PRZEKŁADAJĄC NIEPRZEKŁADALNE IX

MACIEJ KUR

FEASIBILITY OF DEEPL, GOOGLE AND MICROSOFT MT SYSTEMS IMPLEMENTATION INTO THE TRANSLATION PROCESS IN THE ENG -> PL LANGUAGE PAIR

GDAŃSK UNIVERSITY PRESS

FEASIBILITY OF DeepL, GOOGLE AND MICROSOFT MT SYSTEMS IMPLEMENTATION INTO THE TRANSLATION PROCESS IN THE ENG –> PL LANGUAGE PAIR

PRZEKŁADAJĄC NIEPRZEKŁADALNE IX

SERIES EDITORS

OLGA KUBIŃSKA

WOJCIECH KUBIŃSKI

THE BOOK IS ALSO PUBLISHED IN THE SERIES
DISERTATIONES LAUDATISSIMAE UNIVERSITAS GEDANENSIS

MACIEJ KUR

FEASIBILITY OF DeepL, GOOGLE AND MICROSOFT MT SYSTEMS IMPLEMENTATION INTO THE TRANSLATION PROCESS IN THE ENG –> PL LANGUAGE PAIR

GDAŃSK UNIVERSITY PRESS
GDAŃSK 2020

Series Editors
Olga Kubińska
Wojciech Kubiński

Reviewer
Professor Henryk Kardela

Technical proofreading
Katarzyna Jopek

Series design
Karolina Johnson

Typesetting and page layout
Michał Janczewski

This publication is financed by funds provided by the Vice-Rector for Research of the University of Gdańsk as part of a competition for outstanding doctoral dissertations, and by funds provided by the Institute of English and American Studies and the Dean of the Faculty of Languages

ISBN 978-83-8206-099-7

Wydawnictwo Uniwersytetu Gdańskiego
ul. Armii Krajowej 119/121, 81-824 Sopot
tel.: 58 523 11 37; 725 991 206
e-mail: wydawnictwo@ug.edu.pl
www.wyd.ug.edu.pl

Online bookstore: www.kiw.ug.edu.pl

Printing and binding
Zakład Poligrafii Uniwersytetu Gdańskiego
ul. Armii Krajowej 119/121, 81-824 Sopot
tel. 58 523 14 49

Table of contents

Acknowledgements

First and foremost, I would like to express my gratitude to Professor Olga Kubińska. Without her invaluable guidance and constant motivation this publication would never have been completed. I would also like to thank Professor Piotr Fast and Professor Henryk Kardela, who reviewed my PhD thesis and whose kind remarks helped me to significantly improve the quality of this work. My gratitude also goes to all members of the thesis committee, who recommended my PhD thesis for publication. My special thanks go to Professor Wojciech Kubiński, who offered numerous constructive suggestions when reviewing my work, and to Dr Aneta Lica for her substantial help during error categorization process. I also wish to thank Professor Sharon O'Brien, Professor Lucia Specia, Professor Philip Koehn, Professor Alon Lavie, and many more researchers and members of the international MT community whom I had the privilege to talk with or listen to during the last couple of years and who certainly sparked and maintained my interest in this area of translation studies. Last but certainly not least, I would like to thank all translators and professionals who agreed to take part in the experiment described below. I hope that the results of my work will help us to face future challenges in the best possible way.

Introduction

In May 2017, the Translation Automation User Society (TAUS), an independent translation industry organization, issued a report containing projections related to the use of machine translation technologies (MT) in modern translation industry (Joscelyne et al., 2017). On the basis of interviews carried out with over 50 MT vendors, the authors of the report came to the conclusion that "MT is on a journey with no return ticket" and that we should "expect very widespread adoption" of machine translation in the years to come. At the same time, TAUS predicted a rise in importance of post-editing (PE)[1], which is "likely to replace translation memory leveraging as the primary production environment in industrial translation in the next five years".

Such a vision seems to be supported by the contents of financial reports published in 2017 and 2018 by Global Market Insights company (Global Market Insights, 2017), Grand View Research company (Grand View Research, 2018) and by Mordor Intelligence organization (Mordor Intelligence, 2018). All these reports describe various marketing studies of MT industry and provide promising forecasts for the development of the field in the nearest future. With compound annual growth rate between 14.6% and 19%, the estimated value of machine translation market is bound to rise from USD 433.0 million (2016) to at least 983.3 million, or approximately 1.5% of the entire translation market value, by 2022 (Dranch et al., 2018).

Such high expectations must have been influenced by changes occurring within the global translation market and by the rapid development of the MT technology that took place in the recent years and caused an increase in the popularity of various machine translation solutions. Factors like progressive globalization, increasing source content volumes, growing

[1] Process of error elimination and MT quality improvement, carried out by a human post-editor.

market demands, introduction of new legal requirements regulating the types of translated texts and relatively invariable processing capacity of human translators (DePalma, 2009) cause a pressing need for a solution that could allow for the increase in the number of translated texts and for the decrease in the translation costs. Considering constant development of the field and recent introduction of neural architectures into machine translation engines (Wu et al., 2016), MT technology may provide such a solution.

This possibility does not remain unnoticed, as an increasing number of large companies begin to invest in the machine translation industry. The abovementioned reports list IBM, Microsoft, SDL, Lionbridge Technologies, Omniscien Technologies, Lingotek, RWS Holdings, Welocalize, Smart Communications, Systran International, AppTek Partners, Google, Cloudwords and PROMT among the major corporations interested in the development of this technology. At the same time, 18 out of 20 top LSP companies in the world (Nimdzi, 2019) offer machine translation services to their clients. All things considered, it seems that in global perspective, the popularity of MT will grow, while its capabilities will be further developed.

In order to determine whether this global trend can be observed in Poland, a survey with questions about the use of MT technology was prepared and sent to translation agencies associated in Polish Association of Translation Companies[2] offering services in the English → Polish language pair. The analysis of collected answers indicates that only 7.1% of the interviewed agencies use machine translation technologies on a regular basis, 28.6% use them occasionally and as much as 64.3% do not use any kind of MT solution at all. When asked about the reason behind their refusal to use machine translation, 88.9% of these companies expressed their concerns for the quality of the final product, while 11.1% stated that they possess sufficient human resources to successfully cope with the content volumes they process. Furthermore, 55.6% of these companies admitted that they forbid their translators to use MT systems during translation work. At the same time, however, 55.6% of them are interested in the implementation of MT solutions at some point in the future.

Among the agencies that decided to implement machine translation solutions into their processes, 80% use proprietary MT engines trained and

[2] An organization representing Polish translation and localization service providers of various size. For the full list of its members, see: (PSBT).

developed by their own employees, while 20% use a dedicated SaaS solution. None of these companies admitted to using any of the popular, directly accessible, free or relatively cheap systems[3], such as Google Translate, Microsoft Translator or DeepL. As far as prices are concerned, 80% of the interviewed companies offer MT services at a lower price in comparison with traditional translation, while 20% of them have the same rates for traditional and machine translation. At the same time, 60% of such companies have developed dedicated sets of guidelines related to the assessment of quality of machine translated texts.

Apart from the preferences related to machine translation, the survey included several questions about post-editing. The collected answers indicate that only 35.7% of the interviewed companies offer post-editing services to their clients. Most of these agencies have reduced rates for post-editing in comparison with translation (20% of companies offer post-editing at 50% of the translation rate, 60% offer post-editing at 51%–99% of the translation rate and 20% have the same rates for post-editing and translation). Finally, as much as 80% of these companies have developed a set of guidelines and instructions for their post-editors, but only 20% organize special trainings in post-editing techniques.

As the predominant model of operation of the interviewed companies involves outsourcing of translation jobs, a similar survey was sent to freelance translators, in order to learn if there is any difference in attitude towards machine translation. The provided answers indicate that MT systems are used by 58.3% of freelancers on a regular basis. Most of them tend to use free and accessible systems (71.4% – Google Translate, 57.1% – DeepL), with only 14.3% investing in payable solutions (AdaptiveMT was the only option marked by the respondents). At the same time, most of the interviewed translators expressed some concerns related to the quality of machine translated texts, with 42.9% describing it as “Poor” and 28.6% choosing the option “Average”. On the other hand, as much as 28.6% of respondents considered the quality provided by the MT engines they use to be “Good”. When asked about the post-editing assignments, 58.3% of the interviewed freelancers declared that they post-edited machine translated texts at some point in their careers.

The analysis of data provided by the subjects that took part in both surveys point to the general conclusion that there is a difference in opinions

[3] Not exceeding a €20 fee for monthly use.

on machine translation technology among Polish translation companies and Polish freelance translators. Most translation agencies tend to avoid machine translation solutions and post-editing processes due to reservations concerning the quality of the final product. If MT technology is used in Poland at a corporate level, it is usually done through dedicated and proprietary systems, with avoidance of free and cheap engines, and in accordance to a separate set of PE guidelines and QA requirements. On the other hand, most freelance translators take a more positive attitude towards machine translation and post-editing. They use MT systems in their work on a regular basis, mostly choosing free and easily accessible engines. At the same time, they tend to assess the quality of machine translated texts better than the agencies.

The primary aim of this publication is to determine whether it is feasible to implement some of the most popular and free or relatively cheap machine translation systems into the translation workflow in the English → Polish language pair. In order to achieve this goal, a set of outputs provided by DeepL Translator, Google Neural Machine Translation System and Microsoft Translator engines was collected and processed together with translated and post-edited versions of individual segments produced by a group of professional Polish translators. The acquired data were then analyzed with the use of machine translation evaluation metrics (HTER scores, quality rankings and edit time parameters) to determine whether the quality of the provided outputs and amount of effort related to their post-editing was adequate enough to enable efficient implementation of the analyzed engines into the professional environment.

The secondary aim of this publication is to construct a reliable basis for the development of post-editing guidelines that could be used by Polish professionals and translation agencies during quality assessment and improvement of machine translated texts. By detecting, analyzing and categorizing the most common errors made by the studied systems, an attempt was made to create a list of recommendations that could be helpful during the performance of post-editing processes.

The first chapter contains a description of the research work carried out in relation to the MT technology from its very beginnings in the late 1940s, through the early rule-based and transfer-based models and statistical machine translation methods, to the modern day achievements related to neural network technology. Apart from the descriptions of various types of machine translation systems, their advantages and drawbacks,

the chapter contains information about some of the most popular methods of MT quality evaluation, including manual evaluation (accuracy, fluency and comprehension measurements, segment rankings, HTER scores) and automatic evaluation metrics (precision, recall, BLEU, NIST and METEOR) as well as some remarks concerning post-editing techniques and PE effort measurement methods. The main purpose of this chapter is to provide theoretical background for the study and information about solutions implemented during the data analysis process.

The second chapter is focused on the description of the study itself, with much emphasis placed on the details related to the experiment used to collect the translation and post-editing data. Individual sections present information about the preparatory activities, selection and processing of experimental corpora, engines and CAT tool configuration, selection and profiles of participants, details of tasks performed during the experiment and methodology related to data acquisition and analysis.

The third chapter presents the results obtained during the analysis of the collected information with descriptions of conclusions reached at individual stages of the study. It also contains a set of recommendations for Polish translation professionals interested in implementing machine translation technology and post-editing processes in their professional environments.

The last chapter contains the conclusions reached on the basis of the obtained results, a brief summary of the research question and the final opinion concerning the feasibility of implementation of the studied systems into processes carried out in the Polish translation market.

Chapter 1
The status of research

Even though some fundamental concepts and intellectual basis for a machine capable of translating words between various languages appeared as early as in the 17th century (Hutchins, 1995; Hutchins, 2003; Chen, 2016; Schwartz, 2018), the more precise idea for the creation and practical application of such a machine was produced in the early 1930s[1]. After World War II, when the potential of the first computers began to be explored, three scientists from the USA and Great Britain, Warren Weaver, Norbert Wiener and Andrew Booth, discussed the possibility of incorporating computing devices into the translation process. In 1949, the results of their discussions were presented by Weaver in his memorandum entitled simply "Translation", where he quotes his own letter to Wiener:

> Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while.

With such a statement and several more detailed arguments concerning meaning, context and correlation between translation and cryptography (Weaver, 1949), this short document is nowadays widely considered to be a cornerstone for all subsequent elaborations on the subject.

[1] In 1933 two patents on mechanical linguistic devices were granted independently in France (Georges Artsrouni) and in Russia (Petr Trojanskij). Artsrouni's invention was a bilingual dictionary based on paper tape, which could be used to find target language equivalents of source terms. Trojanskij's idea was more elaborate and based on a three-stage translation process involving careful preparation of the source text (pre-editing), mechanical translation into the target language and correction of mistakes made by the device (post-editing) (Schwartz, 2018).

After somewhat discouraging beginnings of research on MT potential[2] and first attempts made to construct fully-operational and useful machine translation systems[3], Weaver's ideas were slowly, but consistently developed by numerous scholars, mathematicians and linguists. The following section presents a brief summary of their research and introduces some of the most important concepts and ideas used during the experiment.

1.1. Basic terms and definitions

The terms "machine translation" and "MT" will be used in this publication interchangeably to describe a fully automatized process of translation conducted by a machine in one or several language pairs. At present, this process in most cases involves the use of computing devices with specialized software called "MT systems" or "MT engines". Even though there are very many MT engines operating on the basis of various ideas, the underlying concept for most of them can be summarized and described with the use of the diagram presented in Figure 1.

Figure 1. General MT engine rule of operation

During these three stages, the source input undergoes various processes, depending on the type of the applied engine and adopted approach. Some of the most fundamental models used in the development of MT systems are described in Section 1.2 below.

[2] In 1959, in his report on the current state of machine translation technology, Yehoshua Bar-Hillel, Israeli philosopher, linguist and organizer of the first MT conference held in June 1952 at the Massachusetts Institute of Technology, claimed that attaining FAHQMT (fully automatic high-quality machine translation) is impossible not only at the present stage of technological progress, but in principle (Arnold et al., 1994).

[3] The results of the most notable one were presented on 8th January 1954 in New York. The system designed by the employees of IBM and prof. Leon Dostert's team of scientists from Georgetown University translated several short messages written in Russian "into good English without human intervention" (Plumb, 1954).

The operation of many modern engines does not require any human activity other than the provision of the source text. Some approaches, however, may involve the performance of some additional actions aimed at the improvement of the target output quality. These actions include, for instance, "pre-editing", "post-editing" and "training".

Pre-editing takes place before input processing and is closely related to the notion of the controlled language (Mitamura, 1999; O'Brien, 2003). It involves the preparation of source input in such a way to make it easier for the machine to process it. During pre-editing, the source text is simplified (made clearer without losing any significant meaning), shortened (any redundant words are eliminated) and rearranged (made more logical and explicit) (Pym, 1988).

Post-editing, on the other hand, is a process that involves the detection and elimination of mistakes made by the MT system and takes place after the target output is synthesized. It is usually performed by trained post-editors or translators[4] with experience in MT output revision. As the concept of post-editing was crucial during the performance of the experiment described in the following chapters, it will be analyzed in much greater detail in Section 1.4.

Some of the modern MT engines can be adjusted or "trained" to improve the quality and precision of their operation. The process of training involves the preparation of mono- or multilingual corpora that are later loaded into the engine to allow for the creation of a database that can be used as a basis for statistical analysis of probability of target language phrases and sentences. This way the amount of resources allocated to the "baseline" system can be increased, which consequently leads to the improvement of the target output general quality.

Numerous researchers active in the MT community have recently focused much of their attention on the problem of the machine translation output quality measurement. Due to the considerable number of various models and ideas, there is a growing need for a reliable method of quality assessment that could provide translators, post-editors and end-users of MT systems with measurable information helpful in choosing an appropriate solution (Snover et al., 2006; Snover et al., 2009; Wisniewski, 2013). In general, the currently available quality evaluation methods can be divided into

[4] Although some more recent scientific advances foreshadow the appearance of efficient automatic post-editing (APE) systems (cf. Simard et al., 2007; Junczys-Dowmunt & Grundkiewicz, 2016).

those that require human evaluators (assessment of fluency and adequacy, sentence ranking, reading comprehension tests) or annotators (analysis of translation error rate or edit distance) and those that are purely automatic (the so-called automatic evaluation metrics, or AEM, such as BLEU, METEOR or NIST). Even though during recent years much progress has been made in this area of MT-related study, it seems that consensus over the most reliable evaluation method is yet to be reached. An in-depth analysis of MT output evaluation methods is provided in Section 1.3 of this chapter.

1.2. Fundamental models

1.2.1. Direct model

Most of the early MT systems created in the 1950s and at the beginning of the 1960s were based on the so-called "direct" or "transformer" architecture (Hutchins & Somers, 1992; Arnold et al., 1994; Trujillo, 1999). This approach is based on a relatively rudimentary concept that involves the simplest possible method of replacing source words with target equivalents, which requires minimum number of processes related to the analysis and synthesis of the translated content. The level of complexity of such engines varied from quite primitive bilingual dictionary look-up systems to some more sophisticated solutions capable of categorizing lexical entities. In its most elaborate form, the direct machine translation process begins with the use of a parser – a program responsible for the preliminary analysis of the source text structure, identification of lexical categories of individual words, determination of their functions, reduction of inflected forms and development of a database with basic-form entities categorized on the basis of their features. These entities are then looked up in a bilingual dictionary to find their equivalents in the target language.

The following step involves rough rearrangement of the equivalents with the use of a "transformer" – another program designed to use the information provided by the parser to render syntactically and grammatically plausible target sentences. In order to achieve this goal, the transformer applies a set of rules that govern, for instance, the order and inflectional forms of target words. As a result, a target language text is obtained, which carries the meaning of individual source words (parser) and conforms to the rules

of the target language (transformer). The general idea of the direct architecture can be summarized with the use of the diagram presented in Figure 2.

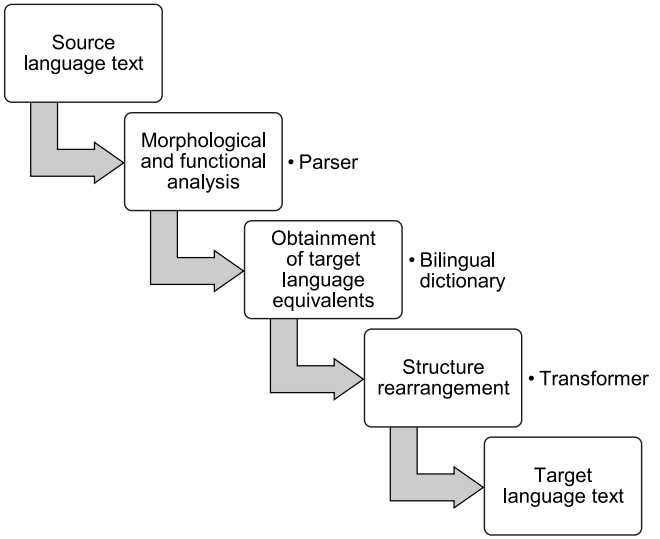

Figure 2. Direct model rule of operation

Even though some engines based on direct architecture were successfully developed and operated efficiently[5], this simple concept in general did not fulfill the expectations related to the target text quality, fluency and accuracy. It quickly became obvious that the linguistic structure of both source and target text is too complicated to allow for such a rudimentary approach. In terms of structure and syntax, the obtained target output was similar to the source input, which resulted in either extremely mistranslated or simply incomprehensible sentences[6]. It became evident that a new type of model was needed.

[5] Most notably the Canadian "Meteo" system, used to translate weather forecasts (Thouin, 1982).

[6] John Hutchins and Harold Sommers (1992) provide some examples of such incorrect outputs of direct machine translation from Russian to English:
Russian source input: Včera my tselyi čas katalis' na lodke.
Human translation: Yesterday we went out boating for a whole hour.
Direct MT output: Yesterday we the entire hour rolled themselves on a boat.

1.2.2. Indirect models

After the period of disillusionment with the results of the direct approach towards machine translation, researchers turned their attention to the possibility of applying the so-called indirect methods (Slocum, 1985; Hutchins & Somers, 1992; Schwartz, 2018). Whereas in the case of direct models the output synthesis process relied heavily on the structure of the source and target languages, the supporters of the indirect approach postulated the need for some intermediate representation, developed as a result of the source input analysis, that could be used to facilitate the transitional processing and improve the results of target output synthesis. Systems based on such an idea were first proposed in the 1960s and were developed in various forms until the 1990s[7]. They can be generally grouped under two categories: transfer-based models and interlingual models.

The transfer-based models reshaped the basic idea of direct approach by introducing an additional element into the processing stage. This element, called "transfer", requires a much more elaborate and detailed analysis of the source input, performed with the use of parsers and sets of syntactic, morphological, semantical and contextual rules of the source language. The results of this analysis are then used to develop an intermediate representation of the source text. This representation is then "transferred" into its target language equivalent, constructed in accordance with the corresponding target language rules. Such a transformed representation could be then used to synthesize the target output. The diagram presented in Figure 3 illustrates the general idea behind the operation of a transfer-based system.

In theory, due to the application of source and target language rules, such an approach should give better results in comparison with less complicated direct systems, which employed only basic rearrangement of target equivalents of source lexical units. Some researchers claim that if an ideally formed set of virtually all linguistic rules could be developed and used in such a system, the generated output should be composed of perfectly well-formed target sentences, free of any syntactical, grammatical and stylistic issues. The only mistakes such an ideal system could make would include errors related to translation accuracy (Arnold et al., 1994).

[7] Although some of their underlying concepts are developed and used in research on MT to this very day (cf. Lu et al., 2018).

PRZEKŁADAJĄC NIEPRZEKŁADALNE (Translating the Untranslatable) is a book series launched at the University of Gdańsk in 2000. It presents different aspects of translation problems – from a theoretical and analytical vantage point – which are particularly vital for translation studies. It covers broadly defined literary translation, translation for publishing and specialized translation. The series – published in English or Polish – includes contributions of Polish authors and Polish translations of texts by foreign authors.

ISBN 978-83-8206-099-7

Wydawnictwo Uniwersytetu Gdańskiego

CENTRE FOR TRANSLATION STUDIES
CENTRUM TRANSLATORYCZNE