Anche l'analisi dei punteggi assegnati da TER ai singoli segmenti supporta questa ipotesi. Ad esempio, i segmenti 3 e 4 sono stati valutati con lo stesso punteggio per entrambi i motori di traduzione automatica, mentre prima Language Weaver aveva ottenuto un punteggio decisamente inferiore: questo dimostra come l'addestramento ha permesso a Language Weaver di raggiungere lo stesso livello di ModernMT.

Comparando i risultati complessivi della seconda valutazione, è possibile confermare che l'IA generativa ottiene risultati migliori rispetto ai motori di traduzione automatica. Soprattutto in seguito al *prompting*, è possibile notare che il risultato è nettamente migliorato. L'addestramento dei motori per la traduzione automatica non ha permesso di raggiungere i buoni risultati che il *prompting* ha permesso di ottenere con l'IA generativa; tuttavia, è stato possibile registrare dei

lievi miglioramenti, segno che questa pratica ha un effetto positivo su questi software.

L'ultimo capitolo è dedicato alle conclusioni. L'analisi condotta ha permesso di dimostrare che i risultati migliori sono stati ottenuti tramite pratiche di *prompt engineering* applicate alle interfacce di Intelligenza Artificiale: rispetto ai risultati ottenuti in seguito all'addestramento dei motori per la traduzione automatica, ChatGPT 3.5 e Microsoft Edge Copilot hanno prodotto traduzioni decisamente migliori, in grado di riprodurre in maniera accettabile gli aspetti umoristici del Testo di partenza nel Testo di arrivo. Per quanto riguarda i *prompt* forniti alla macchina, è risultata evidente la necessità di fornire alla macchina istruzioni quanto più corrette e precise possibile. Innanzitutto, è necessario dare informazioni precise sul contesto; in secondo luogo, potrebbe essere necessario spiegare le sfumature comiche all'IA generativa per ottenere una corretta traduzione. I traduttori hanno inoltre la possibilità di suggerire le parole che sarebbe più opportuno utilizzare.

Il *training* dei motori per la traduzione automatica non ha permesso di ottenere risultati accettabili. Tutti gli errori relativi al significato (Severe e Minor Meaning Error) sono rimasti anche dopo l'addestramento. Nonostante questo, l'addestramento ha dimostrato di essere un buon metodo per risolvere altre problematiche relative alla sintassi o alla locuzione. In particolar modo, il motore RWS Language Weaver è stato quello più sensibile all'addestramento, mostrando netti miglioramenti.

La presente ricerca ha prodotto risultati interessanti, ma apre anche a interessanti prospettive per quanto riguarda gli studi futuri. Ad esempio, sarebbe possibile creare una memoria di traduzione per l'addestramento costituita solo da giochi di parole e altri esempi di linguaggio umoristico; inoltre, potrebbe essere una possibilità quella di creare una memoria che contenga in parti uguali diversi tipi di linguaggio umoristico e le relative traduzioni, in modo tale da testare le capacità dei sistemi per la traduzione automatica di utilizzare diverse strategie traduttive. Un altro problema affrontato durante la fase sperimentale riguarda la mancata (o poco intuitiva) integrazione dei motori addestrati all'interno dell'interfaccia MT Hub resa disponibile da Intento ModernMT: potrebbe essere utile integrare meglio il motore addestrato, in modo tale da rendere più intuibile il suo utilizzo.

Sebbene la fase di *prompt engineering* non abbia presentato particolari problemi, sarebbe comunque interessante proseguire le ricerche su questa pratica. Ad esempio, sarebbe possibile identificare e categorizzare le diverse tipologie di *prompt*. Una volta fatto questo, potrebbe essere possibile applicare le stesse categorie di istruzioni a macchine diverse, in modo tale da sviluppare una sorta di test standardizzato per le diverse interfacce di IA.

Lo studio in questione ha dimostrato come una delle difficoltà più grandi per i motori di traduzione automatica sia quella di identificare gli esempi di linguaggio umoristico. Per questo motivo, sarebbe interessante cercare di integrare a questi motori dei software sviluppati appositamente per il riconoscimento di questo tipo di linguaggio, come PunCAT. In questo modo, i motori per la traduzione automatica sarebbero adeguatamente supportati nel loro lavoro.

La ricerca condotta ha quindi prodotto risultati particolarmente interessanti, sebbene non siano stati raggiunti tutti i risultati sperati. Nonostante questo, i risultati aprono alla possibilità di sviluppare ulteriormente la ricerca su Intelligenza Artificiale generativa e traduzione automatica neurale, le quali possono rappresentare dei validi aiuti per i traduttori.